

**Department of Environment and Agriculture**

**New Approaches to Ancient DNA: Using Novel Substrates to  
Characterise DNA Preservation and Past Biodiversity in Warm-  
Climate Ecosystems**

**Alicia Catherine Grealy**

**This thesis is presented for the Degree of  
Doctor of Philosophy  
of  
Curtin University**

**October 2016**

CURTIN UNIVERSITY

Department of ENVIRONMENT and AGRICULTURE



Curtin University

# NEW APPROACHES TO ANCIENT DNA

**USING NOVEL SUBSTRATES TO  
CHARACTERISE DNA PRESERVATION *and*  
PAST BIODIVERSITY *in* WARM-CLIMATE  
ECOSYSTEMS**

ALICIA C. GREALY  
BSc (HONS)

A thesis submitted in fulfillment of the requirements for the Degree of  
DOCTOR *of* PHILOSOPHY

October 2016

Supervisor: Prof. Mike Bunce  
Co-supervisor: Dr. James Haile

a.grealy@student.curtin.edu.au  
+61 431181057

Trace and Environmental DNA laboratory, Department  
of Environment and Agriculture, Curtin University,  
Bentley, WA 6102

— COPYRIGHT —

I warrant that this thesis contains no material that infringes the copyright of any other persons.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third-party copyright material reproduced in this thesis, or to use any of my own published work (e.g., journal articles) in which copyright is held by another party (e.g., publisher, co-author).

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

*Alicia Greal*  
ALICIA GREALY

**SIGNED** | Alicia Greal

**DATED** | 1 August 2016

—**DECLARATION**—

I declare that I have stated clearly and fully in the thesis the extent of any collaboration with others. To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university.

*Alicia Grealy*  
**ALICIA GREALY**

**SIGNED** | Alicia Grealy

**DATED** | 1 August 2016

---

Dedicated to my partner,  
**- ALEXANDER GOFTON -**

*“My muse, my flame”*

---

## — ABSTRACT —

Ancient DNA (aDNA) is becoming an increasingly useful tool for studying past biodiversity. Such studies rely entirely upon the isolation and characterisation of highly fragmented DNA from fossil material. The advent of Next-Generation Sequencing (NGS) has transformed the field, as it allows deep sequencing of highly fragmented aDNA, either by shotgun or amplicon sequencing. As such, ever older and more degraded samples can be targeted, but it remains difficult to retrieve aDNA from fossils in warm, tropical environments using traditional techniques. This thesis focuses on developing and applying NGS techniques to novel aDNA templates. Here, we further optimise methods to extract and enrich endogenous aDNA from fossil ‘bulk bone’ and eggshell, and demonstrate how, when coupled with NGS, such aDNA can be used to explore past biodiversity and evolutionary processes in climates that are typically not conducive to the preservation of aDNA. Using fossil bulk bone from palaeontological and archaeological sites in South Australia and Madagascar, the utility of bulk bone aDNA analysis as an alternative tool for identifying past faunal assemblages and characterising aDNA preservation is evaluated. Such information is then used to speculate about the factors likely to have caused spatial and temporal changes in biodiversity (for example, human activity). The extraction and targeted enrichment of aDNA from the fossil eggshell of Madagascar’s extinct elephant birds is also developed and used to test hypotheses regarding phylogeny, time-of-divergence, taxonomy, phylogeography, and population history. The studies presented here illustrate the viability of these substrates as a tool for research into evolution, extinction, biodiversity and conservation.

— **TABLE of CONTENTS** —

<b>LIST OF FIGURES</b>	<b>pg xi-xiv</b>
<b>LIST OF TABLES</b>	<b>pg xv-xxi</b>
<b>ABBREVIATIONS</b>	<b>pg xxii-xxvi</b>
<b>EPIGRAPH</b>	<b>pg xxvii</b>
<b>PREFACE</b>	<b>pg xxvii-xxviii</b>
<b>ACKNOWLEDGEMENTS</b>	<b>pg xxix-xxx</b>
<b>CHAPTER 1   INTRODUCTION</b>	<b>pg 1-88</b>

1.1 USING ANCIENT DNA TO STUDY EVOLUTION: PAST, PRESENT AND FUTURE PERSPECTIVES (A LITERATURE REVIEW) . . . . .	pg 2-30
1.1.1 Introducing ancient DNA. . . . .	pg 2-3
1.1.2 Ancient (DNA) History 101: 1984-2000. . . . .	pg 3-6
1.1.3 ‘Damage Control’. . . . .	pg 6-11
1.1.4 The ‘Next-Generation’ Revelation. . . . .	pg 11-20
1.1.5 aDNA Today: 2001-2011. . . . .	pg 20-26
1.1.6 Present Problems and Prospects . . . . .	pg 26-30
1.2 AIMS AND SCOPE OF THESIS. . . . .	pg 30-31
1.3 STUDY SITES AND SPECIES. . . . .	pg 31-33
1.4 SIGNIFICANCE . . . . .	pg 33-34
1.5 REFERENCES. . . . .	pg 34-56
1.6 SUPPLEMENTARY INFORMATION. . . . .	pg 57-88
S1.6.1 Supplementary references. . . . .	pg 70-88

<b>CHAPTER 2   A CRITICAL EVALUATION OF HOW ANCIENT DNA BULK BONE METABARCODING COMPLEMENTS TRADITIONAL MORPHOLOGICAL ANALYSIS OF FOSSIL ASSEMBLAGES</b>	<b>pg 89-180</b>
--	------------------

2.1 PROLOGUE . . . . .	pg 90-92
2.1.1 Acknowledgements. . . . .	pg 91
2.1.2 Author contributions. . . . .	pg 91

2.1.3 Author declarations	pg 91-92
2.2 ABSTRACT	pg 94
2.3 INTRODUCTION	pg 95-98
2.4 MATERIALS AND METHODS	pg 98-104
2.4.1 Study systems and sample collection	pg 98-99
2.4.2 Dating	pg 99
2.4.3 Morphological assessment of fossil assemblage	pg 99
2.4.4 Experimental design	pg 99-100
2.4.5 DNA extraction of bulk bones	pg 100-102
2.4.6 Next-generation sequencing of DNA amplicons	pg 102-103
2.4.7 Sequence identification and bioinformatics analysis	pg 103
2.4.8 Statistical analysis	pg 104
2.5 RESULTS AND DISCUSSION	pg 104-115
2.5.1 Repeatability of the BBM method	pg 106-108
2.5.2 Accuracy of the BBM method	pg 109-112
2.5.3 Overall utility and further applications of the BBM method	pg 112-115
2.6 CONCLUSION	pg 115-116
2.7 REFERENCES	pg 116-124
2.8 SUPPLEMENTARY INFORMATION	pg 125-178
S2.8.1 Study systems and dating	pg 125-126
S2.8.2 Experimental design	pg 127
S2.8.3 DNA extraction of bulk bones	pg 127-130
S2.8.4 qPCR quantification	pg 130-131
S2.8.5 Next-generation sequencing of DNA amplicons	pg 131-133
S2.8.6 Sequence identification and bioinformatic analysis	pg 133-136
S2.8.7 Statistical analysis	pg 136-139
S2.8.8 Does fossil abundance correlate with read abundance?	pg 139-140
S2.8.9 NMDS analysis of the main contributions to variance in family richness	pg 141-142
S2.8.10 Sequencing coverage	pg 143-149
S2.8.11 Morphological taxonomic identification and abundance data	pg 150-155
S2.8.12 Comparison of FF and BC family assemblages identified by BBM	pg 156

S2.8.13 Biases and limitations _ _ _ _ _	pg 156-161
S2.8.14 Practical guidelines for palaeontologists and archaeologists_ _	pg 161-162
S2.8.15 Practical guidelines for geneticists_ _ _ _ _	pg 162-165
S2.8.16 Future directions_ _ _ _ _	pg 166-171
S2.8.17 Supplementary references_ _ _ _ _	pg 171-178
2.9 EPILOGUE _ _ _ _ _	pg 179-180

**CHAPTER 3 | AN ASSESSMENT OF ANCIENT DNA PRESERVATION IN HOLOCENE-PLEISTOCENE FOSSIL EXCAVATED FROM THE WORLD HERIAGE NARACOORTE CAVES, SOUTH AUSTRALIA** pg 181-254

3.1 PROLOGUE _ _ _ _ _	pg 182-
3.1.1 Acknowledgements_ _ _ _ _	pg 183
3.1.2 Author contributions_ _ _ _ _	pg 183
3.1.3 Author declarations_ _ _ _ _	pg 183-184
3.2 ABSTRACT _ _ _ _ _	pg 186
3.3 INTRODUCTION _ _ _ _ _	pg 187-189
3.4 MATERIALS AND METHODS _ _ _ _ _	pg 189-196
3.4.1 Study site and dating_ _ _ _ _	pg 189
3.4.2 Sample collection and bulk bone sampling_ _ _ _ _	pg 189-192
3.4.3 Sample preparation_ _ _ _ _	pg 192
3.4.4 DNA extraction and total DNA quantification_ _ _ _ _	pg 192-193
3.4.5 Amplicon sequencing_ _ _ _ _	pg 193
3.4.6 Shotgun sequencing_ _ _ _ _	pg 194
3.4.7 Bioinformatics and data analysis_ _ _ _ _	pg 194-195
3.4.8 Modeling of aDNA preservation and decay over time _ _ _ _ _	pg 195-196
3.4.9 Base composition analysis_ _ _ _ _	pg 196
3.5 RESULTS AND DISCUSSION _ _ _ _ _	pg 197-213
3.5.1 DNA preservation at RCEC_ _ _ _ _	pg 197-208
3.5.2 Biodiversity_ _ _ _ _	pg 208-213
3.6 CONCLUSION_ _ _ _ _	pg 213-214
3.7 REFERENCES_ _ _ _ _	pg 214-223
3.8 SUPPLEMENTARY INFORMATION_ _ _ _ _	pg 224-252
S3.8.1 Radiocarbon dating_ _ _ _ _	pg 224

S3.8.2 Experimental design	pg 225
S3.8.3 Implementation of standard aDNA protocols	pg 226
S3.8.4 Modifications to Dabney et al's. (2013) aDNA extraction method	pg 226
S3.8.5 Measuring total genomic DNA via spectrophotometry	pg 226
S3.8.6 Visualising genomic DNA with agarose gel electrophoresis	pg 226-227
S3.8.7 qPCR quantification of DNA and fusion-tagging	pg 227-228
S3.8.8 Modifications to the AMPure XP PCR purification protocol	pg 228
S3.8.9 Visualising PCR products and libraries with agarose gel electrophoresis	pg 228-230
S3.8.10 PCR reaction and thermocycling conditions for quantifying the sequencing library	pg 230
S3.8.11 Modifications to the MiSeq 300 v.2 kit	pg 230-232
S3.8.12 Modifications to the single-stranded library building method by Gansauge and Meyer (2013)	pg 232
S3.8.13 PCR reaction and thermocycling conditions for the shotgun library fusion-tag/indexing PCR	pg 232-234
S3.8.14 Modifications to the MiSeq 150 v.3 kit	pg 234
S3.8.15 Amplicon sequencing coverage	pg 235-236
S3.8.16 Base composition analysis from shotgun sequencing data	pg 237-239
S3.8.17 Quality and quantity of genomic DNA as determined by spectrophotometry and gel electrophoresis	pg 240-241
S3.8.18 qPCR amplification results	pg 242-245
S3.8.19 Full data table	pg 246
S3.8.20 Error and taxonomic resolution of the 12SMarsMini primer set	pg 246-251
S3.8.21 Supplementary references	pg 252
3.9 EPILOGUE	pg 253-254

**CHAPTER 4 | TROPICAL ANCIENT DNA FROM BULK ARCHAEOLOGICAL FISH BONE REVEALS THE SUBSISTENCE PRACTICES OF A HISTORIC COASTAL COMMUNITY IN SOUTHWEST MADAGASCAR** pg 255-300

4.1 PROLOGUE	pg 256-258
--------------	------------

4.1.1 Acknowledgements	pg 257
4.1.2 Author contributions	pg 258
4.1.3 Author declarations	pg 258
4.2 ABSTRACT	pg 260
4.3 INTRODUCTION	pg 261-263
4.4 MATERIALS AND METHODS	pg 264-267
4.4.1 Site description and dating	pg 264
4.4.2 Sample preparation	pg 265
4.4.3 aDNA extraction	pg 265
4.4.4 Metabarcoding and next-generation sequencing	pg 265-267
4.4.5 Taxonomic assignment	pg 267
4.5 RESULTS AND DISCUSSION	pg 267-274
4.6 CONCLUSION	pg 274-275
4.7 REFERENCES	pg 275-282
4.8 SUPPLEMENTARY INFORMATION	pg 283-298
S4.8.1 Bias, error and taxonomic resolution of the short <i>12S</i> fish primer set	pg 283-297
S4.8.2 Supplementary references	pg 298
4.9 EPILOGUE	pg 299-300

**CHAPTER 5 | EGGSHELL PALAEOGENOMICS: PALAEOGNATH EVOLUTIONARY HISTORY REVEALED THROUGH ANCIENT NUCLEAR AND MITOCHONDRIAL DNA FROM MADAGASCAN ELEPHANT BIRD (*AEPYORNIS* SP.) EGGSHELL** **pg 301-418**

5.1 PROLOGUE	pg 302-
5.1.1 Acknowledgements	pg 303
5.1.2 Author contributions	pg 304
5.1.3 Author declarations	pg 304
5.2 ABSTRACT	pg 307
5.3 INTRODUCTION	pg 308-311
5.4 MATERIALS AND METHODS	pg 311-317
5.4.1 Site description and collection of eggshell samples	pg 311
5.4.2 Dating and morphological identification of eggshell	pg 312

5.4.3 aDNA extraction	pg 312
5.4.4 Screening samples via amplicon sequencing	pg 312-313
5.4.5 Shotgun sequencing	pg 313
5.4.6 Targeted enrichment through hybridisation capture	pg 313
5.4.7 Quality control	pg 313-314
5.4.8 Mitochondrial genome assembly	pg 314
5.4.9 Identification of nuclear sequences	pg 315
5.4.10 Alignments and partitioning	pg 315-316
5.4.11 Phylogenetic analyses	pg 316-317
5.4.12 Molecular dating	pg 317
5.5 RESULTS	pg 318-328
5.5.1 Sample screening via amplicon sequencing	pg 319
5.5.2 Shotgun sequencing	pg 319
5.5.3 Bait enrichment	pg 320
5.5.4 Mitochondrial genome reconstruction and nuclear mapping	pg 320-321
5.5.5 Phylogeny reconstruction	pg 321-327
5.5.6 Molecular dating	pg 327-328
5.6 DISCUSSION	pg 328-334
5.6.1 Mode and tempo of palaeognath evolution	pg 328-332
5.6.2 Eggshell palaeogenomics and future directions	pg 332-334
5.7 CONCLUSION	pg 334
5.8 REFERENCES	pg 334-343
5.9 SUPPLEMENTARY INFORMATION	pg 344-417
S5.9.1 Dating and morphological identification of eggshell	pg 344
S5.9.2 aDNA extraction	pg 344-348
S5.9.3 Screen samples via amplicon sequencing	pg 348-351
S5.9.4 Quality control	pg 352-353
S5.9.5 Shotgun sequencing	pg 353-355
S5.9.6 Targeted enrichment through hybridisation capture	pg 356-364
S5.9.7 Mitochondrial genome assembly	pg 365-368
S5.9.8 Authenticity	pg 369-370
S5.9.9 Identification of nuclear sequences	pg 370-373
S5.9.10 Alignments and partitioning	pg 373-383

S5.9.11 Phylogenetic analysis	pg 383-399
S5.9.12 Molecular dating	pg 400-412
S5.9.13 Supplementary references	pg 413-417
5.10 EPILOGUE	pg 418

**CHAPTER 6 | INSIGHTS FROM EGGSHELL: ANCIENT DNA FROM FOSSIL EGGSHELL MORPHOTYPES REVISES PHYLOGEOGRAPHY AND SYSTEMATICS OF EXTINCT MADAGASCAN ELEPHANT BIRDS** pg 419-501

6.1 PROLOGUE	pg 420-423
6.1.1 Acknowledgements	pg 421-422
6.1.2 Author contributions	pg 422
6.1.3 Author declarations	pg 422-423
6.2 ABSTRACT	pg 425
6.3 INTRODUCTION	pg 426-428
6.4 MATERIALS AND METHODS	pg 428-435
6.4.1 Eggshell specimen collection	pg 428
6.4.2 Characterisation of thickness morphotypes	pg 428-430
6.4.3 Radiocarbon dating	pg 431
6.4.4 Amino acid racemisation	pg 431
6.4.5 aDNA extraction from eggshell	pg 431-432
6.4.6 Preparation of shotgun sequencing library	pg 432
6.4.7 Hybridisation enrichment of mitochondrial DNA	pg 432
6.4.8 High-throughput DNA sequencing	pg 432
6.4.9 Quality control and filtering	pg 433
6.4.10 Reconstruction of mitochondrial genomes	pg 433
6.4.11 Phylogeny reconstructions and molecular dating	pg 433-434
6.4.12 Assessment of genetic variation in barcoding gene <i>COI</i>	pg 434-435
6.4.13 Characterisation of eggshell microstructure	pg 435
6.4.14 Stable isotope analysis	pg 435
6.5 RESULTS AND DISCUSSION	pg 436-452
6.5.1 Radiocarbon dating	pg 436
6.5.2 Eggshell morphotypes	pg 436-437

6.5.3 Mitochondrial genomes	pg 437-438
6.5.4 Elephant bird phylogeography and taxonomy	pg 438-449
6.5.5 Elephant bird diversification	pg 449-452
6.6 CONCLUSION	pg 452-453
6.7 REFERENCES	pg 453-460
6.8 SUPPLEMENTARY INFORMATION	pg 461-501
S6.8.1 Eggshell specimen collection	pg 461
S6.8.2 Amino acid racemisation	pg 461
S6.8.3 aDNA extraction	pg 462-465
S6.8.4 Preparation of shotgun sequencing library	pg 465-467
S6.8.5 Hybridisation enrichment of mitochondrial DNA	pg 467-469
S6.8.6 Quantitation of the captured libraries and pooling	pg 469
S6.8.7 Size selection and purification of pooled sequencing library	pg 469-470
S6.8.8 Final quantitation and sequencing	pg 470
S6.8.9 Quality control and filtering	pg 470-472
S6.8.10 Reconstruction of mitochondrial genomes	pg 473
S6.8.11 Gene annotation and partitioning	pg 473-477
S6.8.12 RCV analysis	pg 477-479
S6.8.13 Stemminess	pg 480-485
S6.8.14 Model test analysis	pg 485
S6.8.15 Maximum likelihood phylogeny generation	pg 485-486
S6.8.16 Bayesian phylogeny generation	pg 487
S6.8.17 Molecular dating	pg 487-488
S6.8.18 Assessment of genetic variation in barcoding genes	pg 488-490
S6.8.19 Characterisation of eggshell microstructure by micro-CT	pg 490-494
S6.8.20 Stable isotope analysis	pg 495
S6.8.21 Supplementary references	pg 495-499
6.9 EPILOGUE	pg 500-501

**CHAPTER 7 | GENERAL DISCUSSION**

**pg 502-583**

7.1 GENERAL DISCUSSION	pg 503-521
------------------------	------------

7.1.1 Filling the gap_ _ _ _ _	pg 503-506
7.1.2 A grain of salt_ _ _ _ _	pg 506-508
7.1.3 The aDNA sphere_ _ _ _ _	pg 508-509
7.1.4 What is it good for?_ _ _ _ _	pg 509-510
7.1.5 What's next?_ _ _ _ _	pg 510-516
7.1.6 The future of the past_ _ _ _ _	pg 517-521
7.2 CONCLUSIONS_ _ _ _ _	pg 521-522
7.3 REFERENCES_ _ _ _ _	pg 522-540
7.4 SUPPLEMENTARY INFORMATION_ _ _ _ _	pg 541-583
S7.6.1 Supplementary references_ _ _ _ _	pg 560-583

<b>APPENDICES</b>	<b>pg xxxi-xlv</b>
-------------------	--------------------

APPENDIX I RESEARCH OUTPUT_ _ _ _ _	pg xxxi-xxxiv
Appendix I.I Published manuscript facsimilies_ _ _ _ _	pg xxxii
Appendix I.II Conference presentations and awards_ _ _ _ _	pg xxxiii
Appendix I.III Media engagement_ _ _ _ _	pg xxxiii-xxxiv
APPENDIX II SIGNED COAUTHOR PERMISSIONS_ _ _ _ _	pg xxxv-xlv

— INDEX of FIGURES —

<b>FIGURE 1.1</b>	The proportion of Web of Knowledge articles found using the search term ‘ancient DNA OR aDNA’ from each decade since the 1950s_ _ _ _ _ pg 4
<b>FIGURE 1.2</b>	Various types of DNA damage found to occur in aDNA_ _ _ _ _ pg 8
<b>FIGURE 1.3</b>	The NGS workflow_ _ _ _ _ pg 15
<b>FIGURE 2.1</b>	Nested experimental design with replication at three stages_ _ _ _ _ pg 101
<b>FIGURE 2.2</b>	Venn diagram showing the families detected morphologically (blue) and molecularly (yellow)_ _ _ _ _ pg 108
<b>FIGURE 2.3</b>	Rarefaction curves showing how many families would have been detected if fewer bones had been sampled morphologically for both FF (yellow) and BC (grey)_ _ _ _ _ pg 111
<b>FIGURE 2.4</b>	A comparison of the strengths and limitations of both morphological and molecular taxonomic identification of fossils)_ _ _ _ _ pg 114
<b>FIGURE S2.8.1</b>	Location and site information_ _ _ _ _ pg 126
<b>FIGURE S2.8.2</b>	A subsample of 50 bones from the bulk bone sample of morphologically unidentifiable bones from Bat Cave_ _ _ _ _ pg 128
<b>FIGURE S2.8.3</b>	Non-metric multidimensional scaling (NMDS) plot

	showing the main contributions to variance in family richness_ _ _ _ _	pg 141
<b>FIGURE S2.8.4</b>	A comparison of average $C_T$ -values from qPCRs of neat (undiluted) DNA extracted from bleach treated versus non-bleach treated bones_ _ _ _ _	pg 167
<b>FIGURE S2.8.5</b>	Fragment length distribution of 50,000 subsampled reads from sequenced shotgun libraries built on aDNA extracts_ _ _ _ _	pg 169
<b>FIGURE 3.1</b>	Maps of the Naracoorte Caves World Heritage Area in southeast South Australia and Robertson Cave Entrance Chamber_ _ _ _ _	pg 190
<b>FIGURE 3.2</b>	The stratigraphy and dating for Layers 1-6 of the excavation_ _ _ _ _	pg 191
<b>FIGURE 3.3</b>	Pie charts showing the percentage of endogenous and exogenous shotgun sequence reads_ _ _ _ _	pg 200
<b>FIGURE 3.4</b>	Fragment length distributions of unique marsupial shotgun reads, the proportion of endogenous DNA expected to be remaining given a certain time, and the probability of survival of 50,000 years after cell death for various fragment lengths_ _ _ _ _	pg 202
<b>FIGURE S3.8.1</b>	A diagram depicting the workflow of the methods (sequencing and quantitation)_ _ _ _ _	pg 225
<b>FIGURE S3.8.2</b>	Base composition of 15 bases from the 5' end and the 3' end of unique marsupial shotgun reads (percentage of bases) for Layers 1, 3, and 6_ _ _ _ _	pg 239

<b>FIGURE S3.8.3</b>	Cycle threshold ( $C_T$ ) values plotted against log concentration for each layer_ _ _ _ _	pg 244
<b>FIGURE S3.8.4</b>	qPCR amplification plots_ _ _ _ _	pg 245
<b>FIGURE 4.1</b>	Map of Madagascar_ _ _ _ _	pg 266
<b>FIGURE S4.8.1</b>	An alignment of 80 local fish taxa showing the conserved primer-binding regions flanking the <i>12S</i> metabarcoding region_ _ _ _ _	pg 284-286
<b>FIGURE S4.8.2</b>	An alignment of 80 local fish taxa showing the variable metabarcoding region (labeled 'INSERT' in Figure S4.8.1)_ _ _ _ _	pg 287-289
<b>FIGURE 5.1</b>	Details of the eggshell specimen_ _ _ _ _	pg 318
<b>FIGURE 5.2</b>	Mitochondrial, nuclear, and combined palaeognath phylogenetic trees_ _ _ _ _	pg 322
<b>FIGURE 5.3</b>	Combined mitochondrial and nuclear chronogram_ _ _ _ _	pg 325-326
<b>FIGURE S5.9.1</b>	Summary of the steps taken to isolate and reconstruct the mitochondrial genome and retrieve nuclear loci from <i>Aepyornis</i> eggshell_ _ _ _ _	pg 345
<b>FIGURE S5.9.2</b>	Nucleotide misincorporation plots_ _ _ _ _	pg 369
<b>FIGURE S5.9.3</b>	GC content of nuclear third codon position (n3, hb3) across taxa_ _ _ _ _	pg 387
<b>FIGURE S5.9.4</b>	Ratio of C:T frequency of mitochondrial third codon	

	position (m3) across taxa _ _ _ _ _	pg 389
<b>FIGURE S5.9.5  </b>	Chronograms for alternative datasets _ _ _ _ _	pg 403
<b>FIGURE 6.1  </b>	Map of Madagascar depicting the geographic location of eggshell samples collected and analysed _ _ _ _ _	pg 429-430
<b>FIGURE 6.2  </b>	Molecular dated phylogenetic tree _ _ _ _ _	pg 439
<b>FIGURE 6.3  </b>	Eggshell microstructure _ _ _ _ _	pg 446
<b>FIGURE 6.4  </b>	A comparison of mean $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ isotope content of <i>Aepyornis</i> and <i>Mullerornis</i> eggshells from South/Southwest Madagascar _ _ _ _ _	pg 448
<b>FIGURE S6.8.1  </b>	Damage profile _ _ _ _ _	pg 474
<b>FIGURE S6.8.2  </b>	An example of the 20.07 mm <sup>2</sup> (4.45 x 4.45 mm) region of interest (ROI) _ _ _ _ _	pg 493



<b>TABLE S2.8.5  </b>	A comparison of the family and species-level taxa detected in the fossil record (morphological) and by the “bulk bone” method for both FF and BC, respectively_ _ _ _ _ pg 150-155
<b>TABLE S2.8.6  </b>	One-way ANOVA testing whether there is a statistically significant effect of bleach treatment on C <sub>T</sub> -values from a qPCR_ _ _ _ _ pg 165
<b>TABLE 3.1  </b>	Replicate and average qPCR cycle-threshold (C <sub>T</sub> ) for both 12SAO and 12SMarsMini primer sets for Layers 1-6_ _ _ _ _ pg 198
<b>TABLE 3.2  </b>	Empirical decay constant ( $k_{av}$ per site per year) for Layers 1, 3, and 6_ _ _ _ _ pg 205-206
<b>TABLE 3.3  </b>	Summary of the taxa identified from RCEC for Layer 1-6_ _ _ _ _ pg 209-211
<b>TABLE S3.8.1  </b>	Radiocarbon dates for Layers 1 through 6 (square 1) from Robertson Cave Entrance Chamber_ _ _ _ _ pg 224
<b>TABLE S3.8.2  </b>	Primers used for amplicon sequencing_ _ _ _ _ pg 229
<b>TABLE S3.8.3  </b>	Primers used for quantify the final sequence library by qPCR_ _ _ _ _ pg 231
<b>TABLE S3.8.4  </b>	Ligation adapters and sequencing primers used for shotgun sequencing (modifications from Gansauge and Meyer, 2013)_ _ _ _ _ pg 233
<b>TABLE S3.8.5  </b>	Amplicon sequencing read abundance obtained for each extract (12SAO and 12SMarsMini) after quality

filtering, chimera filtering, and abundance filtering\_ \_ \_ \_ \_ pg 235-236

**TABLE S3.8.6** | Base composition of 15 bases from the 5' end and the 3' end of unique marsupial shotgun reads (percentage of bases) for Layers 1, 3, and 6\_ \_ \_ \_ \_ pg 237-238

**TABLE S3.8.7** | Total DNA concentration and purity (260/280) as determined by a Nanodrop spectrophotometer\_ \_ \_ \_ \_ pg 241

**TABLE S3.8.8** | qPCR cycle-threshold ( $C_T$ ) values for each replicate of each dilution per layer, for both 12SAO and 12SMarsMini primer sets\_ \_ \_ \_ \_ pg 242

**TABLE S3.8.9** | The percent amplification efficiency of the qPCR reactions for DNA extracts from Layers 1-6 amplified with 12SAO and 12SMarsMini primer sets\_ \_ \_ \_ \_ .pg 243

**TABLE S3.8.10** | Full data table showing the taxa detected for each sub-sample and each extract replicate within sub-sample (as well as extraction controls) for both 12SAO and 12SMarsMini\_ \_ \_ \_ \_ pg 246

**TABLE S3.8.11** | Identity matrices of sequences amplified with 12SMarsMini in three example samples\_ \_ \_ \_ \_ pg 248-249

**TABLE S3.8.12** | Taxonomic assignments of 1000 sequences generated from mutating one sequence of each genus at a rate of 1%, 2% and 3%\_ \_ \_ \_ \_ pg 250-251

**TABLE 4.1** | Taxa identified through bulk bone metabarcoding of aDNA\_ \_ \_ \_ \_ pg 269-271

**TABLE S4.8.1** | Identity matrices of sequences amplified with the short

	<i>I2S</i> primer set _ _ _ _ _	pg 292-293
<b>TABLE S4.8.2  </b>	Taxonomic assignments of 1000 sequences generated from mutating one sequence of each taxon at a rate of 1%, 2% and 3%_ _ _ _ _	pg 294-297
<b>TABLE 5.1  </b>	Summary of palaeognath divergence dates_ _ _ _ _	pg 324
<b>TABLE S5.9.1  </b>	Metabarcoding primers used, and ion torrent adapter sequences that flank the metabarcoding primers for fusion tagging_ _ _ _ _	pg 350
<b>TABLE S5.9.2  </b>	Primers used for quantify the final sequence library by qPCR_ _ _ _ _	pg 351
<b>TABLE S5.9.3  </b>	Ligation adapters, fusion primers, and sequencing primers used for shotgun sequencing on both the Ion Torrent and MiSeq NGS platforms_ _ _ _ _	pg 354
<b>TABLE S5.9.4  </b>	GenBank accession of the sequences that comprised the avian baits_ _ _ _ _	pg 356-359
<b>TABLE S5.9.5  </b>	Custom blocking primers (for sequencing on both IonTorrent and MiSeq platforms) used in the enrichment protocols_ _ _ _ _	pg 361
<b>TABLE S5.9.6  </b>	The number of reads above 30 bp obtained after trimming, quality filtering, and dereplication for various libraries sequenced on both the Ion Torrent or MiSeq platforms_ _ _ _ _	pg 364
<b>TABLE S5.9.7  </b>	The number of unique reads above 30 bp for shotgun	

and enriched libraries that were endogenous (avian) and mitochondrial (mapped to a mitochondrial genome) \_ \_ \_ \_ \_ pg 367

<b>TABLE S5.9.8</b>	Statistics of the read distributions of several datasets including the mean fragment length and percentage of high quality bases _ _ _ _ _	pg 368
<b>TABLE S5.9.9</b>	Taxon sampling _ _ _ _ _	pg 371
<b>TABLE S5.9.10</b>	Data matrix _ _ _ _ _	pg 374-382
<b>TABLE S5.9.11</b>	RCV calculations _ _ _ _ _	pg 384
<b>TABLE S5.9.12</b>	Nuclear third codon position (n3, hb3) base compositions across taxa _ _ _ _ _	pg 386
<b>TABLE S5.9.13</b>	Mitochondrial third codon position (m3) base compositions _ _ _ _ _	pg 388
<b>TABLE S5.9.14</b>	Uncorrected stemminess values _ _ _ _ _	pg 391
<b>TABLE S5.9.15</b>	Corrected stemminess values _ _ _ _ _	pg 392
<b>TABLE S5.9.16</b>	ModelTest results _ _ _ _ _	pg 394
<b>TABLE S5.9.17</b>	Phylograms from both maximum likelihood and Bayesian analyses of various non-site-stripped datasets _ _ _	pg 396-399
<b>TABLE S5.9.18</b>	Fossil calibrations used for molecular dating _ _ _ _ _	pg 400-401
<b>TABLE S5.9.19</b>	Divergence times (mya) of each major palaeognath clade for multiple datasets, input phylogenetic trees,	

	and rates _ _ _ _ _	pg 404-412
<b>TABLE 6.1</b>	Current accepted species of elephant bird_ _ _ _ _	pg 427
<b>TABLE 6.2</b>	Estimates of average evolutionary divergence over sequence pairs within and between groups across a 596 bp barcoding region of <i>COI</i> for elephant birds and other ratites _ _ _ _ _	pg 440
<b>TABLE S6.8.1</b>	Eggshell specimens prioritised for aDNA extraction_ _ _	pg 463
<b>TABLE S6.8.2</b>	Single-stranded library build primers_ _ _ _ _	pg 466
<b>TABLE S6.8.3</b>	Custom blocking primers_ _ _ _ _	pg 468
<b>TABLE S6.8.4</b>	Custom i5 indexing primer and read 1 sequencing primers used to sequence libraries on the NextSeq platform_ _ _ _ _	pg 471
<b>TABLE S6.8.5</b>	Summary table of the reads lost during quality control, the number of reads mapped and their statistics, and the average coverage of the mitochondrial genome across the number of base-pairs retrieved for each sample _ _ _ _ _	pg 472
<b>TABLE S6.8.6</b>	GenBank accessions of the eight outgroup ratite taxa used for whole mitochondrial genome phylogenetic analyses_ _ _ _ _	pg 476
<b>TABLE S6.8.7</b>	RCV calculations and Chi-Squared test estimates_ _ _ _ _	pg 479
<b>TABLE S6.8.8</b>	Uncorrected stemminess estimates_ _ _ _ _	pg 483

<b>TABLE S6.8.9</b>	Average corrected stemminess estimates_ _ _ _ _	pg 484
<b>TABLE S6.8.10</b>	ModelTest results_ _ _ _ _	pg 486
<b>TABLE S6.8.11</b>	Calibrations used for molecular dating_ _ _ _ _	pg 489
<b>TABLE S6.8.12</b>	Eggshell specimens imaged with micro-CT_ _ _ _ _	pg 491
<b>TABLE S6.8.13</b>	Features of eggshell microstructure_ _ _ _ _	pg 494
<b>TABLE 7.1</b>	Summary of sequencing technologies_ _ _ _ _	pg 494-495
<b>TABLE S7.4.1</b>	Summary of the types of aDNA studies undertaken between 2013 and 2016_ _ _ _ _	pg 522-540

## —ABBREVIATIONS —

$\chi^2$	Chi-Squared
%	Percent
°	Degrees
°C	Degrees Celcius
A	Adenosine
aDNA	Ancient DNA
AMS	Accelerator Mass Spectrometry
ANOVA	Analysis of Variance
APA	Australian Post-graduate Award
ARC	Australian Research Council
aRNA	ancient RNA
AUS	Australia
Av.	Average
BBM	Bulk Bone Metabarcoding
BC	Bat Cave
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
BP	Before Present
BSA	Bovine Serum Albumin
C	Carbon
c. or ca.	Circa
cf.	<i>Conferre</i>
cal	Calendar
cat	Catalogue
chDNA	Chloroplast DNA
CI	Confidence Interval
cm	Centimetre
COI	Cytochrome Oxidase Subunit 1
CT	Cycle Threshold
Cytb	Cytochrome B

ddPCR	Digital Droplet PCR
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide
DTT	Dithiothreitol
dTTP	Thymidine triphosphate
DTU	Distance-based Taxonomic Unit
dUTP	Deoxyuridine Triphosphate
E	East
E-O	Eocene-Oligocene
e.g.	Example
EDTA	Ethylenediaminetetraacetic acid
emPCR	Emulsion PCR
et al.	<i>et alia</i> ; and others
FF	Finsch's Folly
Fwd	Forward
G	Guanine
g	Grams
Gbp or Gb	Gigabase-pairs
GC/MS	Gas Chromatography / Mass spectrometry
GSP	Gene specific primer
HCl	Hydrochloric Acid
HPD	Higher Posterior Density
HPLC	High Performance Liquid Chromatography
hr	hour
HTS	High-throughput Sequencing
i.e.	<i>id est</i> ; that is
ID	Identification
IDT	Integrated DNA Technologies
IT	Ion Torrent
IUCN	International Union for Conservation of Nature
K	Cretaceous
kg	Kilograms

KI	Kangaroo Island
kyr	Thousands of Years / Thousands of years ago
l	Litre
LCA	Last Common Ancestor
LGM	Last Glacial Maximum
LMMA	Locally Managed Marine Area
m	Metre
M	Molar
Ma or mya	Million Years / Millions of years ago
MANOVA	Multivariate Analysis of Variance
MAP	Morombe Archaeological Project
Max	Maximum
MEGAN	Metagenome Analyser
mg	Milligrams
MgCl <sub>2</sub>	Magnesium Chloride
Min	minute
Min	Minimum
ml	Millilitre
mM	Millimolar
mm	Millimetre
MNI	Minimum Number of Individuals
MS	Mean Squares
mtDNA	Mitochondrial DNA
MWCO	Molecular Weight Cut-Off
myr	Million Years
N	North
n.b.	<i>nota bene</i> ; note well
NA	Not Applicable
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
NCWHA	Naracoorte Caves World Hertiage Area
ng	Nanograms
NGS	Next-generation Sequencing

NMDS	Non-metric Multidimensional Scaling
NPMANOVA	Non-parametric MANOVA
nr	Non-redundant
nt	Nucleotide
nuDNA	Nuclear DNA
NZ	New Zealand
OTU	Operational Taxonomic Unit
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PEC	Primer extension capture
PEP	Polymerase extension profiling
PERMANOVA	Permutational MANOVA
Pg	Palaeogene
pg	Page
pM	Picomolar
PTB	N-phenacylthiazolium bromide
Q	Quality
q.v.	<i>quod vide</i> ; for which, see elsewhere
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative PCR
RCEC	Robertson Cave Entrance Chamber
RCV	Relative Composition Variability
Rev	Reverse
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
rRNA	Ribosomal RNA
RT	Room Temperature
s	second
S	South
SA	South Australia
SD or St Dev	Standard Deviation
SDS	Sodium dodecyl sulfate
sedaDNA	Sedimentary Ancient DNA

SI	Supplementary Information
SMRT	Single Molecule Real Time
SNP	Single nucleotide polymorphism
sp.	Species
SPEX	Single primer extension
spp.	Several species
SS	Sums of Squares
SSC	Saline Sodium Citrate
SSPE	Saline Sodium Phosphate EDTA
Std	Standard
T	Thymine
TAE	Tris-acetate EDTA
<i>Taq</i>	<i>Thermus aquaticus</i>
TBA	To be advised
TGS	Third-generation Sequencing
TRACE	Trace Advanced Ultra Clean Environment
TrEnD	Trace and Environmental DNA
tRNA	Transfer RNA
U	Unit
UDG	Uracil Deglycosylase
UNG	Uracil DNA glycosylase; UDG
UV	Ultraviolet
v	Version
V	Volts
Vol	Volume
W	West
WA	Western Australia
X	Times
yr	Years
µg	Micrograms
µl	Microlitre
µm	Micrometre
µM	Micromolar

*The past is the past or so they say—  
'No use to the present or to future days!'—  
But we have the ability to understand past woes,  
and to ignore such evidence could lead to future lows.  
For mankind are blind and have lost so much,  
that we now wish to conserve, and to biodiversity we clutch.  
Thus we need to recognise, understand, and explore  
past biodiversity, phylogenies—and more!  
To this end, dear reader, this thesis is presented to you  
as a small hope for our future (and because it's due!)*

- Tina E Berry

## — PREFACE —

Curtin University,  
Kent St, Bentley, WA

Monday, August 1, 2016

Dear reader,

The following doctoral dissertation is a synthesis of the work undertaken by myself at both Murdoch University and Curtin University over the course of three and a half years between January 2013 and September 2016. Chapter 1 provides an in-depth review of the history and challenges facing the field of ancient DNA up until 2013, and describes the overall aims, scope, and significance of the subsequent work within this context. This thesis includes three articles published in peer-reviewed journals (Chapters 2, 3, and 4), one manuscript in review (Chapter 5), and one manuscript in preparation for submission within the next six months (Chapter 6). The format of this thesis is a typescript of the aforementioned articles, each of which are flanked by a prologue and epilogue that aims to connect each chapter to the next in order to form a cohesive body of work. The body of each chapter is organised into the following sections: Abstract, Introduction, Materials and Methods, Results and Discussion, and Conclusion. Facsimiles of all published manuscripts can be found in the appendices (Appendix I.I). Work that may be of interest but less pertinent to the understanding of the main text of the manuscript has been included at the end of each chapter in a section entitled ‘Supplementary Information’. Finally, Chapter 7 summarises the

major findings of each chapter and discusses how those findings have contributed to the advancement of the field. Future directions are also addressed.

Each chapter is the outcome of collaborations with multiple national and international partners spanning a variety of disciplines (including palaeontology, archaeology, phylogenetics and genomics) across several internationally recognised institutions (including Oxford University and Blue Ventures Conservation (UK); Yale University, the Smithsonian National Museum of Natural History, and the University of Colorado (USA); Canterbury Museum and the University of Otago (NZ); Flinders University, The University of Adelaide, and Queensland University of Technology (AUS); Natural History Museum of Denmark and the University of Copenhagen (DK); and the University of South Africa (SA). The contribution of all co-authors and the role of others (including funding bodies) are acknowledged in the prologue to each chapter. To reiterate that the joint efforts of a team of people made this research possible, the pronoun ‘we’ is often referenced throughout this text.

Although this thesis appears long, it is important to appreciate that a considerable amount of additional experimentation has gone into the generation of each chapter than has been published, including optimisation of new methods, failed samples (in the words of HG Wells in *Aepyornis Island*, “it took me some time to learn how unforgiving and cantankerous an extinct bird can be!”), and abandoned tangential avenues of inquiry. With that, I can assure you that no amount of pain suffered from reading this thesis can compare with the amount that went into writing it.

Happy reading!

Sincerely,

A stylized signature of Alicia Greal, featuring the name in a large, elegant, serif font with decorative flourishes and a cursive script underneath.

Alicia Greal,  
BSc. (Hons.), PhD.

## — ACKNOWLEDGEMENTS —

*“If I have seen further it is by standing on the shoulders of giants”*

- Sir Isaac Newton

My time as a PhD student has been a challenging but rewarding experience, and I could not have completed it without the support, advice, and friendship of my peers and colleagues in the Trace and Environmental DNA group at Curtin University.

In particular, I would like to thank my close friends and fellow students Daithí Murray, Tina Berry, and Dr. Megan Coghlan for their relentless encouragement and unwavering support over the past three and a half years. Not only did they make me feel welcome and included as a new student, but they also became the family that missed so much since moving to Perth to undertake my graduate studies. Our morning coffee runs in which we were able to vent our frustrations and troubleshoot our failed experiments were solely responsible for their eventual success, I am sure. I would also like to thank the TrEnD research assistants, Matthew Power and James

Taylor, without whom life would have deteriorated into a mess of paperwork, primers, and problems. I will be eternally grateful that I was able to work alongside such an intelligent, talented, hard-working, and generous group of people.

To my supervisors Prof. Michael Bunce and Dr. James Haile—thank you so much for your continued guidance throughout my candidature. I have been given countless amazing opportunities to learn new skills, hone my existing abilities, and foster new and lasting collaborations. You have provided me with a solid foundation upon which I look forward to building my career, and I am lucky to have had the chance to pursue my passion for ancient DNA in a project that was engaging and stimulating everyday.

The input of my collaborators and co-authors has been pivotal in helping me reach this point. The enthusiastic, understanding, relaxed, and professional attitudes of Dr. Kristina Douglass and Dr. Amy Macken made working with them effortless and refreshing. A/Prof. Matthew Phillips provided invaluable mentorship above and beyond his role as co-author. I would like to thank everyone who assisted with the collection of samples, including the Morombe Archaeological Project Team who were so accommodating during the 2014 field season in Madagascar.

Above all, my sincerest appreciation goes to my partner (Alexander ‘Panda’ Gofton), my parents (Liz and Bernie Grealy), my sisters (Becci, Tami, and Jaci Grealy) and my friends (Bec Joyce, Lois Shedd, Declan McClure, and more) for putting up with my stress, sneaking in to an international conference to watch me present, pretending to read my “boring” papers, sharing my PhD highs and lows, coming to visit from across the country, and even helping me with the occasional menial laboratory task or two. This would not have been possible without you. Thank you.

The Australian Government’s Australian Postgraduate Award and Curtin University financially supported me in this research.

## — CHAPTER 1 —

---

### INTRODUCTION

---

*Isn't it a neat enough trick to show that DNA can indeed be obtained from ancient specimens? Alas, if ancient DNA is to become a legitimate field of scientific inquiry, then the answer must be no.*

- Stoneking (1995)

## 1.1 USING ANCIENT DNA TO STUDY EVOLUTION: PAST, PRESENT, *and* FUTURE PERSPECTIVES (A LITERATURE REVIEW)

---

### 1.1.1 INTRODUCING ANCIENT DNA:

#### WHAT IS IT AND WHY IS IT IMPORTANT?

**S**INCE their discovery, the study of fossils has been, and remains, the predominant method of examining past life. Specifically, comparative morphology and morphometrics have been the main methods employed to distinguish prehistoric and extinct species, determine their relatedness to one another and extant species, and reconstruct their interactions with each other and the environment (Gatesy and Dial 1996; Rodríguez-Aranda and Calvo 1998; Mardis 2008a; Mardis 2008b; Zhang et al. 2008). However, these methods rely on the assumption that morphological features can distinguish taxa. This can mean that: (i) the identification of some fossil species is impossible if specimens are incomplete or fragmented, which is often the case; and (ii) that the evolutionary relationships inferred can be misguided because often species appear to be different when they are not (Ali 1969; Andre et al. 2013), or conversely, appear to be the same when they are not (Besansky et al. 1994; Yen and Yang 2009). This is largely because divergence can be driven by the evolution of phenotypes that do not tend to fossilise (for example, soft tissue, pigments, and enzymes; Berdan and Fuller 2012; Nydam et al. 2013), or by the evolution of phenotypes other than morphology, such as behavior (Besansky et al. 1994). Similarly, genetics can be used to determine the evolutionary relationships of extant species (e.g., the horse; Vilà et al. 2001), particularly in cases where morphology cannot (for example, where cryptic species exist). However, the use of modern DNA to reconstruct the evolutionary history of extant organisms, and to infer the processes that shaped that history, also has limitations: modern DNA cannot capture the genetic diversity that existed in the past (Willerslev and Cooper 2005). As such, DNA directly derived from ancestral or extinct taxa can be particularly valuable for elucidating the relationships among extant taxa (Cano 1996).

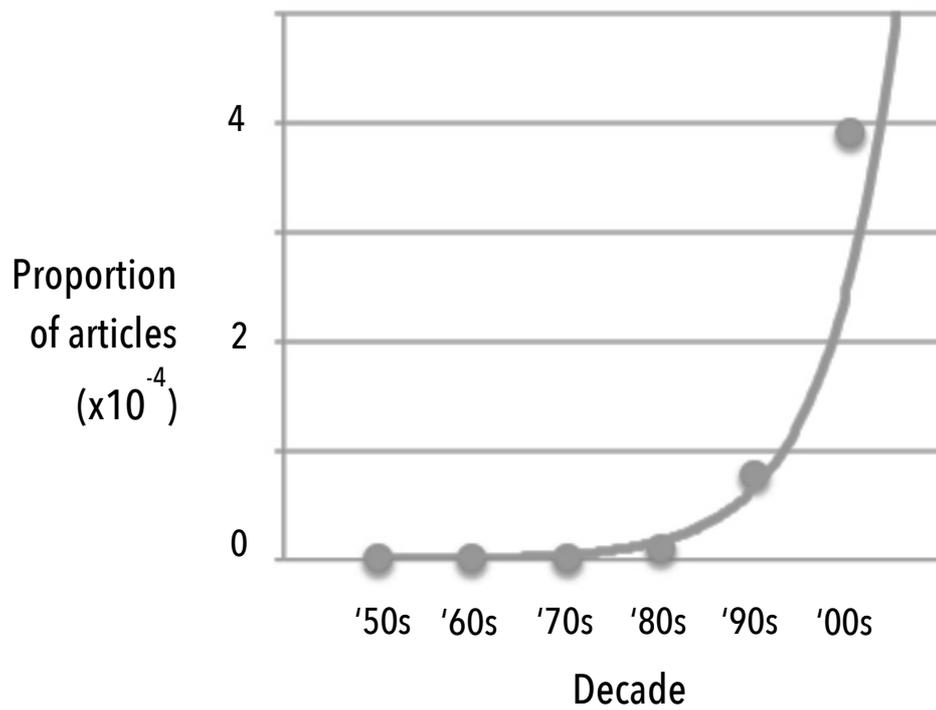
Some of the limitations of both these methods for studying evolutionary history can be addressed using aDNA. Ancient DNA (aDNA) is DNA that has been isolated

from ancient or degraded substrates such as fossils, sediment, or museum specimens. aDNA can be used for the taxonomic identification of morphologically unidentifiable fossils, and can be used to more accurately assess evolutionary relationships between extinct organisms by comparing the differences that exist at the molecular level (Cano 1996). In addition, aDNA can help resolve the evolutionary history of extant organisms by providing direct evidence of the past genetic diversity of species and populations; this can be used to investigate the evolutionary process that formed modern species (such as natural selection, extinction, mutation, gene flow, and genetic drift), as well as to assess species biodiversity across time and space, and provide time-of-divergence estimates through molecular clock dating. Such information gleaned from the study of aDNA has even had valuable implications for the conservation of extant species (e.g., Cooper et al. 1996). In addition, aDNA methods are often employed by other disciplines that deal with highly degraded and mixed DNA, such as metagenomics and forensics. In combination with traditional paleontological and phylogenetic methods, the study of aDNA or 'palaeogenetics' can reveal a great deal more about the processes that shaped the evolution of life than can be gained from the use of traditional methods alone. Since its inception over 30 years ago, the use of aDNA alongside morphology and modern DNA has exponentially grown as genetic methods for isolating and analysing aDNA improve (Figure 1.1). Nevertheless, ancient DNA remains a niche field, exhibiting its own limitations and biases that need to be appreciated and considered when conducting aDNA analysis.

### **1.1.2 ANCIENT (DNA) HISTORY 101:**

#### **HOW HAS ANCIENT DNA TRADITIONALLY BEEN USED? 1984-2000**

In the decade following the first successful extraction of DNA from an ancient specimen (Higuchi et al. 1984), aDNA was recovered from a variety of extinct animals and plants, including moa (Diornithiformes) (Cooper et al. 1992), sabretoothed cat (*Smilodon fatalis*; Janczewski et al. 1992), quagga (*Equus quagga*; Higuchi et al. 1984), giant ground sloth (*Myiodon darwini*; Pääbo 1989), and



---

**FIGURE 1.1** | THE PROPORTION OF WEB OF KNOWLEDGE ARTICLES FOUND USING THE SEARCH TERM 'ANCIENT DNA OR ADNA' FROM EACH DECADE SINCE THE 1950s.

thylacine (*Thylacinus cynocephalus*; Thomas et al. 1989). The DNA obtained from these organisms was used to assess their evolutionary relationships to extant organisms, and to estimate when they diverged from their ancestors. For instance, Higuchi et al. (1984) cloned the first two mitochondrial DNA (mtDNA) sequences from the extinct quagga, and used them to determine the phylogenetic placement of the quagga by comparing the similarity of the sequences to the orthologous sequences from three extant mammals (zebra, cow, and human). Higuchi et al. (1984) also used measures of sequence divergence to estimate the timing of the split between quagga and zebra lineages to approximately 3-4 million years ago (Ma). Similarly, Thomas et al. (1989) compared the mtDNA sequences of the extinct thylacine to extant Australian and South American marsupials, and found that the thylacine shared a common ancestor with Australian carnivorous marsupials about 10-20 Ma. This evidence contradicted morphological analyses that suggested thylacines directly diverged from South American carnivorous marsupials. The authors suggest that this discrepancy could be explained by the evolution of convergent adaptations that arose in response to similar selective pressures (Thomas et al. 1989). These examples demonstrate how, even within the first few years of study, aDNA shed a new light on the evolutionary history of both extinct and extant organisms.

The use of aDNA to inform phylogenies dominated the field well into the 1990s (Cooper 1994; Höss et al. 1996a; Matisoo-Smith et al. 1997; Ozawa et al. 1997; Sorenson et al. 1999). For instance, Yang et al. (1996) extracted mtDNA from the extinct woolly mammoth (*Mammuthus primigenius*) and American mastodon (*Mammut americanum*) to resolve their placement within the elephant phylogeny, which was contentious (Hagelberg et al. 1994). They found that woolly mammoths are more closely related to the Asian elephant than the African elephant (Yang et al. 1996). However, the mid-1990s also saw a shift towards a more applied approach to aDNA work. For example, Cooper et al. (1996) resolved the identity of sub-fossil bones from Hawaii by comparing their DNA to that of the endangered Laysan duck (*Anas laysanensis*). Today, the Laysan duck only exists on the remote island of Laysan, north of Hawaii; however, aDNA analysis of unidentifiable bones showed that these ducks once colonised the main Hawaiian Islands, but were extirpated upon human colonisation (Cooper et al. 1996). Cooper et al. (1996) used this evidence to

justify the reintroduction of the duck to its past range, demonstrating how aDNA could be used for the practical management of vulnerable species. This experiment was the first to highlight how the study of aDNA could have implications for conservation, but more importantly, how much more potential existed in aDNA than was being exploited at the time.

### **1.1.3 ‘DAMAGE CONTROL’:**

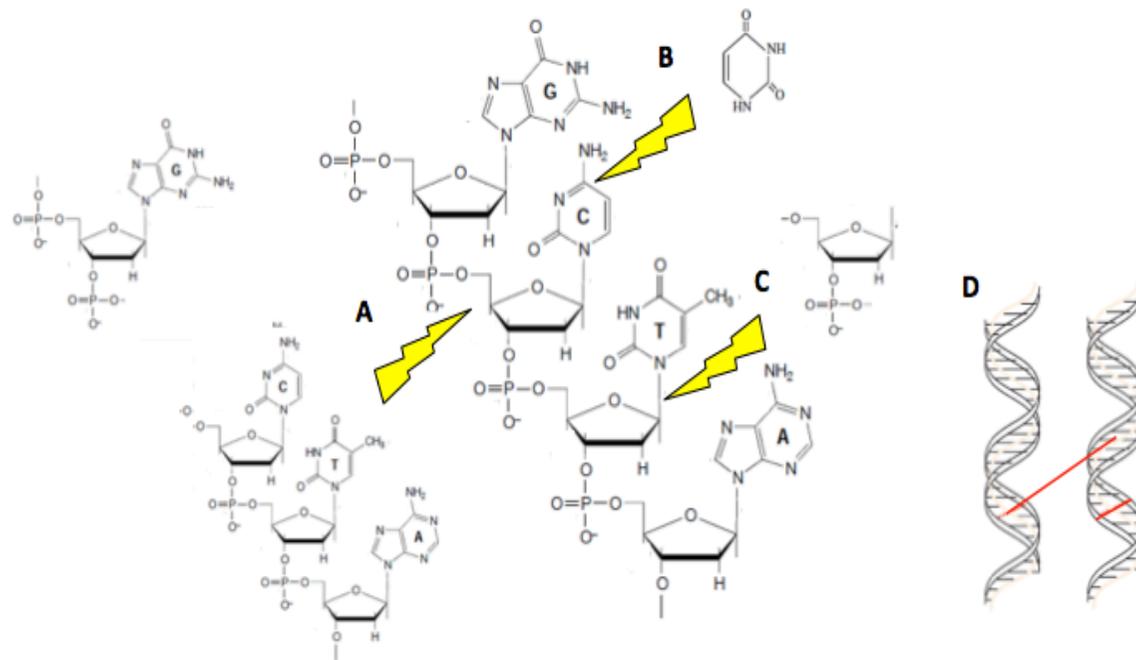
#### **WHAT HAVE THE CHALLENGES BEEN?**

Although many insights into evolution and conservation were gained through the use of aDNA during the 1980s and 1990s, the field was fraught with many challenges. These challenges mostly arose from the inherently damaged, degraded, and contaminated nature of aDNA, as well as from the technical limitations of the time.

Firstly, DNA preservation within ancient specimens that are themselves rare is even rarer. For example, of the 35 specimens of ground sloth analysed by Höss et al. (1996a), only two of those specimens yielded amplifiable DNA (Höss et al. 1996a). This is because in the absence of the repair mechanisms of living cells, over time, DNA is subject to damage and degradation (or ‘post mortem modifications’). Post-mortem modifications may occur by the action of biotic factors, such as enzymes that degrade DNA (e.g., DNases; Darzynkiewicz et al. 1997), and by abiotic factors, such as ionising radiation (e.g., UV exposure), temperature, and pH (Eglinton and Logan 1991; Cano 1996). These factors can cause single-stranded and double-stranded breaks to form, as well as the formation of abasic sites (depurination), blocking or miscoding lesions (deamination), and cross-links (Shapiro and Hofreiter 2012). For instance, DNA hydrolysis and oxidation can result in base modifications, such as the deamination of cytosine to uracil, or adenine to hypoxanthine (Höss et al. 1996b; Hofreiter et al. 2001a; Gilbert et al. 2003). Figure 1.2 shows some examples of post-mortem DNA damage. In addition, DNA spontaneously decomposes (Lindahl 1993), and it has been estimated that under ideal (i.e., frozen) preservation conditions, DNA in bone will degrade to a length of about 700 bp after 10 kyr, and a length of 1 bp after 6.8 Ma (Allentoft 2012). Thus, if aDNA is present at all, it is more often than not highly fragmented, highly damaged, and present in extremely low amounts. This means that only short (50-500 bp) sequences may be amplified (Higuchi et al. 1984;

Pääbo and Wilson 1988; Pääbo et al. 1989), and these could be riddled with damage that will confound signals of evolutionary change (such as true mutations or polymorphisms) (Clark and Whittam 1992; Gilbert et al. 2003; Willerslev and Cooper 2005). Furthermore, if a reference genome for an extinct organism does not exist, it is difficult to distinguish between damage, sequencing errors, and true polymorphism.

Secondly, there is the issue of contamination by both endogenous and exogenous DNA, such as human DNA, bacterial DNA, or DNA from other organisms that may or may not be ancient itself (Soltis and Soltis 1993). These contaminants can arise from handling, from the depositional environment, or from the bacteria that thrived within the live organism, as well as from laboratories or reagents (Kwok and Higuchi 1989; Taylor 1996; Austin et al. 1997b). In one study, 40-65% of fox teeth sampled were contaminated with human DNA resulting from improper handling during sample collection (Wandeler et al. 2003). Contaminating DNA is problematic because it is present in far greater quality and quantity than endogenous ancient DNA, and could completely prevent ('swamp') the detection of aDNA by downstream applications because undamaged modern DNA will be preferentially amplified over damaged aDNA during PCR (Gilbert et al. 2005). Alternatively, the amplification of contaminating DNA can be confused for aDNA, especially if there is not much divergence between ancient and modern sequences (this is particularly true for human DNA). There have been several examples of studies that claimed to have obtained DNA from ancient substrates that are millions of years old, such as 80 Ma bone fragments (Woodward et al. 1994), dinosaur eggs (Zou et al. 1995), 17-20 Ma plant fossil material (Golenberg et al. 1990), and a 120-135 Ma weevil embedded in amber (Cano et al. 1993); these results were later found to be either irreproducible (Handt et al. 1994) or the consequence of contamination (Austin et al. 1997a; Austin et al. 1997b) by human DNA (Zischler et al. 1995) and fungal DNA (Gutiérrez and Marín 1998). Another more insidious example of DNA contamination resulting in false positives was that of human DNA extracted from teeth (later found to be from cave bears) found at a fossil deposit in China (Hofreiter et al. 2001b): in this case, sequencing and independent replication were unable to detect the error. These examples clearly demonstrate how pervasive DNA contamination can be, and how it can lead to the spread of misinformation. Compounds from the environment that



**FIGURE 1.2** | VARIOUS TYPES OF DNA DAMAGE FOUND TO OCCUR IN ADNA. **A** hydrolysis causing single stranded and double stranded breaks, resulting in fragments. **B** oxidation and hydrolysis causing deamination, resulting in a miscoding lesion (here, cytosine is deaminated to form uracil) and sequencing errors. **C** oxidation and hydrolysis causing depurination, resulting in an abasic site. **D** alkylation causing the formation of intra- and inter-molecular crosslinks, resulting in PCR inhibition and no amplification. Adapted from Hofreiter et al. (2001b) and Willerslev and Cooper (2005).

inhibit downstream applications (such as PCR and sequencing) can also contaminate aDNA. For example, the presence of tannins or humic acids (Tsai 1992), as well as other compounds (Wilson 1997), can prevent DNA polymerase from binding to template DNA (Bruce 1992), resulting in no amplification of the target aDNA (Watson and Blackwell 2000). Often this type of inhibition can be overcome by diluting the extract, and thereby diluting the inhibitor (Soltis and Soltis 1993; Wilson 1997), though there are limits to diluting a sample that already contains a low concentration of template molecules. In combination with inhibiting compounds, blocking lesions terminate amplification, and cross-links that can form within and between DNA molecules result in a change in secondary structure that can inhibit strand separation during PCR as well as the binding of DNA polymerases (Pääbo and Wilson 1988). Thus, even if aDNA is present, it may not be amenable to PCR amplification due to a variety of post-mortem processes.

Finally, the molecular techniques required to extract and subsequently analyse such severely damaged, degraded and contaminated DNA had not been developed. Most early studies relied almost entirely on error-prone techniques such as molecular cloning and PCR. In these workflows, short fragments of aDNA are PCR amplified (Mullis and Faloona 1987) using primers complementary to a conserved region of the gene of interest, and the products are cloned into bacterial vectors for sequencing, one-by-one. Sequences are then aligned to determine the consensus sequence (Krings et al. 1997). However, molecular cloning is highly inefficient and can introduce artifacts (Pääbo and Wilson 1988; Pääbo et al. 1989), and—although the development of PCR revolutionized the field because it allows the amplification of a targeted sequence of interest from, theoretically, a single starting copy (Willerslev and Cooper 2005)—PCR can introduce base misincorporations (Pääbo et al. 1989), replicate damage (Soltis and Soltis 1993), and, because of its sensitivity, amplify even low-level contamination (Cano 1996). In addition, the fidelity of some PCR enzymes may be low: for example, *Taq* DNA polymerase can have an error rate of up to 0.25% in a standard PCR (Pääbo and Wilson 1988). Length and GC content biases in sequencing data have also been attributed to the use of certain DNA polymerases (Dabney and Meyer 2012).

Because of these inherent problems with aDNA, and because of the limits of genetic technology at the time, the types of organisms, substrates, and sequences that could be studied were limited: aDNA was almost always only retrievable from robust, single-source substrates such as bones and museum tissue specimens (Thomas et al. 1989; Thomas et al. 1990). As such, much of the DNA isolated was obtained from either relatively recently extinct animals (the past several hundred years, such as the quagga and thylacine) (Higuchi et al. 1984; Krajewski et al. 1997), or from specimens that were ‘snap-frozen’ in cold climates, such as permafrost (for instance, woolly mammoth) (Ozawa et al. 1997; Greenwood et al. 1999). This was because such substrates offer the best opportunity for DNA preservation: that is, DNA is less likely to be degraded if it is young, and similarly, DNA is less likely to be damaged if it is subjected to a constant, low-temperature environment (Lindahl 1993; Hofreiter et al. 2001b; Willerslev and Cooper 2005). For instance, Höss et al. (1996) used gas chromatography/ mass spectrometry (GC/MS) to measure the amount of oxidative base modifications in specimens representing different tissue types, ages, and preservation conditions. They found that specimens from arctic and sub-arctic regions exhibited lower amounts of oxidative base modifications compared with those from warmer regions (Höss et al. 1996b), supporting the idea that low temperature is crucial for the long-term preservation of aDNA. Thus, there was not a great diversity of organisms with adequate DNA preservation to be studied.

In addition, the DNA that was obtained from these substrates was mostly mitochondrial, and therefore, studies were limited to the analysis of short regions of mitochondrial or chloroplast genes, such as *12S rRNA*, *cytb*, or *rbcl* (Cooper et al. 1992; Ozawa et al. 1997). This is because mitochondria and plastids have a higher copy number per cell than nuclear DNA and so loci located on these genomes have a better chance of being amplified. However, these markers may not accurately reflect deep evolutionary relationships (for instance, the mitochondria can be prone to saturation), and accordingly, they may not be an adequate proxy for species evolution (Willerslev and Cooper 2005). Thus, in many cases, the evolutionary relationships inferred from only one or a few genes located on these organelle genomes were still branded with an overarching question mark (Cooper 1994).

#### **1.1.4 THE ‘NEXT-GENERATION’ REVELATION:**

##### **HOW HAVE THESE CHALLENGES BEEN OVERCOME?**

In the past decade, the development of new protocols and techniques has helped overcome many of the challenges of working with aDNA (Table 1.1). Firstly, the adoption of stringent protocols for the study of aDNA as an accepted minimum standard was arguably the greatest strategy toward combating issues such as false positives arising from contamination, and sequence errors arising from technical error or post-mortem damage (Kwok and Higuchi 1989; Willerslev and Cooper 2005; Shapiro and Hofreiter 2012). The ‘criteria of authenticity’ (Cooper and Poinar 2000) for aDNA include: working in a designated pre-PCR ‘clean room’ that is physically separated from a post-PCR facilities (i.e., a room with regular UV irradiation, isolated ventilation, air filtration, positive pressure, limited access with full personal protective equipment, and regular cleaning with bleach; Knapp et al. 2012); incorporating negative control amplification reactions, such as ‘blanks’ containing water instead of DNA; reproducing results by performing multiple PCR reactions and extractions, as well as replicating results in an independent laboratory; decontaminating equipment, reagents and specimens using UV irradiation, bleach, or ultra-filtration (Champlot et al. 2010); cloning of PCR products to assess damage introduced by PCR, and assess contamination; assessing the plausibility of DNA preservation through amino acid racemisation (Poinar et al. 1996) or GC/MS (Poinar and Stankiewicz 1999); the finding of DNA in associated remains; and quantifying the number of starting aDNA molecules in an extract using competitive PCR or quantitative PCR (Cooper and Poinar 2000; Hofreiter et al. 2001b; Willerslev and Cooper 2005; Shapiro and Hofreiter 2012). These criteria are designed to ensure that researchers destroy as much contaminating DNA as possible, prevent the movement of contaminating amplicons between specimens or extracts, rule-out the possibility of within-laboratory contamination, highlight sequence errors, and demonstrate beyond a reasonable doubt that aDNA is present (Willerslev and Cooper 2005). Adherence to these criteria helped ensure the legitimacy of future aDNA claims.

Secondly, the development of quantitative PCR (qPCR) allowed the quantification of the number of starting aDNA template copies (either absolute or relative), as well as the rapid detection of contamination and inhibition (Pruvost and Geigl 2004;

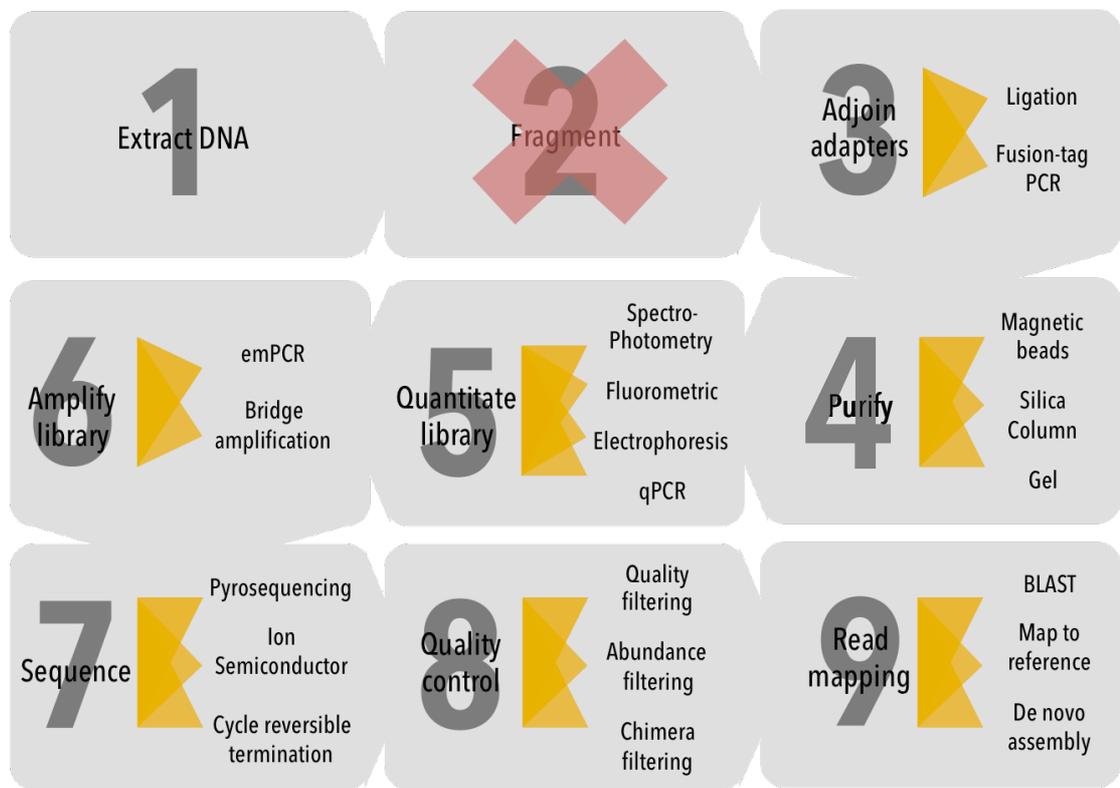
**TABLE 1.1 | A SUMMARY OF COMMON PROBLEMS ASSOCIATED ADNA AND THE TECHNICAL DEVELOPMENTS MADE TO MITIGATE THEM.**

<b>aDNA problem</b>	<b>Cause</b>	<b>Solutions</b>	<b>References</b>
Low yield	Inherent	PCR quantitation and amplification	(Heid et al. 1996; Morin et al. 2001; Eid et al. 2009)
	Storage	Use freshly excavated specimens	(Pruvost et al. 2007)
	Substrate	Optimised DNA extraction protocols	(Rohland and Hofreiter 2007b, a; Oskam et al. 2010)
Inhibition	Strand breaks	Amplify short, overlapping fragments	(Shapiro and Hofreiter 2012)
		Next-generation sequencing	(Margulies et al. 2005; Bentley et al. 2008)
	Blocking lesions	Polymerase extension profiling	(Heyn et al. 2010)
	Cross-links	PTB treatment	(Vasan et al. 1996)
	Inhibiting compounds	Repeat silica extraction	(Kemp et al. 2006)
		Dilution and evaluation with qPCR	(Soltis and Soltis 1993; Pruvost and Geigl 2004; Kontanis and Reed 2006)
		Use of specialised polymerases	(Kermekchiev et al. 2009)
	Use of PCR enhancers	(Kreader 1996; Zhang et al. 2010)	
Sequence errors	Miscoding lesions	UNG treatment	(Briggs et al. 2010)
		Multiple extractions and amplifications	(Shapiro and Hofreiter 2012)
		Cloning	(Willerslev and Cooper 2005)
		High fidelity polymerases	(d'Abbadie et al. 2007; Gloeckner et al. 2007; Molak and Ho 2011; Dabney and Meyer 2012; Overballe-Petersen et al. 2012)
		High sequencing coverage	(Overballe-Petersen et al. 2012)
		Statistical and bioinformatics models	(Bower et al. 2005; Helgason et al. 2007; Ho et al. 2007; Rambaut et al. 2009; Ginolhac et al. 2011)

Contamination	Amplicon carry-over	uPCR	(Longo et al. 1990; Pruvost et al. 2005)
		Primer tagging	(Binladen et al. 2007)
	Endogenous or exogenous DNA	Adhere to the "criteria of authenticity" and follow strict decontamination procedures	(Willerslev and Cooper 2005; Champlot et al. 2010)
		Cloning	(Willerslev and Cooper 2005)
		Identify with qPCR	(Pruvost and Geigl 2004)
	Modern DNA	C-T transition profiling	(Sawyer et al. 2012)
		Standard operating procedure for specimen collection	(Pruvost et al. 2007; Bollongino et al. 2008; Fortea et al. 2008; Pruvost et al. 2008)
	Specific	Blocking primers	(Gigli et al. 2009)

Kontanis and Reed 2006). Unlike conventional PCR, qPCR measures the fluorescence generated when a reporter is cleaved from double stranded DNA during each cycle of amplification; this fluorescence is proportional to the amount of DNA present, and therefore, the absolute initial quantity of template DNA present in the extract can be extrapolated by comparison with a standard of known concentration (Heid et al. 1996). Alternatively, the samples can be quantified relative to one another by means of comparing cycle threshold ( $C_T$ ) values. The  $C_T$ -value is the cycle at which fluorescence exceeds the background, and is inversely related to the amount of starting template. Extracts containing a greater number of aDNA templates will have a low  $C_T$ -value, whereas extracts containing a low number of aDNA templates will have a high  $C_T$ -value. This is because the reaction will exceed background levels sooner in the presence of more DNA copies (Pruvost and Geigl 2004). Although there are several methods available to quantify extracts, including UV spectrophotometry (such as NanoDrop), capillary gel electrophoresis (such as *Agilent* Bioanalyzer), and fluorometry (such as the Qubit fluorometer; (White et al. 2009; Zheng et al. 2010) (see White et al. 2009 for a comparison of available quantitation methods), these methods require a large amount of DNA, and cannot quantify amplifiable DNA alone (Zheng et al. 2010); that is, contaminating DNA and RNA may also be measured. As such, quantification using these methods is often unreliable, and to date, qPCR remains the most accurate method for quantifying the number of template copies in aDNA extracts and sequencing libraries (Meyer et al. 2008). Being able to visualise the dynamics of amplification also allows levels of inhibition to be assessed by examining the reaction efficiency; the accumulation of DNA as it is amplified from inhibited extracts will not appear ‘sigmoidal’ as it would in non-inhibited extracts. Thus, extracts may be diluted to reduce the concentration of inhibiting compounds in the extract, or a specialised polymerase may be used to overcome inhibition (Kermekchiev et al. 2009). These attributes make qPCR an ideal way to screen specimens for those containing the highest quantity and quality of amplifiable DNA for subsequent sequencing (Morin et al. 2001).

Most importantly, the advent of next-generation sequencing (NGS) revolutionised the field of ancient DNA, making it faster, cheaper, and easier than ever to study aDNA. Unlike traditional sequencing technology (Sanger sequencing) that requires



**FIGURE 1.3** | THE NGS WORKFLOW FOR USE WITH ANCIENT DNA. Step 2 is not usually needed when dealing with aDNA, due to the already fragmented nature of the template molecules.

long stretches of DNA (Sanger et al. 1977), NGS is optimally suited for sequencing short DNA fragments (< 50-500 bp), even those too short to be amplified by PCR. In this way, the highly fragmented nature of aDNA can be exploited by using NGS. In the NGS workflow (Figure 1.3), sequencing adapters, which are binding sites for the sequencing primer, are either ligated or amplified (using ‘fusion primers’) onto the ends of DNA fragments or PCR products to create a library containing many sequences, which is quantified (by qPCR, fluorometry, spectrophotometry, or gel electrophoresis) to determine the concentration of template molecules in the library. The library is then amplified in parallel either on the surface of beads within micro-reactors created by emulsion PCR (Nakano et al. 2003), or on the surface of a silica matrix by ‘bridge amplification’ (Millar et al. 2008). Next, the library can be sequenced, or enriched to recover template-positive beads and sequenced. Simultaneous sequencing of the many different DNA fragments within a library can be achieved using one of several unique chemistries. Over the past decade, four NGS platforms have dominated in the field of ancient DNA: these utilise (i) pyrosequencing (light) technology (454/Roche; Margulies et al. 2005; Margulies et al. 2007), (ii) fluorescence-based sequencing-by-synthesis technology (*Illumina*; Bentley et al. 2008), (iii) sequencing-by-oligo-ligation technology (SOLiD), and (iv) ion semi-conductor (pH) technology (IonTorrent, *Life Technologies*). While some platforms (e.g., *Illumina*) may be more accurate than others (e.g., IonTorrent), there is a trade-off between the quality of the output, the length of the output, and the time, cost, and labor input (Loman et al. 2012). All things considered, the choice of platform depends largely on the question that is being addressed, and financial restrictions (Quail et al. 2012): for instance, metagenomic analyses may require the unparalleled high coverage and affordability of the IonTorrent, while taxonomic identifications may require the high sequence accuracy of *Illumina*, and *de novo* genome assembly is simplified by the long read length provided by 454/Roche (Shokralla et al. 2012).

NGS can help overcome some of the challenges of working with aDNA. For instance, barcoding (or ‘indexing’) largely eliminates the issue of carry-over contamination because the PCR products of every amplification reaction are given a unique tag such that carry-over contamination can be traced back to its source at any time. Another advantage of using NGS to analyse aDNA is that error can typically be

uncovered because it will only appear in a small proportion of the reads compared with the true variant (Overballe-Petersen et al. 2012). NGS also allows ‘shotgun’ sequencing—the sequencing of DNA extracts without targeted amplification (Millar et al. 2008). Shotgun sequencing captures the entire assemblage of DNA sequences present in a sample, including contamination. After sequencing, overlapping reads are aligned to recreate a contiguous sequence (Margulies et al. 2005; Bentley et al. 2008). In this way, long stretches of both mitochondrial and nuclear DNA may be recovered to rapidly reconstruct whole genomes—something that was not easily achievable before. Contamination can be identified by comparison of reads to known reference sequences within a database such as GenBank (Benson et al. 2006) using computer algorithms such as ‘basic local alignment search tool’ (BLAST; Altschul et al. 1990; Millar et al. 2008) or BRONX (Little 2011).

Finally, the problems generated by contamination and post-mortem damage to aDNA templates have also been mitigated by other developments that are collated in Table 1.1 and are discussed briefly below:

**PTB.** The pre-treatment of extracts with PTB (N-phenacylthiazolium bromide) will cleave the intra- and inter- molecular cross-links that form within and between DNA molecules, as well as between DNA and proteins (Vasan et al. 1996). These cross-links are responsible for a large amount of PCR inhibition, so cleavage of these cross-links allows PCR amplification of aDNA targets to proceed uninhibited.

**UNG.** UNG removes uracils that were the result of deamination of cytosine residues, and repairs the abasic sites (Briggs et al. 2010). Treatment with UNG will therefore remove the post-mortem C-T transitions that can mimic evolutionary change. However, Sawyer et al. (2012) examined the number of C-T transitions in 86 specimens ranging from 18 to 60,000 yr, and found that the frequency of C-T transitions at the 3’ end of aDNA fragments increases with time. Thus, the pattern of C-T base misincorporations in an ancient sequence can be used as proof of its antiquity (Sawyer et al. 2012; Shapiro and Hofreiter 2012), but this would not be possible with prior treatment with UNG. On the other hand, there has been some suggestion that UNG treatment following an initial PCR amplification with dUTP as opposed to dTTP (i.e., uPCR), but prior to a follow-up PCR, can prevent carry-over

contamination from previous amplification products (Longo et al. 1990; Pruvost et al. 2005); however, this method is cannot eliminate contamination arising from other sources (Champlot et al. 2010).

**OPTIMISED SAMPLE COLLECTION AND EXTRACTION PROTOCOLS.** The use of optimised extraction protocols and stringent collection procedures can greatly increase the yield of aDNA as well as reduce the likelihood of modern contamination. Several protocols have been tailored for aDNA extraction from specific substrates such as bone (Höss and Pääbo 1993; Yang et al. 1998; Kalmár et al. 2000; Ye et al. 2004; Campos et al. 2012) and teeth (Rohland and Hofreiter 2007a), or from substrates that have exhibited a high amount of PCR inhibition (Hänni et al. 1995; Kemp et al. 2006). Others have described multiple alternative methods for efficient aDNA extraction (Rohland and Hofreiter 2007b); having options increases the likelihood that one protocol will be suitable if another is not. In addition, it has been found that improper handling and storage of specimens can cause as much DNA degradation after excavation as occurred during the entire length of burial; as such, freshly excavated fossils yield more DNA than those that have been stored in less-than-ideal conditions (Pruvost et al. 2007). Guidelines have also been proposed for the excavation and subsequent storage of specimens for aDNA analysis (Fortea et al. 2008; Pruvost et al. 2008), the most important considerations being the prevention of further DNA degradation by storage in a cool, dry place without major fluctuations in ambient temperature, and the prevention of contamination by excavating with personal protective equipment such as gloves and face masks (Bollongino et al. 2008b).

**POLYMERASE EXTENSION PROFILING AND SINGLE PRIMER EXTENSION.** Polymerase extension profiling (PEP) can be used to identify the sites of blocking lesions in aDNA templates. In this method, a primer-binding sequence is ligated to the 5' end of the aDNA templates, and an 'end-of-template' recognition sequence is ligated to the 3' end. A primer is annealed and extended, with extension terminating when it comes to the site of a blocking lesion. The single primer extension products are then sequenced and aligned to show the sites of blocking lesions, and to identify full-length (undamaged) amplicons (i.e., those containing the 'end-of-template' recognition sequence) (Heyn et al. 2010). In a similar method, single primer

extension (SPEX) can be used on aDNA templates to obtain detailed information about post-mortem base modifications and, consequently, more accurate sequence data (Brotherton et al. 2007).

**MULTIPLEX AMPLIFICATION OR DIRECT MULTIPLEX SEQUENCING.** This method can be used to amplify the entire mitochondrial genome of an organism with two rounds of amplification. Primer pairs are designed that will amplify overlapping regions of the mitochondrial genome. All primers are placed together in one PCR reaction, and many products are generated. The PCR product is diluted and divided into as many new reactions as primer pairs, and a second amplification targets each product, one per reaction (Krause et al. 2006). This approach uses only as much DNA template as one PCR reaction would, thereby making efficient use of limited aDNA templates (Stiller and Fulton 2012). Direct sequencing of multiplex PCR products helps eliminate some of the biases introduced by the second round of PCR, and reduces labor (Stiller et al. 2009).

**HIGH FIDELITY POLYMERASES AND PCR ENHANCERS.** The use of high-fidelity polymerases, or polymerases with other specific properties (such as proof-reading) during PCR amplification of aDNA templates can alleviate issues such as the termination of strand synthesis, base misincorporations, and the presence of length and GC biases (Overballe-Petersen et al. 2012). For instance, *Taq* polymerase will treat uracil as thymine, resulting in C-T transitions, whereas *Pfu* polymerase will terminate synthesis at uracil residues (Molak and Ho 2011). However, some polymerases have been synthesised to tolerate highly damaged DNA by bypassing abasic sites or template mismatches (d'Abbadie et al. 2007; Gloeckner et al. 2007). Such polymerases have been shown to out-perform *Taq* polymerase by three times, and increase PCR sensitivity in the presence of low yields of aDNA (d'Abbadie et al. 2007). Similarly, some commercial polymerases have been shown to introduce less bias (such as AccuPrimePfx polymerase) than others (such as Phusion and AmpliTaqGold polymerases) (Dabney and Meyer 2012). In addition, the use of PCR enhancers, such as bovine serum albumin (BSA), betaine, and dimethylsulfoxide (DMSO) have been shown to help PCR reactions overcome inhibition or increase the specificity and strength of primer-binding to their targets (Kreider 1996).

**STATISTICAL MODELS AND BIOINFORMATICS.** Statistical models can be applied to sequence data to help detect endogenous DNA sequences over contaminating sequences (Helgason et al. 2007), and to map nucleotide misincorporations for assessing damage and validating the antiquity of a sequence (Ginolhac et al. 2011). Because sequence errors and damage can obscure true mutations, programs have been developed to take damage into account during phylogenetic analyses (Ho et al. 2007; Rambaut et al. 2009).

**BLOCKING PRIMERS.** Blocking primers can be used to prevent the amplification of specific contaminating sequences, such as DNA from a predator, host, or human. Blocking primers are modified to terminate extension when bound to the contaminant DNA, preventing amplification of the contaminant and allowing the amplification of target aDNA (Gigli et al. 2009). For instance, without the use of blocking primers, only human contamination was amplified from permafrost samples; however, when blocking primers were used, DNA from the extinct woolly rhino (*Coelodonta antiquitatis*) was detected (Boessenkool et al. 2012). In another example, Gigli et al. (2009) found that the recovery of Neanderthal DNA from bone was increased by 65% through the use of primers that block the amplification of modern human DNA.

### **1.1.5 ADNA TODAY:**

#### **WHAT IS ANCIENT DNA CURRENTLY BEING USED TO INVESTIGATE?**

##### **2001-2012**

The development of these new protocols and techniques over the past decade has made it possible for both nuclear and mitochondrial aDNA to be extracted from older specimens and a broader range of organisms (Table S1.6.1). This in turn allowed a variety of novel and diverse evolutionary questions to be addressed (Table S1.6.1), taking the field to a new level of scientific inquiry.

Advances in extraction protocols have allowed DNA has been recovered from a variety of novel substrates (Gilbert et al. 2004; Lee and Hallam 2009; Oskam et al. 2010). For example, DNA has been extracted from moa eggshell (Huynen et al.

2010; Oskam et al. 2011; Oskam et al. 2012), feathers (Rawlence et al. 2009), and ancient plants (Palmer et al. 2012), and woolly mammoth (Gilbert et al. 2007) and ancient human hair (Gilbert et al. 2008b; Rasmussen et al. 2010; Rasmussen et al. 2011). In many cases, the DNA recovered from these substrates displayed a marked improvement in quality and quantity over traditional substrates such as bone and tissue. For instance, Oskam et al. (2010) recovered both nuclear and mitochondrial aDNA from moa eggshell, showing that eggshell is conducive to DNA preservation. Furthermore, it was found that DNA retrieved from eggshell has a bacterial load 125 times less than bone of the same age: this may be because eggshell is water resistant and therefore more protected from microbial decay and hydrolytic damage than bone is, due to its porosity (Oskam et al. 2010). Hair has also been found to be an excellent substrate for aDNA preservation, has low levels of contamination, and is a nondestructive way to obtain DNA from specimens (Bengtsson et al. 2012). Gilbert et al. (2007) sequenced DNA from woolly mammoth hair specimens and found that the amount of damage in endogenous DNA was much lower than the amount of damage found in DNA obtained from bone. This is largely because the keratinisation of cells forms a barrier against both microbial attack and moisture that causes hydrolytic damage (Gilbert et al. 2004; Gilbert et al. 2007).

In addition, aDNA has been retrieved from a variety of mixed-source samples, including ground sloth coprolites (Hofreiter et al. 2000; Poinar et al. 2001; Hofreiter et al. 2003; Poinar et al. 2003; Wood et al. 2008), herbivore middens (Murray et al. 2012), and permafrost sediments (Willerslev et al. 2003; Haile et al. 2009; Jørgensen et al. 2012). The advent of NGS metabarcoding has been instrumental in being able to sequence these complex mixtures of aDNA. Metabarcoding involves the use of highly conserved primers that are able to bind to DNA from multiple different species, yet amplify a region (a DNA “barcode”) that is variable enough to distinguish between species based on its sequence (Taberlet et al. 2012). Metabarcoding has facilitated an expansion in the types of substrates able to be used for aDNA analysis, especially from environmental samples, and therefore has broadened the types of questions that can be addressed (Table S1.6.1). For example, the DNA from coprolites can be used to study palaeodiet. Wood et al. (2008) used coprolites of moa to investigate the make up their diet, and to compare their diet to the diet of introduced ungulates. They found that the diet of moa does not overlap

with the diet of the introduced ungulates, refuting the possibility that they may fill the role once played by moa in the ecosystem (Wood et al. 2008). In other cases, the DNA recovered from environmental samples has allowed the study of organisms that are not represented in the macrofossil record. For instance: reconstructions of palaeoenvironments has been possible using plant DNA amplified from coprolites and sediments (Hofreiter et al. 2003; Willerslev et al. 2003; Jørgensen et al. 2012); late survival hypotheses were tested through the amplification of megafaunal aDNA from permafrost cores, allowing Haile et al. (2009) to more accurately assess extinction causes; and a previously unrecognised sloth species was characterised from DNA recovered from a coprolite (Hofreiter et al. 2003).

The advances in PCR and sequencing described in section 1.1.4 have improved reconstructions of population history: this includes population movements, extinctions, and how changes in genetic diversity correlate to changes in geography or the environment (e.g., climate; Willerslev and Cooper 2005). aDNA is vital to making accurate reconstructions of population history because:

*several processes can result in similar patterns of modern genetic data [for example, population subdivision could indicate recent divergence or old divergence followed by recent gene flow]. Thus, true demographic history can be difficult to reconstruct based on modern genetic data alone.*

(Ramakrishnan and Hadly 2009)

The population histories of many organisms can be studied through analysis of their genetic diversity over time and space, or by comparison to another species (Table S1.6.1). This involves sequencing regions of the genome in individuals from a historic population or species, and comparing the sequences to those from other populations or species, which may be from another area or time (for example, comparison with an extant species). This approach has been used to trace the domestication of cattle (Troy et al. 2001; Götherström et al. 2005; Achilli et al. 2008), investigate human dispersals (Gilbert et al. 2008b; Rasmussen et al. 2011), test extinction hypotheses (Nyström et al. 2010; Nyström et al. 2012), and justify conservation strategies (Larson et al. 2002; Dalén et al. 2007; Bray et al. 2013). For

example, Nyström et al. (2010) analysed 741 bp of mtDNA of a specimen from the last surviving population of mammoths. They found the population underwent a bottleneck, possibly a founder effect, resulting in the loss of genetic diversity; however, this genetic diversity remained stable for 5000 yr before sudden extinction, suggesting another possible cause for extinction, such as the arrival of humans or disease (Nyström et al. 2010; Nyström et al. 2012). This demonstrates how population data derived from aDNA can be used to test extinction hypotheses. Similarly, 34 pre-fur trade otter specimens were genotyped for four nuclear microsatellites and one mitochondrial locus, and the genetic variation was compared to five extant populations of sea otter genotyped at the same loci (Larson et al. 2002). They found that the mean heterozygosity (a measure of genetic diversity) in pre-fur trade otters was greater than extant otters, indicating that extant populations have undergone a population bottleneck resulting from over-hunting that massively reduced genetic diversity (Larson et al. 2002). Their risk of extinction may be mitigated by assisting gene flow between extant populations (Larson et al. 2002). These studies show how of the population history of an extant species, which is directly examinable using aDNA, can be important for conservation.

Advances in PCR and sequencing technology have also opened up a new field of palaeogenomics, the study of ancient genomes:

*Genome wide sampling [is required] to identify and to separate locus-specific effects from genome-wide effects...this is crucial because only genome-wide effects inform us reliably about population demography and phylogenetic history, whereas locus-specific effects help identify genes that important for fitness and adaptation.*

(Luikart et al. 2003)

It has been possible to sequence the full mitochondrial genome of many extinct organisms, such as Moa sp. (Cooper et al. 2001), mammoth (*Mammuthus primigenius*; Krause et al. 2006), and thylacine (*Thylacinus cynocephalus*; Miller et al. 2009). Complete mitochondrial genomes provide a larger dataset on which to base phylogenies and more accurately determine rates of molecular evolution. For example, Cooper et al. (2001) sequenced the mitochondrial genome of two moa

species: these genomes were later used to estimate the time of divergence of the moa lineage from other birds (Phillips et al. 2010). Phillips et al. (2010) estimated this time to be around 70 Ma ago, but more surprisingly, they found that moa are most closely related to the volant (flighted) tinamous of South America, contrary to existing evidence.

However, as mentioned in section 1.1.4, even whole mitochondrial genomes still only represent one genetic locus, and therefore do not necessarily reflect the evolution of the entire genome. Indeed, many phylogenetic relationships and divergence estimates differ depending on the genome from which they are inferred (Miller et al. 2008). Thus, many phylogenies remain disputable. The advent of NGS, combined with the extraction of aDNA from samples with a relatively high endogenous content (e.g., hair) has allowed nuclear DNA to be recovered. Nuclear loci can better inform phylogenies and more accurately determine rates of molecular evolution (Miller et al. 2008). For instance, nuclear loci from mammoths and mastodons have been used to confirm their placement within the elephant phylogeny (Miller et al. 2008), and to model the demographic history of elephants, including their divergence times (Rohland et al. 2010). These studies revealed surprising relationships within the Elephantidae, namely, that species of African elephant are as equally divergent from one another as mammoths are from Asian elephants, justifying current taxonomic assignments and putting to rest the controversy surround their classification (Rohland et al. 2010). This finding would not have been possible without the insight that mammoth nuclear DNA provided. Nuclear microsatellite markers have also been developed for extinct organisms, such as the moa (Allentoft et al. 2009; Allentoft et al. 2011). Microsatellites are sequences consisting of short tandem repeats, and different alleles are characterised by the number of repeats at a locus. Because of this, microsatellites have greater resolution to detect population-level changes than other types of polymorphic markers (such as SNPs; Allentoft et al. 2011). Thus, genotyping many nuclear microsatellites in individuals offers a potentially rapid method of detecting small differences in genetic diversity between populations, which is essential for fine-scale studies of phylogeography.

Nuclear DNA has also allowed unfossilisable phenotypes to be investigated for the first time, as many phenotypes are encoded by nuclear genes. These phenotypes include appearance such as pigmentation (Römpler et al. 2006; Lalueza-Fox et al. 2007; Rasmussen et al. 2010; Draus-Barini et al. 2013), plumage (Rawlence et al. 2009), behavior (Huynen et al. 2010), and protein function (Krause et al. 2007; Lalueza-Fox et al. 2008; Lalueza-Fox et al. 2009; Campbell et al. 2010). Phenotypes such as these allow the investigation of adaptations that are not encoded by mitochondrial genes. For example, a region of the *MC1R* gene associated with reduced skin pigmentation in humans was amplified and sequenced in two Neanderthal specimens (Lalueza-Fox et al. 2007). A variant of this gene was detected that is not found in modern human populations: this variant also reduces the function of *MC1R* such that it would change skin pigmentation. The authors conclude that it is likely that Neanderthals had light skin and red hair, but that the phenotype arose independently in humans (Lalueza-Fox et al. 2007). These results support a role for convergent evolution, and therefore suggest that this phenotype was adaptive. This is a novel and powerful approach to studying the processes that have shaped the evolution and adaptation of extinct organisms.

Finally, the recovery of nuclear DNA and the advent of NGS allowed the draft sequencing of the full nuclear genome of two extinct archaic humans, the Neanderthal (*Homo neanderthalensis*; Green et al. 2010), and the Denisovan hominin (Reich et al. 2010). Having large regions of the nuclear genome allows the identification of adaptive molecular variation (Luikart et al. 2003), and hypotheses about the processes that have shaped the evolution of the species to be tested—processes such as natural selection, genetic drift, and gene flow. For instance, Green et al. (2010) identified signals of selective sweeps in modern humans by screening the genome for regions of difference between humans and Neanderthals that arose after their divergence. One candidate gene for positive selection in humans that was identified is *RUNX2*, which affects the morphology of the skull, shoulders, ribcage and teeth. This gene may have been under selection in modern humans, and could explain the differences we see in these aspects of morphology between humans and Neanderthals (Green et al. 2010). In addition, Green et al. (2010) tested the ‘out-of-Africa’ model of human origins by determining the extent and direction of gene flow between Neanderthals and modern humans. They found evidence of gene flow from

Neanderthals into non-African humans, which supports a more complex model for human origins than the traditional ‘out-of-Africa’ model. This study is the first to fulfill one of the ultimate goals of aDNA research: a detailed history of the population dynamics, evolutionary processes, genes, and interactions that have shaped the evolution of an extinct organism, as well as the evolution of our own species. This study epitomises just how much information can be gleaned from the study of ancient genomes.

### **1.1.6 PRESENT PROBLEMS *and* PROSPECTS:**

#### **WHAT CHALLENGES REMAIN NOW, AND HOW MIGHT THEY BE OVERCOME IN THE FUTURE? 2013—**

With new developments in technology, new challenges have arisen. These challenges vary depending on the sequencing technology used, methods used for interrogating the data, the degree and nature of DNA preservation, and the sample type. Two of the most pertinent challenges facing the field of aDNA include: (1) overcoming or mitigating the biases introduced by new technologies and methodologies (in particular, metabarcoding workflows), and (2) retrieving aDNA from warm-climate ecosystems. The scope of these biases and limitations are largely under-appreciated.

In the first category, new technologies and methodologies have their own unique biases and limitations that are introduced from the experimental design phase through to sequencing and data analysis. These biases are particularly pronounced when dealing with aDNA extracted from environmental samples (such as sediments) and other mixed-sources (such as coprolites). Firstly, the strategies employed to sample and subsample substrates for aDNA extraction can introduce bias: in addition to the contamination and further degradation that can be caused by improper handling and storage of samples (discussed in section 1.1.3), sample preparation and inadequate sampling can affect the repeatability and accuracy of aDNA studies. Secondly, during aDNA extraction, there can be a bias toward the size of fragments that are retrieved simply based on the method employed (Dabney et al. 2013), resulting in a larger proportion of contaminating DNA relative to endogenous DNA being recovered. Next, PCR amplification biases are exacerbated when amplifying highly complex, mixed aDNA samples: for instance, Murray et al. (2015) found that

**TABLE 1.2 | A BRIEF COMPARISON OF NGS PLATFORMS CURRENTLY AVAILABLE (updated from Millar et al. 2008).**

<b>Platform</b>	<b>Library amplification</b>	<b>Sequencing chemistry</b>	<b>Reaction surface</b>	<b>Average read length (bp)</b>	<b>Average throughput/run (Gb)</b>	<b>Average # Reads/Run</b>	<b>Accuracy (%)</b>	<b>Average run time (h)</b>
454/Roche GS Junior	emPCR	Pyrosequencing (light emission and detection)	PicoTiter Plate	400	35	100,000	99.5	10
454/Roche GS FLX+	emPCR	Pyrosequencing (light emission and detection)	PicoTiter Plate	700	700	1,000,000	99.997	23
<i>Illumina</i> MiSeq	Solid-phase anchored oligo bridge	Cycle reversible termination (fluorescence)	Flow cell	250	8.5	15,000,000	99	4-39
<i>Illumina</i> HiSeq	Solid-phase anchored oligo bridge	Cycle reversible termination (fluorescence)	Flow cell	100-150	90-600	600,000,000-6,000,000,000	98.5	7 h- 11 d
SOLiD 5500W	emPCR	Sequencing by ligation (fluorescence)	Flowcell	50-75	80-320	< 6,000,000,000	99.99	2-7 d
Ion Torrent PGM	emPCR	Ion semiconductor (pH)	Silica Chip	200-400	0.3-2	400,000-5,500,000	98.5	2.3-7.3
Ion Torrent Proton	emPCR	Ion semiconductor (pH)	Silica chip	200	10	60,000,000-80,000,000	98.5	2-4

the presence of inhibitors and low starting template in extracts containing a high proportion of *Sardinops* DNA relative to *Engraulis* resulted in the inconsistent detection of *Engraulis* in replicate PCRs. There are also unique biases introduced depending on the NGS sequencing platform used (Table 1.2 summaries some of the ‘pros and cons’ of the four NGS platforms currently available<sup>1</sup>): *Illumina* misincorporates bases at a high frequency, resulting in single-nucleotide changes that may be mistaken for true polymorphisms (Loman et al. 2012), and conversely, *IonTorrent* and *454/Roche* have trouble distinguishing bases that form homopolymer run stretches, resulting in insertion or deletion errors (Loman et al. 2012). Although most of these data issues can be overcome by stringent filtering of the data after sequencing (including trimming, the removal of low-quality and low-abundant reads, and identifying chimeras using programs such as *Galaxy*, *UCHIME*, and *Geneious*; Kircher 2012; Edgar et al. 2011), there are no set criteria for quality control, which should be evaluated on a case-by-case basis. Whatever platform is used, the library preparation, sequencing workflow, and subsequent data analysis requires empirical optimisation. Finally, the taxonomic identification of reads can be obscured by aDNA damage and sequencing errors, short read lengths, a lack of a reference database for many extinct organisms, and an incomplete genetic database of modern species (Schubert et al. 2012). These factors are extremely important to consider when working with aDNA, yet by-and-large are still not given enough attention. As the use of mixed-source samples and metabarcoding workflows increase in popularity, more consideration needs to be given to mitigating the associated biases. Thus, developing practices to achieve this remains key.

In the second category, aDNA from warm environments is often more severely degraded and contaminated than aDNA preserved in cool environments, making it much more challenging to isolate informative DNA from warm-climate ecosystems. Although DNA has been retrieved from environmental deposits known to be unfavorable, or even detrimental, for DNA preservation (including hot, arid environments; Poinar et al. 2003; Murray et al. 2012; Oskam et al. 2012), there

---

<sup>1</sup> See Mardis et al. (2008a, 2008b) for a detailed explanation of sequencing chemistries, and Miller et al. (2008), Loman et al. (2012), Shokralla et al. (2012), Pareek et al. (2011) for a detailed comparison of the currently available platforms.

remains a bias in the literature towards aDNA studies that centre around cool-climate ecosystems, especially from Europe, Siberia, and other places in Northern Hemisphere. For example, as of 2013, “only three aDNA studies of sub-Saharan African material have been published...[and] no ancient DNA study of sub-Saharan archaeological material has been published” (Campana et al. 2013). Furthermore, of the 152 aDNA studies summarised in Table S1.6.1, over 89% focus on cool-climate ecosystems. The divide is even greater when considering the number of studies that were able to isolate whole mitochondrial genomes or nuclear loci, or aDNA from samples older than a few thousand years. The use of ‘novel’ substrates will increase the probability of retrieving aDNA from warm-climate ecosystems and other unfavorable environments. For instance, the protective qualities of eggshell means that the DNA within is less degraded by conditions such as heat and moisture than DNA within bone (see section 1.1.5; Oskam et al. 2010), which suggests that eggshell is a good candidate for retrieving aDNA from fossil birds that lived in tropical climates. Other studies have employed a metabarcoding approach to aDNA extracted from bulk fossil bone in order to retrieve information about past biodiversity from sites within arid climates (Murray et al. 2013, Haouchar et al. 2014). Thus, optimisation of aDNA extraction from these kinds of substrates is paramount. Another way of increasing the chances of aDNA recovery from warm climates is via the preferential capture and amplification of target aDNA over contaminating DNA through the enrichment of target DNA prior to sequencing. Some library preparation methods maximise the yield of aDNA in the sequencing library by capturing the single-stranded molecules (Gansauge and Meyer 2013) that would otherwise be lost due to single-stranded modifications and ‘breaks’ that are present in aDNA (Zheng et al. 2010): in this way, Gansauge and Meyer (2013) demonstrated a six-fold increase in yield of Neanderthal DNA compared to double-stranded library preparation methods. Other enrichment methods include: primer extension capture (PEC), where 5'-biotinylated primers capture target DNA from a library and allow contamination to be washed away (Briggs et al. 2009); bait capture, where long-range PCR products bound to beads hybridise with targeted mitochondrial genomes to extract them from mixed DNA samples (Gnirke et al. 2009; Maricic et al. 2010; Carpenter et al. 2013); and array-based sequence capture, where target DNA hybridises to oligo probes bound to a microarray (Bau et al. 2009; Burbano et al. 2010). Using these methods, Neanderthal sequences could be retrieved

from extracts consisting of 99.8% microbial DNA (Briggs et al. 2009; Burbano et al. 2010). However, these methods have not been tested on a variety of older fossils from more ‘unfavourable’ environmental conditions.

Assessing and applying the recently developed NGS techniques discussed above to novel substrates from warm climate ecosystems is one of the next frontiers in aDNA research, and is the overarching focus of the following thesis.

## **1.2 AIMS AND SCOPE *of* THESIS**

---

As discussed in Chapter 1.1, some of the limitations of traditional methods for studying evolutionary history, especially of extinct species, can be addressed by using ancient DNA (aDNA). However, the field of ancient DNA is still in its infancy, and many challenges arising from the inherently damaged, degraded, and contaminated nature of aDNA have yet to be overcome (Willerslev & Cooper 2005). As such, it is important to develop and apply new methods in order to exploit the full potential of aDNA, particularly from warm or tropical climates that are not conducive to DNA preservation.

In this thesis, we aimed to further optimise methods to extract DNA from a variety of novel ancient substrates including fossil bulk bone and fossil eggshell in order to study past biodiversity and evolutionary history in warm-climate ecosystems. This involved comparing several new aDNA extraction methods with established methods to maximise the yield and quality of template aDNA, then further develop the use of NGS techniques such as metabarcoding, shotgun sequencing, and target enrichment for use with highly degraded aDNA from these ancient substrates. At the same time, this molecular kit was applied to a variety of warm-climate settings in Australia and Madagascar in order to assess their viability as a tool for addressing outstanding palaeontological and archaeological questions, such as examining the processes that have caused spatial and temporal changes in biodiversity (section 1.3). Firstly, Chapter 2 set out to optimise the recently developed bulk bone sampling methodology (Murray et al. 2013), which involved conducting experiments to maximise the yield and quality of DNA, optimise efficiency, determine sample sizes

that will maximise power, determine the best bioinformatic analyses to employ, and compare how well species assemblages correlate with palaeontological records. Secondly, in Chapter 3, the single-stranded shotgun library build method (Gansauge and Meyer 2013) was applied to bulk bone for the first time, which was used to evaluate DNA preservation and identify past biodiversity. Next, Chapter 4 applied bulk bone metabarcoding to a warm-climate archaeological site to identify past fish assemblages and explore the subsistence practices of past peoples. In Chapter 5, the single-stranded shotgun library build method was applied to fossil eggshell for the first time and used to examine DNA preservation as well as test hypotheses regarding avian phylogenetic relationships and time-of-divergence. Finally, in Chapter 6, DNA extraction from fossil eggshell (Oskam et al. 2010) was improved upon by co-opting recently-developed ultra-short DNA extraction techniques (Dabney et al. 2013) and enriching for endogenous aDNA targets through the use of hybridisation capture in order to better understand the taxonomy, phylogeography, and population history of an extinct megafaunal bird.

### **1.3 STUDY SITES, SUBSTRATES *and* SPECIES**

---

In order to develop and apply methods to retrieve aDNA from warm or tropical environments that are not conducive to the long-term preservation of aDNA, we examined fossil bulk bone and eggshell from a variety of warm-climate settings within Australia and Madagascar.

Australia today is an arid island with a climate that ranges from temperate to tropical. Australia is home to rich fauna and flora that are found nowhere else in the world, and has a deep history of biodiversity fluctuations through time. Palaeontological sites in Australia offer an ideal opportunity to develop aDNA methods for use with highly degraded samples, and test whether these methods can be used to characterise past biodiversity. In this thesis, we use fossil bulk bone material from two palaeontological sites in South Australia (one on Kangaroo Island and one within the Naracoorte Caves World Heritage Area on the mainland) to examine past biodiversity and characterise aDNA preservation. Kangaroo Island (KI) just off the coast of South Australia is the only known land-bridge island with a diverse 100,000

year-old palaeontological record that includes species of marsupial megafauna such as *Zygomaturus*. Megafauna, and many other endemic species in Australia, went extinct approximately 50,000 years ago (Johnson 2006), and little is known about the cause of the extinction. KI is one of the only sites in Australia where there is a fossil record spanning the time before, during, and after this major decline in biodiversity occurred. As such, being able to retrieve aDNA from fossil deposits on KI is of great interest to both palaeontologists and conservation biologists. Although aDNA studies have been carried out at sites on KI before (Haouchar et al. 2014), it remains an excellent candidate for further development of bulk bone metabarcoding methods (Chapter 2). Likewise, no aDNA has been successfully retrieved from faunal fossils at World Heritage listed Naracoorte Caves on the mainland of South Australia, despite extremely good preservation of a variety of vertebrate fossils (including megafauna) as far back as half a million years. Bulk bone material from this site therefore offers a low-risk opportunity to not only characterise and predict the limits of aDNA preservation at Naracoorte, but also to demonstrate how genetic information can add to our understanding of past biodiversity and faunal turnover in Australia (Chapter 3). The methods developed here will be crucial for future aDNA studies to be carried out at these significant sites, as well as other palaeontological and archaeological sites located in warm-climates around the world, such as Madagascar.

Similar to Australia, Madagascar is a highly biodiverse island that has also undergone large changes in biodiversity over time. Many of these changes are thought to have occurred as a consequence of human activity. Because humans arrived in Madagascar relatively recently (i.e., a few thousand years ago) in comparison to many places around the world, including Australia, it is a good system to examine the impact of human activity on faunal change. The impact of humans on the marine environment is of particular interest as coastal people today struggle to obtain a sustainable livelihood; however, it is difficult to achieve an understanding marine faunal turnover over the past several hundred years because not only is fish bone especially difficult to identify morphologically, it is also difficult to identify genetically as aDNA preservation in tropical climates such as Madagascar is poor. As such, it is appropriate to test the application of a bulk bone approach to the

identification of archaeological fish bone, in order to glean information about how past peoples exploited marine resources (Chapter 4).

Madagascar's historical fauna also included unique terrestrial megafaunal species, including several genera of large, flightless elephant birds (members of the palaeognathae). These birds only went extinct less than a thousand years ago. Like the Australian megafauna, the cause of their extinction remains contentious, as does their origin and evolution, including their relation to other birds and the number of species. One reason for this is because only a few DNA sequences have been retrieved due to the lack of bone specimens with well-preserved DNA—presumably a function of the tropical environment being suboptimal for the preservation of bone. In 2010, Oskam et al. demonstrated that ancient DNA was preserved in fossil eggshell (q.v., section 1.1.5). Elephant birds laid the largest egg of any bird, and consequently, the eggshell is extremely robust and well-preserved. In this project, we aimed to retrieve genomic information from the elephant bird using aDNA extracted from its eggshell, with the ultimate goal of understanding their evolutionary relationship to other birds (Chapter 5). We also aimed at sequencing multiple mitochondrial genomes from elephant bird eggshell specimens across Madagascar in order to assess their genetic diversity and how it correlates with geographic distribution (i.e., assess phylogeographic patterns) and the current taxonomic framework (Chapter 6).

#### 1.4 SIGNIFICANCE

---

The methods developed during the course of this project, and the results obtained from the application of those methods, complement the traditional disciplines of palaeontology and archaeology, and may benefit other disciplines that also deal with highly mixed and degraded DNA, such as bacterial metagenomics, molecular ecology, and forensics. New aDNA methods such as these will allow DNA to be retrieved from older, more degraded substrates, and will also increase the cost-efficacy and accessibility of aDNA projects by maximising data output while decreasing labour, time, and resource input.

In addition, the results obtained provide new insights into the processes that have shaped the evolution and extinction of species. In this way, the results of this project will have implications for conservation because they aid in the understanding of historical faunal turnover against the backdrop of anthropogenic climate change and habitat modification. Finally, South Australia and Madagascar are of considerable cultural and economic import due largely to their unique flora and fauna. As such, the knowledge of past biodiversity gained through this project will be crucial for maintaining (or restoring) the biodiversity in these locations today.

## 1.5 REFERENCES

---

- Achilli A, Olivieri A, Pellecchia M, Uboldi C, Colli L, et al. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology* **18**: R157-R158.
- Ali SI. 1969. *Senecio lautus* complex in Australia .V. Taxonomic interpretations. *Australian Journal of Botany* **17**: 161-176.
- Allentoft ME, Schuster SC, Holdaway RN, Hale M, McLay E, et al. 2009. Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. *Biotechniques* **46**: 195-200.
- Allentoft ME, Oskam C, Houston J, Hale ML, Gilbert MT, et al. 2011. Profiling the Dead: Generating Microsatellite Data from Fossil Bones of Extinct Megafauna-Protocols, Problems, and Prospects. *PLOS One* **6**.
- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B-Biological Sciences* **279**: 4724-4733.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

- Andre A, Weiner A, Quillévéré F, Aurahs R, Morard R, et al. 2013. The cryptic and the apparent reversed: lack of genetic differentiation within the morphologically diverse plexus of the planktonic foraminifer *Globigerinoides sacculifer*. *Paleobiology* **39**: 21-39.
- Austin JJ, Ross AJ, Smith AB, Fortey RA, Thomas RH. 1997a. Problems of reproducibility - Does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society B-Biological Sciences* **264**: 467-474.
- Austin JJ, Smith AB, Thomas RH. 1997b. Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends in Ecology and Evolution* **12**: 303-306.
- Bau S, Schracke N, Kranzle M, Wu H, Stähler PF, et al. 2009. Targeted next-generation sequencing by specific capture of multiple genomic loci using low-volume microfluidic DNA arrays. *Analytical and Bioanalytical Chemistry* **393**: 171-175.
- Bengtsson CF, Olsen ME, Brandt LØ, Bertelsen MF, Willerslev E, et al. 2012. DNA from keratinous tissue. Part I: Hair and nail. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 17-25.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**: 16-20.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.
- Berdan EL and Fuller RC. 2012. Interspecific divergence of ionoregulatory physiology in killifish: insight into adaptation and speciation. *Journal of Zoology* **287**: 283-291.

- Besansky NJ, Powell JR, Caccone A, Hamm DM, Scott JA, et al. 1994. Molecular phylogeny of the Anopheles-Gambiae complex suggests genetic introgression between principal malaria vectors. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 6885-6888.
- Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, et al. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLOS One* **2**.
- Boessenkool S, Epp LS, Haile J, Bellemain E, Edwards M, et al. 2012. Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology* **21**: 1806-1815.
- Bollongino R, Tresset A, Vigne JD. 2008b. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* **7**: 91-98.
- Bower MA, Spencer M, Matsumura S, Nisbet RER, Howe CJ. 2005. How many clones need to be sequenced from a single forensic or ancient DNA sample in order to determine a reliable consensus sequence? *Nucleic Acids Research* **33**: 2549-2556.
- Bray SCE, Austin JJ, Metcalf JL, Østbye K, Østbye E, et al. 2013. Ancient DNA identifies post-glacial recolonisation, not recent bottlenecks, as the primary driver of contemporary mtDNA phylogeography and diversity in Scandinavian brown bears. *Diversity and Distributions* **19**: 245-256.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, et al. 2009. Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. *Science* **325**: 318-321.
- Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, et al. 2010. Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Research* **38**.

- Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, et al. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research* **35**: 5717-5728.
- Bruce KD, Hiorons WD, Hobman JL, Osborn AM, Strike P, et al. 1992. Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Applied and Environmental Microbiology* **58**: 3413-3416.
- Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, et al. 2010. Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture. *Science* **328**: 723-725.
- Campbell KL, Roberts JE, Watson LN, Stetefeld J, Sloan AM, et al. 2010. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genetics* **42**: 536-591.
- Campana MG, Bower MA, Crabtree PJ. 2013. Ancient DNA for the archaeologist: the future of African research. *African Archaeological Review* **30**: 21-37.
- Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E, et al. 2012. DNA in ancient bone - Where is it located and how should we extract it? *Annals of Anatomy-Anatomischer Anzeiger* **194**: 7-16.
- Cano RJ. 1996. Analysing ancient DNA. *Endeavour* **20**: 162-167.
- Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar GO. 1993. Amplification and sequencing of DNA from a 120-135 million-year-old weevil. *Nature* **363**: 536-538.
- Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, et al. 2013. Pulling out the 1%: whole-genome capture for the targeted enrichment of

- ancient DNA sequencing libraries. *American Journal of Human Genetics* **93**: 852-864.
- Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, et al. 2010. An Efficient Multistrategy DNA Decontamination Procedure of PCR Reagents for Hypersensitive PCR Applications. *PLOS One* **5**.
- Clark AG and Whittam TS. 1992. Sequencing errors and molecular evolutionary analysis. *Molecular Biology and Evolution* **9**: 744-752.
- Cooper A. 1994. Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthistidae). *Experientia* **50**: 558-563.
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**: 704-707.
- Cooper A, Mourer-Chauviré C, Chambers GK, von Haeseler A, Wilson AC, et al. 1992. Independent origins of New Zealand moas and kiwis. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 8741-8744.
- Cooper A and Poinar HN. 2000. Ancient DNA: Do it right or not at ALL. *Science* **289**: 1139-1139.
- Cooper A, Rhymer J, James HF, Olson SL, McIntosh CE, et al. 1996. Ancient DNA and island endemics. *Nature* **381**: 484-484.
- d'Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, et al. 2007. Molecular breeding of polymerases for amplification of ancient DNA. *Nature Biotechnology* **25**: 939-943.

- Dabney J and Meyer M. 2012. Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**: 87-94.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15758-15763.
- Dalén L, Nyström V, Valdiosera C, Germonpré M, Sablin M, et al. 2007. Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6726-6729.
- Darzynkiewicz Z, Juan G, Li X, Gorczyca W, Murakami T, et al. 1997. Cytometry in cell necrobiology: Analysis of apoptosis and accidental cell death (necrosis). *Cytometry* **27**: 1-20.
- Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, et al. 2013. *Bona fide* colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative genetics* **4**: 3-3.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194-2200.
- Eglinton G and Logan GA. 1991. Molecular preservation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **333**: 315-328.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, et al. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**: 133-138.

- Fortea J, de la Rasilla M, García-Tabernero A, Gigli E, Rosas A, et al. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidron Cave (Asturias, Spain). *Journal of Human Evolution* **55**: 353-357.
- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Gatesy SM and Dial KP. 1996. Locomotor modules and the evolution of avian flight. *Evolution* **50**: 331-340.
- Gigli E, Rasmussen M, Civit S, Rosas A, de la Rasilla M, et al. 2009. An improved PCR method for endogenous DNA retrieval in contaminated Neanderthal samples based on the use of blocking primers. *Journal of Archaeological Science* **36**: 2676-2679.
- Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, et al. 2003. Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics* **72**: 48-61.
- Gilbert MTP, Grønnow B, Andersen PK, Metspalu E, Reidla M, et al. 2008b. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**: 1787-1789.
- Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, et al. 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science* **32**: 785-793.
- Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, et al. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**: 1927-1930.
- Gilbert MTP, Wilson AS, Bunce M, Hansen AJ, Willerslev E, et al. 2004. Ancient mitochondrial DNA from hair. *Current Biology* **14**: 463-464.

- Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**: 2153-2155.
- Gloeckner C, Sauter KBM, Marx A. 2007. Evolving a thermostable DNA polymerase that amplifies from highly damaged templates. *Angewandte Chemie-International Edition* **46**: 3115-3117.
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, et al. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**: 182-189.
- Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, et al. 1990. Chloroplast DNA sequence from a Miocene magnolia species. *Nature* **344**: 656-658.
- Götherström A, Anderung C, Hellborg L, Elburg R, Smith C, et al. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society B-Biological Sciences* **272**: 2345-2350.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A Draft Sequence of the Neandertal Genome. *Science* **328**: 710-722.
- Greenwood AD, Capelli C, Possnert G, Pääbo S. 1999. Nuclear DNA sequences from late Pleistocene megafauna. *Molecular Biology and Evolution* **16**: 1466-1473.
- Gutiérrez G and Marín A. 1998. The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Molecular Biology Evolution* **15**: 926-929.

- Hagelberg E, Thomas MG, Cook CE, Sher AV, Baryshnikov GF, et al. 1994. DNA from ancient mammoth bones. *Nature* **370**: 333-334.
- Haile J, Froese DG, MacPhee RDE, Roberts RG, Arnold LJ, et al. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 22352-22357.
- Handt O, Höss M, Krings M, Pääbo S. 1994. Ancient DNA - Methodological Challenges. *Experientia* **50**: 524-529.
- Hänni C, Brousseau T, Laudet V, Stehelin D. 1995. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Research* **23**: 881-882.
- Haouchar D, Haile J, McDowell MC, Murray DC, White NE, et al. 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews* **84**: 56-64.
- Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative PCR. *Genome Research* **6**: 986-994.
- Helgason A, Pálsson S, Lalueza-Fox C, Ghosh S, Sigurdardóttir S, et al. 2007. A statistical approach to identify ancient template DNA. *Journal of Molecular Evolution* **65**: 92-102.
- Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M, et al. 2010. Road blocks on paleogenomes-polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Research* **38**.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**: 282-284.

- Ho SYW, Heupink TH, Rambaut A, Shapiro B. 2007. Bayesian estimation of sequence damage in ancient DNA. *Molecular Biology Evolution* **24**: 1416-1422.
- Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG. 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quaternary Research* **59**: 364-378.
- Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001a. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* **29**: 4793-4799.
- Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, et al. 2000. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9**: 1975-1984.
- Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. 2001b. Ancient DNA. *Nature Reviews Genetics* **2**: 353-359.
- Höss M, Dilling A, Currant A, Pääbo S. 1996a. Molecular phylogeny of the extinct ground sloth *Myiodon darwinii*. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 181-185.
- Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. 1996b. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24**: 1304-1307.
- Höss M and Pääbo S. 1993. DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research* **21**: 3913-3914.
- Huynen L, Gill BJ, Millar CD, Lambert DM. 2010. Ancient DNA reveals extreme egg morphology and nesting behavior in New Zealand's extinct moa.

*Proceedings of the National Academy of Sciences of the United States of America* **107**: 16201-16206.

Janczewski DN, Yuhki N, Gilbert DA, Jefferson GT, O'Brien SJ. 1992. Molecular phylogenetic inference from saber toothed cat fossil of Rancho-La-Brea. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 9769-9773.

Johnson. 2006. Australia's Mammal Extinctions: a 50,000 Year History. New York, Cambridge University Press.

Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, et al. 2012. A comparative study of ancient sedimentary DNA, pollen and microfossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* **21**: 1989-2003.

Kalmár T, Bachrati CZ, Marcsik A, Raskó I. 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Research* **28**.

Kemp BM, Monroe C, Smith DG. 2006. Repeat silica extraction: a simple technique for the removal of PCR inhibitors from DNA extracts. *Journal of Archaeological Science* **33**: 1680-1689.

Kermekchiev MB, Kirilova LI, Vail EE, Barnes WM. 2009. Mutants of *Taq* DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Research* **37**.

Kircher M. 2012. Analysis of High-Throughput Ancient DNA Sequencing Data. In: Shapiro B, Hofreiter M, editors. Ancient DNA: Methods and Protocols. p. 197-228.

- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage - Building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Kontanis EJ and Reed FA. 2006. Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors. *Journal of Forensic Sciences* **51**: 795-804.
- Krajewski C, Buckley L, Westerman M. 1997. DNA phylogeny of the marsupial wolf resolved. *Proceedings of the Royal Society B-Biological Sciences* **264**: 911-917.
- Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, et al. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* **439**: 724-727.
- Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, et al. 2007. The derived FOXP2 variant of modern humans was shared with neandertals. *Current Biology* **17**: 1908-1912.
- Kreader CA. 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology* **62**: 1102-1106.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, et al. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* **90**: 19-30.
- Kwok S and Higuchi R. 1989. Avoiding False Positives with PCR. *Nature* **339**: 237-238.
- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A. 2009. Bitter taste perception in Neanderthals through the analysis of the TAS2R38 gene. *Biology Letters* **5**: 809-811.

- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A, et al. 2008. Genetic characterization of the ABO blood group in Neandertals. *BMC Evolutionary Biology* **8**.
- Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* **318**: 1453-1455.
- Larson S, Jameson R, Etnier M, Fleming M, Bentzen P. 2002. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Molecular Ecology* **11**: 1899-1903.
- Lee S and Hallam SJ. 2009. Extraction of high molecular weight genomic DNA from soils and sediments. *Journal of visualized experiments* **33**: 1569.
- Lindahl T. 1993. Instability and Decay of the Primary Structure of DNA. *Nature* **362**: 709-715.
- Little DP. 2011. DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLOS One* **6**.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* **30**: 562-562.
- Longo MC, Berninger MS, Hartley JL. 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain-reactions. *Gene* **93**: 125-128.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**: 981-994.

- Mardis ER. 2008a. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133-141.
- Mardis ER. 2008b. Next-generation DNA sequencing methods. In. Annual Review of Genomics and Human Genetics **9**: 387-402.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.
- Margulies M, Jarvie TP, Knight JR, Simons JF. 2007. The 454 Life Sciences Picoliter Sequencing System. Mitchelson KR, editor. New High Throughput Technologies for DNA Sequencing and Genomics. p. 153-186.
- Maricic T, Whitten M, Pääbo S. 2010. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLOS One* **5**.
- Matisoo-Smith E, Allen JS, Ladefoged TN, Roberts RM, Lambert DM. 1997. Ancient DNA from Polynesian rats: Extraction, amplification and sequence from single small bones. *Electrophoresis* **18**: 1534-1537.
- Meyer M, Briggs AW, Maricic T, Höber B, Höffner BH, et al. 2008. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research* **36**.
- Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM. 2008. New developments in ancient genomics. *Trends in Ecology and Evolution* **23**: 386-393.
- Miller W, Drautz DI, Janecka JE, Lesk AM, Ratan A, et al. 2009. The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19**: 213-220.

- Miller W, Drautz DI, Ratan A, Pusey B, Qi J, et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**: 387-390.
- Molak M and Ho SYW. 2011. Evaluating the Impact of Post-Mortem Damage in Ancient DNA: A Theoretical Approach. *Journal of Molecular Evolution* **73**: 244-255.
- Morin PA, Chambers KE, Boesch C, Vigilant L. 2001. Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**: 1835-1844.
- Mullis KB and Faloona FA. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain-reaction. *Methods in Enzymology* **155**: 335-350.
- Murray DC, Pearson SG, Fullagar R, Chase BM, Houston J, et al. 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews* **58**: 135-145.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. 2013. Scrapheap Challenge: A novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific reports* **3**.
- Murray DC, Coghlan M, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS One* **10**: e0124671.
- Nakano M, Komatsu J, Matsuura S, Takashima K, Katsura S, et al. 2003. Single-molecule PCR using water-in-oil emulsion. *Journal of Biotechnology* **102**: 117-124.
- Nydam ML, Hoang TA, Shanley KM, De Tomaso AW. 2013. Molecular evolution of a polymorphic HSP40-like protein encoded in the histocompatibility locus

of an invertebrate chordate. *Developmental and comparative immunology* **41**: 128-136.

Nyström V, Dalén L, Vartanyan S, Lidén K, Ryman N, et al. 2010. Temporal genetic change in the last remaining population of woolly mammoth. *Proceedings of the Royal Society B-Biological Sciences* **277**: 2331-2337.

Nyström V, Humphrey J, Skoglund P, McKeown NJ, Vartanyan S, et al. 2012. Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Molecular Ecology* **21**: 3391-3402.

Oskam CL, Allentoft ME, Walter R, Scofield RP, Haile J, et al. 2012. Ancient DNA analyses of early archaeological sites in New Zealand reveal extreme exploitation of moa (Aves: Dinornithiformes) at all life stages. *Quaternary Science Reviews* **52**: 41-48.

Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, et al. 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **277**: 1991-2000.

Oskam CL, Jacomb C, Allentoft ME, Walter R, Scofield RP, et al. 2011. Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *Journal of Archaeological Science* **38**: 2589-2595.

Overballe-Petersen S, Orlando L, Willerslev E. 2012. Next-generation sequencing offers new insights into DNA degradation. *Trends in Biotechnology* **30**: 364-368.

Ozawa T, Hayashi S, Mikhelson VM. 1997. Phylogenetic position of Mammoth and Steller's sea cow within Tethytheria demonstrated by mitochondrial DNA sequences. *Journal of Molecular Evolution* **44**: 406-413.

- Pääbo S. 1989. Ancient DNA - extraction, charactersiation, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 1939-1943.
- Pääbo S, Higuchi RG, Wilson AC. 1989. Ancient DNA and the polymerase chain-reaction - the emerging field of molecular archaeology. *Journal of Biological Chemistry* **264**: 9709-9712.
- Pääbo S and Wilson AC. 1988. Polymerase Chain-Reaction Reveals Cloning Artifacts. *Nature* **334**: 387-388.
- Palmer S, Smith O, Allaby RG. 2012. The blossoming of plant archaeogenetics. *Annals of Anatomy* **194**:146-156.
- Pareek CS, Smoczynski R, Tretyn A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**: 413-435.
- Phillips MJ, Gibb GC, Crimp EA, Penny D. 2010. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Systematic Biology* **59**: 90-107.
- Poinar H, Kuch M, McDonald G, Martin P, Pääbo S. 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. *Current Biology* **13**: 1150-1152.
- Poinar HN, Höss M, Bada JL, Pääbo S. 1996. Amino acid racemization and the preservation of ancient DNA. *Science* **272**: 864-866.
- Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, et al. 2001. A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 4317-4322.

- Poinar HN and Stankiewicz BA. 1999. Protein preservation and DNA retrieval from ancient tissues. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 8426-8431.
- Pruvost M and Geigl EM. 2004. Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *Journal of Archaeological Science* **31**: 1191-1197.
- Pruvost M, Grange T, Geigl EM. 2005. Minimizing DNA contamination by using UNG-coupled quantitative real-time PCR on degraded DNA samples: application to ancient DNA studies. *Biotechniques* **38**: 569-575.
- Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, et al. 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 739-744.
- Pruvost M, Schwarz R, Correia VB, Champlot S, Grange T, et al. 2008. DNA diagenesis and palaeogenetic analysis: Critical assessment and methodological progress. *Palaeogeography Palaeoclimatology Palaeoecology* **266**: 211-219.
- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**.
- Ramakrishnan U and Hadly EA. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology* **18**: 1310-1330.
- Rambaut A, Ho SYW, Drummond AJ, Shapiro B. 2009. Accommodating the Effect of Ancient DNA Damage on Inferences of Demographic Histories. *Molecular Biology and Evolution* **26**: 245-248.

- Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, et al. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334**: 94-98.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**: 757-762.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A. 2009. DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B-Biological Sciences* **276**: 3395-3402.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**: 1053-1060.
- Rodríguez-Aranda JP and Calvo JP. 1998. Trace fossils and rhizoliths as a tool for sedimentological and palaeoenvironmental analysis of ancient continental evaporite successions. *Palaeogeography Palaeoclimatology Palaeoecology* **140**: 383-399.
- Rohland N and Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. *Nature Protocols* **2**: 1756-1762.
- Rohland N and Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**: 343-352.
- Rohland N, Reich D, Mallick S, Meyer M, Green RE, et al. 2010. Genomic DNA Sequences from Mastodon and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants. *PLOS Biology* **8**.
- Römpler H, Rohland N, Lalueza-Fox C, Willerslev E, Kuznetsova Tet al. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**: 62-62.

- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.
- Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS One* **7**.
- Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, et al. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**.
- Shapiro B and Hofreiter M. 2012. Ancient DNA: methods and protocols. New York Humana Press (Springer Science).
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**: 1794-1805.
- Soltis PS and Soltis DE. 1993. Ancient DNA - prospects and limitations. *New Zealand Journal of Botany* **31**: 203-209.
- Sorenson MD, Cooper A, Paxinos EE, Quinn TW, James HF, et al. 1999. Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **266**: 2187-2193.
- Stiller M and Fulton TL. 2012. Multiplex PCR Amplification of Ancient DNA. In: Shapiro B, Hofreiter M, editors. *Ancient DNA: Methods and Protocols*. p. 133-141.

- Stiller M, Knapp M, Stenzel U, Hofreiter M, Meyer M. 2009. Direct multiplex sequencing (DMPS)-a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Research* **19**: 1843-1848.
- Stoneking M. 1995. Ancient DNA - How do you know when you have it and what can you do with it. *American Journal of Human Genetics* **57**: 1259-1262.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**: 2045-2050.
- Taylor PG. 1996. Reproducibility of ancient DNA sequences from extinct pleistocene fauna. *Molecular Biology Evolution* **13**: 283-285.
- Thomas RH, Schaffner W, Wilson AC, Pääbo S. 1989. DNA Phylogeny of the Extinct Marsupial Wolf. *Nature* **340**: 465-467.
- Thomas WK, Pääbo S, Villablanca FX, Wilson AC. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *Journal of Molecular Evolution* **31**: 101-112.
- Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, et al. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**: 1088-1091.
- Tsai YL and Olson B. 1992. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology* **58**: 2292-2295.
- Vasan S, Zhang X, Zhang XN, Kapurniotu A, Bernhagen J, et al. 1996. An agent cleaving glucose-derived protein crosslinks *in vitro* and *in vivo*. *Nature* **382**: 275-278.
- Vilà C, Leonard JA, Götherström A, Marklund S, Sandberg K, et al. 2001. Widespread origins of domestic horse lineages. *Science* **291**: 474-477.

- Wandeler P, Smith S, Morin PA, Pettifor RA, Funk SM. 2003. Patterns of nuclear DNA degeneration over time - a case study in historic teeth samples. *Molecular Ecology* **12**: 1087-1093.
- Watson RJ and Blackwell B. 2000. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Canadian Journal of Microbiology* **46**: 633-642.
- White RA III, Blainey PC, Fan HC, Quake SR. 2009. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10**.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795.
- Wilson IG. 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology* **63**: 3741-3751.
- Wood JR, Rawlence NJ, Rogers GM, Austin JJ, Worthy TH, et al. 2008. Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quaternary Science Reviews* **27**: 2593-2602.
- Woodward SR, Weyand NJ, Bunnell M. 1994. DNA sequence from cretaceous period bone fragments. *Science* **266**: 1229-1232.
- Yang DY, Eng B, Wayne JS, Dudar JC, Saunders SR. 1998. Technical note: Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology* **105**: 539-543.

- Yang H, Golenberg EM, Shoshani J. 1996. Phylogenetic resolution within the elephantidae using fossil DNA sequence from the American mastodon (*Mammut americanum*) as an outgroup. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 1190-1194.
- Ye J, Ji AQ, Parra EJ, Zheng XF, Jiang CT, et al. 2004. A simple and efficient method for extracting DNA from old and burned bone. *Journal of Forensic Sciences* **49**: 754-759.
- Yen C and Yang JL. 2009. Historical review and prospect of taxonomy of tribe Triticeae Dumortier (Poaceae). *Breeding Science* **59**: 513-518.
- Zhang X, Shi GR, Gong Y. 2008. Middle jurassic trace fossils from the ridang formation in sajia county, south Tibet, and their palaeoenvironmental significance. *Facies* **54**: 45-60.
- Zhang Z, Kermekchiev MB, Barnes WM. 2010. Direct DNA amplification from crude clinical samples using a PCR enhancer cocktail and novel mutants of *Taq*. *The Journal of Molecular Diagnostics* **12**: 152-161.
- Zheng Z, Advani A, Melefors O, Glavas S, Nordstrom H, et al. 2010. Titration-free massively parallel pyrosequencing using trace amounts of starting material. *Nucleic Acids Research* **38**.
- Zischler H, Höss M, Handt O, Vonhaeseler A, van der Kuyl AC, et al. 1995. Detecting dinosaur DNA. *Science* **268**: 1192-1193.
- Zou YP, Yin Z, Fang XS, Li GL, Zhang Y. 1995. Ancient DNA in Late Cretaceous dinosaur eggs from Xixia County, Henan providence. *Chinese Science Bulletin* **40**: 856-860.

## 1.6 SUPPLEMENTARY INFORMATION

---

**TABLE S1.6.1** | SUMMARY OF THE TYPES OF ADNA STUDIES UNDERTAKEN BETWEEN 2000 AND 2013. Asterisked references are good examples of the type. Studies marked by ^ indicate that the DNA was extracted from a warm-climate ecosystem. Adapted from (Ramakrishnan and Hadly 2009) (modified, updated, and extended).

Topic of investigation	Organism	Substrate	DNA	Sequence	Reference
Population history, conservation	Otter ( <i>Enhydra lutris</i> )	Bone	nuDNA, mtDNA	Amplicon	(Larson et al. 2002; Valentine et al. 2008)
	Salmon ( <i>Salmo salar</i> )	Bone	mtDNA	Amplicon	(Consuegra et al. 2002)
	Beach tiger beetle ( <i>Cincindela d. dorsalis</i> )	Museum specimens	mtDNA	Amplicon	(Goldstein and Desalle 2003)
	Scots pine ( <i>Picea sylvestris</i> ), Norway spruce ( <i>Picea abies L.</i> )	Pollen	clDNA	Amplicon	(Parducci et al. 2005)
	Brown bears ( <i>Ursus arctos</i> )	Bone	mtDNA	Amplicon	(Leonard et al. 2000; Barnes et al. 2002; Calvignac et al. 2008; Valdiosera et al. 2008;

					Bray et al. 2013)
	Brown bears ( <i>Ursus arctos</i> )	Bone	nuDNA mtDNA	Amplicon	(Miller and Waits 2003)
	Bowhead whales ( <i>Balaena mysticetus</i> )	Bone	mtDNA	Amplicon	(Borge et al. 2007)
	Arctic fox ( <i>Alopex lagopus</i> )	Bone	mtDNA	Amplicon	(Dalén, et al. 2007)
	Ancient kiwi ( <i>Apteryx</i> sp.)	Museum specimens	mtDNA	Amplicon	(Shepherd and Lambert 2008)
	Caspian tiger ( <i>Panthera tigris virgate</i> )	Museum specimens	mtDNA	Amplicon	(Driscoll et al. 2009)
	California Tule Elk ( <i>Cervus elaphus nannodes</i> )	Bone	mtDNA	Amplicon	(Broughton et al. 2013)
	Depuch Island rock wallaby ( <i>Petrogale</i> sp.)	Bone	mtDNA	Amplicon	(Haouchar et al. 2013)^
Population history, human	Neolithic humans ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Haak et al. 2005)
	Andaman Islanders ( <i>Homo sapiens</i> )	Teeth	mtDNA	Amplicon	(Endicott et al. 2006)^

Sytho-Siberian ( <i>Homo sapiens</i> )	Bone	nuDNA mtDNA	Amplicon	(Ricaud et al. 2003)
Ancient Nubians ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Lalueza Fox 1997)^
Egyptian mummies ( <i>Homo sapiens</i> )	Bone	nuDNA	Amplicon	(Hawass et al. 2010, 2012)^
Auersperg Tomb ( <i>Homo sapiens</i> )	Bone, teeth	nuDNA mtDNA	Amplicon	(Pajnic et al. 2013)
Tianyuan Cave, China ( <i>Homo sapiens</i> )	Bone	mtDNA nuDNA	Shotgun	(Fu et al. 2013)
Ancient Ilberians ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Sampietro et al. 2005)^
Aboriginal Australians ( <i>Homo sapiens</i> )	Hair	nuDNA	Whole genome	(Rasmussen et al. 2011)^
Aboriginal Australians ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Adcock et al. 2001)^
Cro-Magnon ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Caramelli et al. 2003; Caramelli et al. 2008)
Andaman Islanders ( <i>Homo sapiens</i> )	Bone, teeth	mtDNA	Amplicon	(Endicott et al. 2003)^
Etruscans ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Vernesi et al. 2004)
Ancient Sardinians	Tooth	mtDNA	Amplicon	(Caramelli et al. 2007)

<i>(Homo sapiens)</i>					
	Tyrolean Iceman <i>(Homo sapiens)</i>	Tissue	mtDNA	Whole genome	(Ermini et al. 2008)
	Ancient eskimo <i>(Homo sapiens)</i>	Hair	mtDNA	Whole genome	(Gilbert et al. 2008b)
	European hunter-gatherers <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Bramanti et al. 2009)
	Medieval Tuscans <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Guimaraes et al. 2009)
	Neolithic farmer <i>(Homo sapiens)</i>	Bone	mtDNA nuDNA	Amplicon	(Haak et al. 2010)
	Nagano, Japan <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Adachi et al. 2013)
	North East Europeans <i>(Homo sapiens)</i>	Bone, tooth	mtDNA	Amplicon	(Sarkissian et al. 2013)
	Sanganji, Japan <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Kanzawa-Kiryama et al. 2013)
Population history, domestication	Wild aurochsen <i>(Bos primigenius)</i>	Bone, teeth	mtDNA	Amplicon	(Troy et al. 2001; Anderung et al. 2005; Beja-Pereira et al. 2006) (Edwards et al. 2004; Edwards et al. 2007; Mona

					et al. 2010)
Aurochs ( <i>Bos primigenius</i> )	Bone	nuDNA	Amplicon	(Götherström et al. 2005; Bollongino et al. 2008a)	
Wild aurochsen ( <i>Bos primigenius</i> )	Bone	mtDNA	Amplicon, Whole genome	(Achilli et al. 2008; Lari et al. 2011; Zeyland et al. 2013)	
Neolithic cattle ( <i>Bos Taurus</i> )	Bone	mtDNA	Amplicon	(Bollongino et al. 2006)	
African cattle	Bone, teeth	mtDNA	Amplicon	(Anderung et al. 2005; Ascunce et al. 2007; Bradely et al. 1996)^	
African wild ass ( <i>Equus africanus</i> )	Bone	mtDNA	Amplicon	(Kimura et al. 2011, 2013)^	
Domestic horse ( <i>Equus sp.</i> )	Bone	nuDNA	Amplicon	(Lippold et al. 2011)	
Sheep and goat	Bone	mtDNA	Amplicon	(Loreille et al. 1997)	
Wheat	Seeds	nuDNA	Amplicon	(Oliveira et al. 2012)	
Nubian barley ( <i>Hordeum vulgare L.</i> )	Seeds	nuDNA	Amplicon	(Palmer et al. 2009)^	
Pig ( <i>Sus scrofa</i> )	Bone	mtDNA	Amplicon	(Larson et al. 2007)	
Pig ( <i>Sus scrofa</i> )	Museum specimens	mtDNA	Amplicon	(Larson et al. 2005)	

	Canids	Bone, teeth	mtDNA	Amplicon	(Byrd et al. 2013)
	Southern Levantine Pigs ( <i>Sus scrofa</i> )	Bone	mtDNA	Amplicon	(Meiri et al. 2013)
	Horse ( <i>Equus</i> sp.)	Bone, teeth	mtDNA	Amplicon	(Cieslak et al. 2010)
Population history, phylogeography, evolutionary processes	Mammoth ( <i>Mammuthus primigenius</i> )	?	mtDNA	Amplicon	(Debruyne et al. 2008)
	Mammoth ( <i>Mammuthus primigenius</i> )	Bone, tooth or tusk	mtDNA	Amplicon	(Barnes et al. 2007; Nyström et al. 2010; Palkopoulou et al. 2013)
	Mammoth ( <i>Mammuthus primigenius</i> )	Bone, tooth or tusk	nuDNA, mtDNA	Amplicon	(Nyström et al. 2012)
	Mammoth ( <i>Mammuthus primigenius</i> )	Hair	mtDNA	Whole genome	(Gilbert et al. 2008a)
	American megafauna sp.	Permafrost core	mtDNA	Amplicon	(Haile et al. 2009)*
	Pleistocene fauna	Bones, teeth	mtDNA	Amplicon	(Hofreiter et al. 2004)
	Cave bear	Bone , tooth	mtDNA	Amplicon	(Hofreiter et al. 2002;

	<i>(Ursus spelaeus)</i>				Orlando et al. 2002; Hofreiter et al. 2007; Stiller et al. 2010)
	Moa sp.	Bone	mtDNA	Amplicon	(Bunce et al. 2009) (McCallum et al. 2013)
	Denisovan <i>(Homo sp.)</i>	Bone	nuDNA	Whole genome	(Reich et al. 2010)
	Neanderthal <i>(Homo neanderthalensis)</i>	Tooth	mtDNA	Amplicon	(Lalueza-Fox et al. 2005; Orlando et al. 2006)
	Neanderthal <i>(Homo neanderthalensis)</i>	Bone, tooth	nuDNA	Shotgun	(Green et al. 2006) (Noonan et al. 2006)
	Neanderthal <i>(Homo neanderthalensis)</i>	Bone	nuDNA	Whole genome	(Green et al. 2010)
	Ancient muskox <i>(Ovibos moschatus)</i>	Bone	mtDNA	Amplicon	(MacPhee et al. 2005)
	Adelie penguin <i>(Pygoscelis adeliae)</i>	Bone	nuDNA	Amplicon	(Shepherd et al. 2005)
	Ancient dog	Bone?	mtDNA	Amplicon	(Leonard et al. 2002)

(*Canis lupus*)

	Ancient Maize ( <i>Zea mays</i> )	Cob	nuDNA	Amplicon	(Jaenicke-Després et al. 2003)
	Rodent ( <i>Ctenomys sociabilis</i> )	Bone, tooth	mtDNA	Amplicon	(Chan et al. 2006)
	Bison ( <i>Bison bison</i> )	Bone	mtDNA	Amplicon	(Shapiro et al. 2004)
	Brown bear ( <i>Ursus arctos</i> )	Bone, teeth	mtDNA	Amplicon	(Valdiosera et al. 2007)
	Atlantic surgeon ( <i>Acipenser</i> spp.)	Museum specimens	mtDNA	Amplicon	(Chassaing et al. 2013)
	Richardson's collared lemming ( <i>Dicrostonyx richardsoni</i> )	Bone, teeth	mtDNA	Amplicon	(Fulton et al. 2013)
	Blue antelope ( <i>Hippotragus leucophaeus</i> )	Bone	mtDNA	Amplicon	(Robinson et al. 1996)^
	Norway Spruce ( <i>Picea abies</i> )	Sediment	clDNA	Amplicon	(Magyari et al. 2011)
Phylogeny, time of divergence	Moa sp.	Bone, soft tissue	mtDNA	Amplicon	(Cooper et al. 1992)
	Moa sp.	Bone	mtDNA	Amplicon	(Baker et al. 2005)
	Moa sp.	Bone	mtDNA	Whole genome	(Cooper et al. 2001)

Ratite sp. including Moa sp.	Bone	mtDNA	Whole genome	(Haddrath and Baker 2001)
Adelie penguin ( <i>Pygoscelis adeliae</i> )	Bone	mtDNA	Amplicon	(Lambert et al. 2002; Ritchie et al. 2004)
Quagga ( <i>Equus quagga</i> )	Museum specimens	mtDNA	Amplicon	(Leonard et al. 2005)
Ancient canids ( <i>Canis</i> sp.)	Bone	mtDNA	Amplicon	(Verginelli et al. 2005)
Altai Canid ( <i>Canis</i> sp.)	Bone	mtDNA	Amplicon	(Druzhkova et al. 2013)
Neanderthal ( <i>Homo neanderthalensis</i> )	Bone	mtDNA	Amplicon	(Caramelli et al. 2006)
Mammoth ( <i>Mammuthus primigenious</i> )	Bone	mtDNA	Whole genome	(Krause et al. 2006)
Mammoth ( <i>Mammuthus primigenious</i> )	Bone	nuDNA	Shotgun	(Poinar et al. 2006; Miller et al. 2008)
Mammoth	Tooth	nuDNA	Shotgun	(Rohland et al. 2010)

(*Mammuthus primigenius*),  
Mastodon  
(*Mammut americanum*)

Columbian mammoth ( <i>Mammuthus columbi</i> )	Bone, tusk	mtDNA	Whole genome	(Enk et al. 2011)
Ancient gray wolf ( <i>Canis lupus</i> )	Bone	mtDNA	Amplicon	(Leonard et al. 2007)
Bear sp.	Bone	mtDNA	Whole genome	(Krause et al. 2008)
Archaeolemur sp.	Bone, teeth	mtDNA	Amplicon	(Orlando et al. 2008)^
Thylacine ( <i>Thylacinus cynocephalus</i> )	Museum specimens	mtDNA	Whole genome	(Miller et al. 2009)^
New Zealand quail ( <i>Coturnix novaezealandiae</i> )	Museum specimens	mtDNA	Amplicon	(Seabrook-Davison et al. 2009)
Denisovan	Bone	mtDNA	Whole	(Krause et al. 2010)

	( <i>Homo</i> sp.?)			genome	
	Polar bear ( <i>Ursus maritimus</i> )	Bone, tooth	mtDNA	Whole genome	(Lindqvist et al. 2010)
	Shasta sloth ( <i>Nothrotherops shastensis</i> )	Coprolite	nuDNA	Amplicon	(Poinar et al. 2003)
	Cave bear ( <i>Ursus spelaeus</i> )	Bone	mtDNA	Amplicon	(Loreille et al. 2001)
	Cave bear ( <i>Ursus spelaeus</i> )	Bone	nuDNA, mtDNA	Shotgun	(Noonan et al. 2005)
	Lemur ( <i>Megaladapsis edwardsi</i> )	Bone	mtDNA	Amplicon	(Montagnon et al. 2001; Karanth et al. 2005)^
	Moa sp.	Bone	nuDNA	Amplicon	(Bunce et al. 2003)
Taxonomy	Moa sp.	Bone	nuDNA	Amplicon	(Huynen et al. 2003)
	Tasman booby ( <i>Sula tasmani</i> )	Bone	mtDNA	Amplicon	(Steeves et al. 2010)
Palaeodiet,	Ground sloth	Coprolite	mtDNA	Amplicon	(Hofreiter et al. 2000)

palaeoenvironment	<i>(Nothotherips shastensis)</i>		cDNA		
	Ground sloth <i>(Lagidium sp.)</i>	Coprolite	mtDNA cDNA	Amplicon	(Hofreiter et al. 2003)
	Moa sp.	Coprolite	mtDNA	Amplicon	(Wood et al. 2008; Wood et al. 2012; Wood et al. 2013)
	Ancient native Americans <i>(Homo sapiens)</i>	Coprolite	mtDNA cDNA	Amplicon	(Poinar et al. 2001)
Palaeoenvironment	Megafauna sp., plant sp.	Permafrost core	mtDNA cDNA	Amplicon	(Willerslev et al. 2003)
	Plant sp.	Permafrost core	mtDNA cDNA	Amplicon	(Jørgensen et al. 2012)
	Fungi	Permafrost core	cDNA	Amplicon	(Kochkina et al. 2012)
	Foraminifera and Radiolaria	Sea sediments	mtDNA	Amplicon	(Lejzerowicz et al. 2013)
	Plant sp.	Lake sediments	cDNA	Amplicon	(Pedersen et al. 2013; Parducci et al. 2013)
	Plant sp., bacteria	Permafrost core	mtDNA cDNA	Amplicon	(Porter et al. 2013)
	Fungi	Permafrost core	ITS	Amplicon	(Bellemain et al. 2013)

Phenotype, adaptation	Neanderthal ( <i>Homo neanderthalensis</i> )	Bone	nuDNA	Amplicon	(Krause et al. 2007) (Lalueza-Fox et al. 2007; Lalueza-Fox et al. 2008; Lalueza-Fox et al. 2009)
	Ancient eskimo ( <i>Homo sapiens</i> )	Hair	nuDNA, mtDNA	Whole genome	(Rasmussen et al. 2010)
	Mammoth ( <i>Mammuthus primigenius</i> )	Bone	nuDNA	Amplicon	(Römpler et al. 2006; Campbell et al. 2010)
	Moa sp.	Feathers	mtDNA	Amplicon	(Rawlence et al. 2009)
	Moa sp.	Eggshell	mtDNA	Amplicon	(Huynen et al. 2010)
	Ancient canids ( <i>Canis</i> spp.)	Bone	nuDNA	Amplicon	(Ollivier et al. 2013)
	Pre-domestic horse ( <i>Equus</i> sp.)	Bone	nuDNA	Amplicon	(Ludwig et al. 2009)
	Human ( <i>Homo sapiens</i> )	Bone	nuDNA	Amplicon	(Draus-Barini et al. 2013)

### **S1.6.1 SUPPLEMENTARY REFERENCES**

- Achilli A, Olivieri A, Pellecchia M, Uboldi C, Colli L, et al. 2008. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology* **18**: 157-158.
- Adachi N, Sawada J, Yoneda M, Kobayashi K, Iroh S. 2013. Mitochondrial DNA analysis of the human skeleton of the initial Jomon phase excavated at the Yugura cave site, Nagano, Japan. *Anthropological Science* **121**: 137-143.
- Adcock GJ, Dennis ES, Eastaugh S, Huttley GA, Jeremiin LS, et al. 2001. Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 537-542.
- Anderung C, Bouwman A, Persson P, Carretero JM, Ortega AI, et al. 2005. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 8431-8435.
- Ascunce MS, Kithen A, Schmidt PR, Miyamoto MM, Mulligan CJ. 2007. An unusual pattern of ancient mitochondrial DNA haplogroups in northern African cattle. *Zoological Studies* **46**: 123-125.
- Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: The giant moas of New Zealand. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 8257-8262.
- Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A. 2002. Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**: 2267-2270.

- Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, et al. 2007. Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology* **17**: 1072-1075.
- Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, et al. 2006. The origin of European cattle: Evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 8113-8118.
- Bellemain E, Davey ML, Kauserud H, Epp LS, Boessenkool S, et al. 2013. Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. *Environmental Microbiology* **15**: 1176-1189.
- Bollongino R, Edwards CJ, Alt KW, Burger J, Bradley DG. 2006. Early history of European domestic cattle as revealed by ancient DNA. *Biology Letters* **2**: 155-159.
- Bollongino R, Elsner J, Vigne JD, Burger J. 2008a. Y-SNPs Do Not Indicate Hybridisation between European Aurochs and Domestic Cattle. *PLOS One* **3**.
- Borge T, Bachmann L, Bjørnstad G, Wiig O. 2007. Genetic variation in Holocene bowhead whales from Svalbard. *Molecular Ecology* **16**: 2223-2235.
- Bradley DG, MacHugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 5131-5135.
- Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, et al. 2009. Genetic Discontinuity Between Local Hunter-Gatherers and Central Europe's First Farmers. *Science* **326**: 137-140.
- Bray SCE, Austin JJ, Metcalf JL, Østbye K, Østbye E, et al. 2013. Ancient DNA identifies post-glacial recolonisation, not recent bottlenecks, as the primary

driver of contemporary mtDNA phylogeography and diversity in Scandinavian brown bears. *Diversity and Distributions* **19**: 245-256.

Broughton JM, Beck RK, Coltrain JB, Rourke DH, Rogers AR. 2013. A Late Holocene population bottleneck in the California Tule Elk (*Cervus elaphus nannodes*): provisional support from ancient DNA. *Journal of Archaeological Method and Theory* **20**: 495-524.

Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, et al. 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**: 172-175.

Bunce M, Worthy TH, Phillips MJ, Holdaway RN, Willerslev E, et al. 2009. The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 20646-20651.

Byrd BE, Cornellas A, Eerkens JW, Rosenthal JS, Carpenter TR, et al. 2013. The role of canids in ritual and domestic contexts: new ancient DNA insights from complex hunter-gatherer sites in prehistoric central California. *Journal of Archaeological Science* **40**: 2176-2189.

Calvignac S, Hughes S, Tougaard C, Michaux J, Thevenot M, et al. 2008. Ancient DNA evidence for the loss of a highly divergent brown bear clade during historical times. *Molecular Ecology* **17**: 1962-1970.

Campbell KL, Roberts JE, Watson LN, Stetefeld J, Sloan AM, et al. 2010. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genetics* **42**: 536-591.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, et al. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 6593-6597.

- Caramelli D, Lalueza-Fox C, Condemi S, Longo L, Milani L, et al. 2006. A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Current Biology* **16**: 630-632.
- Caramelli D, Vernesi C, Sanna S, et al. 2007. Genetic variation in prehistoric Sardinia. *Human Genetics* **122**: 327-336.
- Caramelli D, Milani L, Vai S, Modi A, Pecchioli E, et al. 2008. A 28,000 Years Old Cro-Magnon mtDNA Sequence Differs from All Potentially Contaminating Modern Sequences. *PLOS One* **3**.
- Chan YL, Anderson CNK, Hadly EA. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLOS Genetics* **2**: 451-460.
- Chassaing O, Desse-Berset N, Duffraisse M, Hughes S, Hanni C, et al. 2013. Paleogenetics of western French sturgeons spotlights the relationships between *Acipenser sturio* and *Acipenser oxyrinchus*. *Journal of Biogeography* **40**: 382-393.
- Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, et al. 2010. Origin and History of Mitochondrial DNA Lineages in Domestic Horses. *PLOS One* **5**.
- Consuegra S, García de Leániz C, Serdio A, González\_ Morales M, Straus LG, et al. 2002. Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology* **11**: 2037-2048.
- Cooper A, Mourer-Chauviré C, Chambers GK, von Haeseler A, Wilson AC, et al. 1992. Independent origins of New Zealand moas and kiwis. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 8741-8744.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**: 704-707.

Dalén L, Nyström V, Valdiosera C, Germonpré M, Sablin M, et al. 2007. Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6726-6729.

Debruyne R, Chu G, King CE, Bos K, Kuch M, et al. 2008. Out of America: Ancient DNA evidence for a new world origin of late quaternary woolly mammoths. *Current Biology* **18**: 1320-1326.

Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, et al. 2013. *Bona fide* colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative genetics* **4**: 3-3.

Driscoll CA, Yamaguchi N, Bar-Gal GK, Roca AL, Luo S, et al. 2009. Mitochondrial Phylogeography Illuminates the Origin of the Extinct Caspian Tiger and Its Relationship to the Amur Tiger. *PLOS One* **4**.

Druzhkova AS, Thalmann O, Trifonov VA, Leonard JA, Vorobieva NV, et al. 2013. Ancient DNA analysis affirms the candid from Altai as a primitive dog. *PLOS One* **8**.

Edwards CJ, MacHugh DE, Dobney KM, Martin L, Russell N, et al. 2004. Ancient DNA analysis of 101 cattle remains: limits and prospects. *Journal of Archaeological Science* **31**: 695-710.

Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, et al. 2007. Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society B-Biological Sciences* **274**: 1377-1385.

- Endicott P, Gilbert MTP, Stringer C, Lalueza-Fox C, Willerslev E, et al. 2003. The genetic origins of the Andaman Islanders. *American Journal of Human Genetics* **72**: 178-184.
- Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, et al. 2006. Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. *Plos One* **1**: e81.
- Enk J, Devault A, Debruyne R, King CE, Treangen T, et al. 2011. Complete Columbian mammoth mitogenome suggests interbreeding with woolly mammoths. *Genome Biology* **12**.
- Ermini L, Olivieri C, Rizzi E, Corti G, Bonnal R, et al. 2008. Complete Mitochondrial Genome Sequence of the Tyrolean Iceman. *Current Biology* **18**: 1687-1693.
- Fu QM, Meyer M, Gao X, Stenzel U, Burbano HA, et al. 2013. DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 2223-2227.
- Fulton TL, Norris RW, Graham RW, Semken HA, Shapiro B. 2013. Ancient DNA supports southern survival of Richardson's collared lemming (*Dicrostonyx richardsoni*) during the last glacial maximum. *Molecular Ecology* **22**: 2540-2548.
- Gilbert MTP, Drautz DI, Lesk AM, Ho SY, Qi J, et al. 2008a. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 8327-8332.
- Gilbert MTP, Grønnow B, Andersen PK, Metspalu E, Reidla M, et al. 2008b. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**: 1787-1789.

- Goldstein PZ and Desalle R. 2003. Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Molecular Ecology* **12**: 1993-1998.
- Götherström A, Anderung C, Hellborg L, Elburg R, Smith C, et al. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society B-Biological Sciences* **272**: 2345-2350.
- Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**: 330-336.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, et al. 2010. A Draft Sequence of the Neandertal Genome. *Science* **328**: 710-722.
- Guimaraes S, Ghirotto S, Benazzo A, Milani L, Lari M, et al. 2009. Genealogical Discontinuities among Etruscan, Medieval, and Contemporary Tuscans. *Molecular Biology and Evolution* **26**: 2157-2166.
- Haak W, Forster P, Bramanti B, Matsumura S, Brandt G, et al. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* **310**: 1016-1018.
- Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, et al. 2010. Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLOS Biology* **8**.
- Haddrath O and Baker AJ. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society B-Biological Sciences* **268**: 939-945.

- Haile J, Froese DG, MacPhee RDE, Roberts RG, Arnold LJ, et al. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 22352-22357.
- Haouchar D, Haile J, Spencer PBS, Bunce M. 2013. The identity of the Depuch Island rock wallaby revealed through ancient DNA. *Australian Mammalogy* **35**: 101-106.
- Hawass Z, Gad YZ, Ismail S, Khairat R, Fathalla D, et al. 2010. Ancestry and pathology in King Tutankhamun's family. *Journal of American Medical Association* **303**: 638-647.
- Hofreiter M, Poinar HN, Spaulding WG, Bauer K, Martin PS, et al. 2000. A molecular analysis of ground sloth diet through the last glaciation. *Molecular Ecology* **9**: 1975-1984.
- Hofreiter M, Capelli C, Krings M, Waits L, Conard N, et al. 2002. Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. *Molecular Biology Evolution* **19**: 1244-1250.
- Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG. 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quaternary Research* **59**: 364-378.
- Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, et al. 2004. Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 12963-12968.
- Hofreiter M, Muenzel S, Conard NJ, Pollack J, Slatkin M, et al. 2007. Sudden replacement of cave bear mitochondrial DNA in the late Pleistocene. *Current Biology* **17**: R122-R123.

- Huynen L, Millar CD, Scofield RP, Lambert DM. 2003. Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**: 175-178.
- Huynen L, Gill BJ, Millar CD, Lambert DM. 2010. Ancient DNA reveals extreme egg morphology and nesting behavior in New Zealand's extinct moa. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 16201-16206.
- Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, et al. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206-1208.
- Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, et al. 2012. A comparative study of ancient sedimentary DNA, pollen and microfossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* **21**: 1989-2003.
- Kanzawa-Kiriyama H, Saso A, Suwa G, Saitou N. 2013. Ancient mitochondrial DNA sequences of Jomon teeth samples from Sanganjū Tohoku district, Japan. *Anthropological Science* **121**: 89-103.
- Karanth KP, Delefosse T, Rakotosamimanana B, Parsons TJ, Yoder AD. 2005. Ancient DNA from giant extinct lemurs confirms single origin of Malagasy primates. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 5090-5095.
- Kimura B, Marshall FB, Chen S, Rosenbom S, Moehlman PD, et al. 2011. Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B: Biological Sciences* **278**: 50-57.
- Kimura B, Marshall F, Beja-Pereira A, Mulligan C. 2013. Donkey domestication. *African Archaeological Review* **30**: 1.

- Kochkina G, Ivanushkina N, Ozerskaya S, Chigineva N, Vasilenko O, et al. 2012. Ancient fungi in Antarctic permafrost environments. *Fems Microbiology Ecology* **82**: 501-509.
- Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, et al. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* **439**: 724-727.
- Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, et al. 2007. The derived FOXP2 variant of modern humans was shared with neandertals. *Current Biology* **17**: 1908-1912.
- Krause J, Unger T, Noçon A, Malaspinas AS, Kolokotronis SO, et al. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology* **8**.
- Krause J, Fu Q, Good JM, Viola B, Shunkov MV et al. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**: 894-897.
- Lalueza-Fox C. 1997. MtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Saharan and North Africa Nile valley. *Annals of Human Biology* **24**: 217-227.
- Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, et al. 2005. Neandertal evolutionary genetics: Mitochondrial DNA data from the Iberian Peninsula. *Molecular Biology and Evolution* **22**: 1077-1081.
- Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* **318**: 1453-1455.

- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A, et al. 2008. Genetic characterization of the ABO blood group in Neandertals. *BMC Evolutionary Biology* **8**.
- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, Rosas A. 2009. Bitter taste perception in Neanderthals through the analysis of the TAS2R38 gene. *Biology Letters* **5**: 809-811.
- Lambert DM, Ritchie PA, Millar CD, Holland B, Drummond AJ, et al. 2002. Rates of evolution in ancient DNA from Adelie penguins. *Science* **295**: 2270-2273.
- Lari M, Rizzi E, Mona S, Corti G, Catalano G, et al. 2011. The Complete Mitochondrial Genome of an 11,450-year-old Aurochs (*Bos primigenius*) from Central Italy. *BMC Evolutionary Biology* **11**.
- Larson S, Jameson R, Etnier M, Fleming M, Bentzen P. 2002. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Molecular Ecology* **11**: 1899-1903.
- Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, et al. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**: 1618-1621.
- Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, et al. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 15276-15281.
- Lejzerowicz F, Esling P, Majewski W, Szczuciński W, Decelle J, et al. 2013. Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology Letters* **9**.

- Leonard JA, Wayne RK, Cooper A. 2000. Population genetics of Ice age brown bears. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1651-1654.
- Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, et al. 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**: 1613-1616.
- Leonard JA, Rohland N, Glaberman S, Fleischer RC, Caccone A, et al. 2005. A rapid loss of stripes: the evolutionary history of the extinct quagga. *Biology Letters* **1**: 291-295.
- Leonard JA, Vilà C, Fox-Dobbs K, Koch PL, Wayne RK, et al. 2007. Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. *Current Biology* **17**: 1146-1150.
- Lindqvist C, Schuster SC, Sun Y, Talbot SL, Qi J, et al. 2010. Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5053-5057.
- Lippold S, Knapp M, Kuznetsova T, Leonard JA, Benecke N, et al. 2011. Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature Communications* **2**.
- Loreille O, Vigne JD, Hardy C, Callou C, Treinen-Claustre F, et al. 1997. First distinction of sheep and goat archaeological bones by the means of their fossil mtDNA. *Journal of Archaeological Science* **24**: 33-37.
- Loreille O, Orlando L, Patou-Mathis M, Philippe M, Taberlet P, et al. 2001. Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Current Biology* **11**: 200-203.
- Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, et al. 2009. Coat colour variation at the beginning of horse domestication. *Science* **324**: 485.

- MacPhee RDE, Tikhonov AN, Mol D, Greenwood AD. 2005. Late Quaternary loss of genetic diversity in muskox (*Ovibos*). *BMC Evolutionary Biology* **5**: 49-49.
- Magyari EK, Major A, Balint M, Nédli J, Braun M, et al. 2011. Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evolutionary Biology* **11**.
- McCallum J, Hall S, Lissone I, Anderson J, Huynen L, et al. 2013. Highly Informative Ancient DNA 'Snippets' for New Zealand Moa. *PLOS One* **8**.
- Meiri M, Huchon D, Bar-Oz G, Boaretto E, Horwitz LK, et al. 2013. Ancient DNA and population turnover in Southern Levantine Pigs—Signature of the sea peoples migration? *Scientific Reports* **3**.
- Miller W, Drautz DI, Janecka JE, Lesk AM, Ratan A, et al. 2009. The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19**: 213-220.
- Miller CR and Waits LP. 2003. The history of effective population size and genetic diversity in the Yellowstone grizzly (*Ursus arctos*): Implications for conservation. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4334-4339.
- Miller W, Drautz DI, Ratan A, Pusey B, Qi J, et al. 2008. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**: 351-387.
- Mona S, Catalano G, Lari M, Larson G, Boscato P, et al. 2010. Population dynamic of the extinct European aurochs: genetic evidence of a north-south differentiation pattern and no evidence of post-glacial expansion. *BMC Evolutionary Biology* **10**.

- Montagnon D, Ravaoarimanana B, Rakotosamimanana B, Rumpler Y. 2001. Ancient DNA from *Megaladapis edwardsi* (Malagasy subfossil): Preliminary results using partial cytochrome b sequence. *Folia Primatologica* **72**: 30-32.
- Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, et al. 2005. Genomic sequencing of Pleistocene cave bears. *Science* **309**: 597-600.
- Noonan JP, Coop G, Kudaravalli S, Kudaravalli S, Smith D, et al. 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science* **314**: 1113-1118.
- Nyström V, Dalén L, Vartanyan S, Lidén K, Ryman N, et al. 2010. Temporal genetic change in the last remaining population of woolly mammoth. *Proceedings of the Royal Society B-Biological Sciences* **277**: 2331-2337.
- Oliveira HR, Civán P, Morales J, Rodríguez-Rodríguez A, Lister DL, et al. 2012. Ancient DNA in archaeological wheat grains: preservation conditions and the study of pre-Hispanic agriculture on the island of Gran Canaria (Spain). *Journal of Archaeological Science* **39**: 828-835.
- Ollivier M, Tresset A, Hitte C, Petit C, Hughes S, et al. 2013. Evidence of coat color variation sheds new light on ancient canids. *PLOS One* **8**.
- Orlando L, Bonjean D, Bocherens H, Thenot A, Argant A, et al. 2002. Ancient DNA and the population genetics of cave bears (*Ursus spelaeus*) through space and time. *Molecular Biology Evolution* **19**: 1920-1933.
- Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, et al. 2006. Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Current Biology* **16**: 400-402.
- Orlando L, Calvignac S, Schnebelen C, Douady CJ, Godfrey LR, et al. 2008. DNA from extinct giant lemurs links archaeolemurids to extant indriids. *BMC Evolutionary Biology* **8**.

- Pajnic IZ. 2013. Molecular genetic analyses of 300-year old skeletons from Auersperg tomb. *Zdravniski Vestnik-Slovenian Medical Journal* **82**: 796-808.
- Palkopoulou E, Dalén L, Lister AM, Vartanyan S, Sablin M. 2013. Holarctic genetic structure and range dynamics in the woolly mammoth. *Proceedings of the Royal Society B Biological Sciences* **280**.
- Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG. 2009. Archaeogenetic evidence of ancient Nubian barley evolution from six to two row indicates local adaptation. *PLOS One* **4**.
- Parducci L, Suyama Y, Lascoux M, Bennett KD. 2005. Ancient DNA from pollen: a genetic record of population history in Scots pine. *Molecular Ecology* **14**: 2873-2882.
- Parducci L, Matetovici I, Fontana SL, Bennett KD, Suyama Y, et al. 2013. Molecular and pollen based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology* **22**: 3511-3524.
- Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, et al. 2013. A comparative study of ancient environmental DNA to pollen and microfossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Reviews* **75**: 161-168.
- Poinar HN, Kuch M, Sobolik KD, Barnes I, Stankiewicz AB, et al. 2001. A molecular analysis of dietary diversity for three archaic Native Americans. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 4317-4322.
- Poinar H, Kuch M, McDonald G, Martin P, Pääbo S. 2003. Nuclear gene sequences from a late Pleistocene sloth coprolite. *Current Biology* **13**: 1150-1152.

- Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, et al. 2006. Metagenomics to paleogenomics: Large-scale sequencing of mammoth DNA. *Science* **311**: 392-394.
- Porter TM, Golding GB, King C, Froese D, Zazula G, et al. 2013. Amplicon pyrosequencing late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale reproducibility. *Molecular Ecology Resources* **13**: 798-810.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**: 757-762.
- Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, et al. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334**: 94-98.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A. 2009. DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B-Biological Sciences* **276**: 3395-3402.
- Reich D, Green RE, Kircher M, Krause J, Patterson N, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**: 1053-1060.
- Ricaud FX, Keyser-Tracqui C, Cammaert L, Crubézy E, Ludes B. 2003. Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. *American Journal of Physical Anthropology* **123**: 351-360.
- Ritchie PA, Millar CD, Gibb GC, Baroni C, Lambert DM. 2004. Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adelie penguin lineages in Antarctica. *Molecular Biology and Evolution* **21**: 240-248.

- Robinson TJ, Bastos AD, Halanych KM, Herzig B. 1996. Mitochondrial DNA sequence relationships of the extinct blue antelope *Hippotragus leucophaeus*. *Naturwissenschaften* **83**: 178-182.
- Rohland N, Reich D, Mallick S, Meyer M, Green RE, et al. 2010. Genomic DNA Sequences from Mastodon and Woolly Mammoth Reveal Deep Speciation of Forest and Savanna Elephants. *PLOS Biology* **8**.
- Römpler H, Rohland N, Lalueza-Fox C, Willerslev E, Kuznetsova T, et al. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**: 62-62.
- Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, et al. 2005. The genetics of the pre-Roman Iberian Peninsula: A mtDNA study of ancient Iberians. *Annals of Human Genetics* **69**: 535-548.
- Der Sarkissian CD, Balanovsky O, Brandt G, Khartanovich V, Buzhilova A, et al. 2013. Ancient DNA reveals prehistoric gene flow from Siberia in the Complex Human Population History of North East Europe. *PLOS Genetics* **9**.
- Seabrook-Davison M, Huynen L, Lambert DM, Brunton DH. 2009. Ancient DNA Resolves Identity and Phylogeny of New Zealand's Extinct and Living Quail (*Coturnix* sp.). *PLOS One* **4**.
- Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, et al. 2004. Rise and fall of the Beringian steppe bison. *Science* **306**: 1561-1565.
- Shepherd LD and Lambert DM. 2008. Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* **17**: 2174-2184.
- Shepherd LD, Millar CD, Ballard G, Ainley DG, Wilson PR, et al. 2005. Microevolution and mega-icebergs in the Antarctic. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 16717-16722.

- Steeves TE, Holdaway RN, Hale ML, McLay E, McAllan IAW, et al. 2010. Merging ancient and modern DNA: extinct seabird taxon rediscovered in the North Tasman Sea. *Biology Letters* **6**: 94-97.
- Stiller M, Baryshnikov G, Bocherens H, Grandal d'Anglade A, Hilpert B, et al. 2010. Withering Away-25,000 Years of Genetic Decline Preceded Cave Bear Extinction. *Molecular Biology Evolution* **27**: 975-978.
- Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, et al. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**: 1088-1091.
- Valdiosera CE, García N, Anderung C, Dalén L, Crégut-Bonnoure E, et al. 2007. Staying out in the cold: glacial refugia and mitochondrial DNA phylogeography in ancient European brown bears. *Molecular Ecology* **16**: 5140-5148.
- Valdiosera CE, García-Garitagoitia JL, García N, Doadrio I, Thomas MG, et al. 2008. Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences of the United States of America* **105**: 5123-5128.
- Valentine K, Duffield DA, Patrick LE, Hatch DR, Butler VL, et al. 2008. Ancient DNA reveals genotypic relationships among Oregon populations of the sea otter (*Enhydra lutris*). *Conservation Genetics* **9**: 933-938.
- Verginelli F, Capelli C, Coia V, Musiani M, Falchetti M, et al. 2005. Mitochondrial DNA from prehistoric canids highlights relationships between dogs and South-East European wolves. *Molecular Biology Evolution* **22**: 2541-2551.
- Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, et al. 2004. The Etruscans: A population-genetic study. *American Journal of Human Genetics* **74**: 694-704.

- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MT, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795.
- Wood JR, Rawlence NJ, Rogers GM, Austin JJ, Worthy TH, et al. 2008. Coprolite deposits reveal the diet and ecology of the extinct New Zealand megaherbivore moa (Aves, Dinornithiformes). *Quaternary Science Reviews* **27**: 2593-2602.
- Wood JR, Wilmshurst JM, Worthy TH, Cooper A. 2012. First coprolite evidence for the diet of *Anomalopteryx didiformis*, an extinct forest ratite from New Zealand. *New Zealand Journal of Ecology* **36**: 164-170.
- Wood JR, Wilmshurst JM, Richardson SJ, Rawlence NJ, Wagstaff SJ, et al. 2013. Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes). *Proceedings of the National Academy of Sciences of the United States of America* **110**: 16910-16915.
- Zeyland J, Wolko L, Bocianowski J, Szalata M, Slomski R, et al. 2013. Complete mitochondrial genome of wild aurochs (*Bos primigenius*) reconstructed from ancient DNA. *Polish Journal of Veterinary Sciences* **16**: 265-273.

— CHAPTER 2 —

---

**A CRITICAL EVALUATION *of* HOW ANCIENT DNA BULK BONE  
METABARCODING COMPLEMENTS TRADITIONAL  
MORPHOLOGICAL ANALYSIS *of* FOSSIL ASSEMBLAGES**

---

*A man with two watches never knows the time. A man with one watch only thinks he knows the time.*

- Segal's law

## 2.1 PROLOGUE

---

Chapter 1 of this thesis introduced several recently developed methods that make use of ancient DNA extracted from novel substrates. ‘Bulk bone metabarcoding’ is one such method that has been used to rapidly explore past biodiversity through the simultaneous molecular identification of hundreds of morphologically un-diagnostic bone fragments. While it has been argued that the bulk bone metabarcoding method complements traditional morphological methods, this has yet to be quantified. The preceding epigraph encapsulates the chief issue concerning the use of this method: namely, that most studies implementing the method look at just “one watch”. However, we are unable to identify or estimate error without replication, and without an estimation of uncertainty, how confident can we be that the method allows an accurate characterisation of past biodiversity? As the method has the potential to enrich knowledge across various disciplines, and may be widely used in the future, it is vital to investigate the limits of its utility.

For the first time, we critically evaluate this method, highlighting the benefits and caveats that need to be considered when implementing it. Using carefully designed experiments that unite palaeontological, statistical, ecological, and genetic principles, we show that the inclusion of the bulk bone metabarcoding method alongside morphological analysis of fossils can give greater insight into past biodiversity than either approach alone, which can have profound implications for the conservation and management of species. This chapter also develops stringent protocols and standards for applying the bulk bone metabarcoding method that bolster the legitimacy of its application across multiple diverse fields of study.

The study presented in this chapter resulted in a manuscript published in a 2015 issue of *Quaternary Science Reviews* (Grealy et al., Vol. 128, Pg. 37-47), a facsimile of which can be found in Appendix I. This chapter is a reproduction of the aforementioned manuscript (formatting, including in-text referencing, cross referencing and headings excepted).

### **2.1.1 ACKNOWLEDGEMENTS**

Work was funded by Australian Research Council grants (DP120104435). MB and GP were supported in this research by an ARC Future Fellowships FT0991741 and FT130101728. We thank the Pawsey supercomputer facility for assistance in data analysis. PS wishes to thank K Davis and YP Montelle for help with FF collection and analysis, and J Wood for facilitating AMS dating of FF.

### **2.1.2 AUTHOR CONTRIBUTIONS**

AG designed experiments and wrote the manuscript, with methods written by AG, MM, DF, and PS. All authors contributed to editing of the manuscript. AG, DM, MM, MB assisted with analysis. Samples were collected by MM, DM, PS, and GP. Morphological identification was carried out by MM, DF, PS.

### **2.1.3 AUTHOR DECLARATIONS**

All necessary permits were obtained for the described study, which complied with all relevant regulations.

The authors declare no competing interests.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work (e.g., journal articles) in which the copyright is held by another party (e.g., publisher, co-author). Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged. Reprinted from Publication *Quaternary Science Reviews*, Vol. 128, Grealy AC et al., A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages, Pg. 37-47, Copyright (2015), with permission from Elsevier.

Supplementary data related to this article can be accessed online at:

<http://dx.doi.org/10.1016/j.quascirev.2015.09.014>, as well as section 2.8.

Sequencing data has been deposited on the online data repository DataDryad, and is available at: <http://dx.doi.org/10.5061/dryad.gt377>.

**A CRITICAL EVALUATION OF HOW ANCIENT DNA BULK BONE METABARCODING  
COMPLEMENTS TRADITIONAL MORPHOLOGICAL ANALYSIS OF FOSSIL  
ASSEMBLAGES**

—*in*—

***QUATERNARY SCIENCE REVIEWS* (2015) | VOL. 128 | PG. 37-47**

Alicia C. Grealy<sup>1,2</sup>, Matthew C. McDowell<sup>3</sup>, Paul Scofield<sup>4</sup>, Dáithí C. Murray<sup>1,2</sup>,  
Diana A. Fusco<sup>3</sup>, James Haile<sup>1,2</sup>, Gavin J. Prideaux<sup>3</sup>, and Michael Bunce<sup>1,2\*</sup>

<sup>1</sup> Trace and Environmental DNA (TrEnD) laboratory, Department of Environment  
and Agriculture, Curtin University, Kent St, Bentley, WA, Australia, 6102

<sup>2</sup> Ancient DNA laboratory, School of Veterinary and Life sciences, Murdoch  
University, South St, Murdoch, WA, 6150

<sup>3</sup> School of Biological Sciences, Flinders University, Bedford Park, SA, Australia,  
5042

<sup>4</sup> Canterbury Museum, Christchurch 8013, New Zealand

\* Corresponding author: michael.bunce@curtin.edu.au

**Key words:** aDNA ancient, DNA, archaeology, biodiversity, bulk bone,  
experimental error, fossil, metabarcoding, next-generation sequencing, palaeontology

## 2.2 ABSTRACT

---

When pooled for extraction as a bulk sample, the DNA within morphologically unidentifiable fossil bones can, using next-generation sequencing, yield valuable taxonomic data. This method has been proposed as a means to rapidly and cost-effectively assess general ancient DNA preservation at a site, and to investigate temporal and spatial changes in biodiversity; however, several caveats have yet to be considered. We critically evaluated the bulk bone metabarcoding (BBM) method in terms of its: (i) repeatability, by quantifying sampling and technical variance through a nested experimental design containing subsamples and replicates at several stages; (ii) accuracy, by comparing morphological and molecular family-level identifications; and (iii) overall utility, by applying the approach to two independent Holocene fossil deposits, Bat Cave (Kangaroo Island, Australia) and Finsch's Folly (Canterbury, New Zealand). For both sites, bone and bone powder subsampling were found to contribute significantly to variance in molecularly identified family assemblage, while the contribution of library preparation and sequencing was almost negligible. Nevertheless, total variance was small. Sampling over 80% fewer bones than was required to morphologically identify the taxonomic assemblages, we found that the families identified molecularly are a subset of the families identified morphologically and, for the most part, represent the most abundant families in the fossil record. In addition, we detected a range of extinct, extant and endangered taxa, including some that are rare in the fossil record. Given the relatively low sampling effort of the BBM approach compared with morphological approaches, these results suggest that BBM is largely consistent, accurate, sensitive, and therefore widely applicable. Furthermore, we assessed the overall benefits and caveats of the method, and suggest a workflow for palaeontologists, archaeologists, and geneticists that will help mitigate these caveats. Our results show that DNA analysis of bulk bone samples can be a universally useful tool for studying past biodiversity, when integrated with existing morphology-based approaches. Despite several limitations that remain, the BBM method offers a cost-effective and efficient way of studying fossil assemblages, offering complementary insights into evolution, extinction, and conservation.

## 2.3 INTRODUCTION

---

**F**OR over a century, the study of fossils has played a major role in understanding prehistoric life and evolutionary processes. In particular, morphological analyses of fossils can reveal species that existed in the past, help elucidate the evolutionary relationships of extinct and extant species (e.g., Donoghue et al. 1989; Deméré et al. 2005; Manos et al. 2007), and assist the development of palaeoenvironment reconstructions that provide insights into the evolutionary and ecological impacts of environmental changes (e.g., Rodríguez-Aranda and Calvo, 1998; Zhang et al. 2008). However, such traditional methods have limitations. For instance, taxonomic assignments of fossils have been necessarily reliant on morphological distinctions, making the identification of fragmented or taxonomically-mixed fossil material challenging, if not impossible. This limitation can be partially overcome in some Late Quaternary contexts with the application of ancient DNA (aDNA) techniques. Over the past two decades, aDNA has proved to be a useful complement to the morphological study of fossils, and is rapidly growing in popularity, accessibility, and applicability. In combination with next-generation sequencing (NGS), aDNA has been used to test phylogenetic relationships, and timing of speciation and extinction trajectories (e.g., Krause et al. 2010), resolve taxonomy (e.g., Rohland et al. 2010), reconstruct palaeoenvironments (e.g., Willerslev et al. 2003), and measure historic genetic diversity (e.g., Larson et al. 2002; Allentoft et al. 2014).

Despite the utility of aDNA analysis, unidentifiable bone fragments that are retrieved from palaeontological and archaeological excavations are often too numerous and small to justify the expense of aDNA analysis. However, if such bones are pooled for aDNA extraction as one bulk sample, the pool may be sequenced cost-effectively to yield valuable systematic data useful for assessing past biodiversity over time and space (Murray et al. 2013). In addition, bulk bone samples may be useful for evaluating general aDNA preservation at a site, without requiring the destruction of complete or precious fossil specimens (Murray et al. 2013). The bulk bone method employs a metabarcoding approach (Taberlet et al. 2012), which involves: (1) simultaneous extraction of aDNA from multiple unidentifiable fragments of bone; (2) amplification of short, ‘diagnostic’ regions of mitochondrial genes by polymerase chain reaction (PCR); and (3) sequencing (via NGS) of these amplicons to identify

the species present by comparison with a genetic database of known species (e.g., GenBank; Altschul et al. 1990; Benson et al. 2006). Metabarcoding has been used to evaluate both present and past biodiversity (Epp et al. 2012) through the analysis of environmental samples such as sediments (e.g., Jørgensen et al. 2011; Andersen et al. 2012; Pedersen et al. 2013; Epp et al. 2015; Pansu et al. 2015), seawater (e.g., Minamoto et al. 2012; Thomsen et al. 2012), coprolites (e.g., Hofreiter et al. 2003), and middens (e.g., Murray et al. 2012), and has even been able to detect taxa that were considered extinct based on the macrofossil record (e.g., Haile et al. 2009; Haouchar et al. 2014). Using a metabarcoding approach to generate biodiversity data has the potential to significantly reduce workload and costs compared with a morphological approach that can be labor intensive, or require large amounts of taxonomic expertise and time investment (Ji et al. 2013). When combined with the use of indexing (Binladen et al. 2007; Meyer et al. 2007; Kircher et al. 2012) (where DNA from each bulk sample is ‘tagged’ with a few unique bases), multiple DNA samples can be combined with equimolarity and sequenced in parallel (i.e., ‘sample multiplexing’) on an NGS platform, increasing throughput and further reducing cost and time.

Although the bulk bone metabarcoding (BBM) method has been implemented in several recent studies (e.g., Murray et al. 2013; Haouchar et al. 2014), several caveats and biases of this method have yet to be addressed. Firstly, the amount of variance attributable to experimental error in the BBM method, as well as other environmental metabarcoding methods (Andersen et al. 2012; Pedersen et al. 2013; Porter et al. 2013), has not been measured. As such, it is unknown at what step, and to what extent, efforts need to be concentrated to minimise experimental error (Earp et al. 2011), and an optimal experimental protocol has not been developed. This is important if we wish to confidently compare how biodiversity has changed across time or space, in order to reliably determine what has driven those changes (Wooley et al. 2010). The “ability of the researcher to obtain a statistically significant result” (Kitchen et al. 2010) is influenced by the treatment effect, and repeatability (or precision), which is affected by biological variability and technical noise. For the BBM method, the treatment effect is the variance in biodiversity *between* samples that arises from differences between paleontological collection sites (space) or stratigraphic layers (time); biological variability refers to the differences in

biodiversity *within* samples resulting from subsampling effort and differential DNA preservation in the fossils; and technical noise is the variability in biodiversity introduced by the experimental protocol itself (including subsampling bone powder for DNA extraction, human error, random contamination, stochastic variations in quantitative PCR and amplification biases, aDNA damage, PCR and sequencing errors, and amplicon pooling during the creation of NGS libraries). In order to isolate the treatment effect from the background (Kitchen et al. 2010), we can quantify the contribution that each of these factors makes to the total variance in the data (the ‘experimental error’) through a careful experimental design containing multiple subsamples, biological and technical replicates, stringent laboratory protocols, and the use of multiple blank controls at each step (Kuehl 2000; Macgregor 2007; Kitchen et al. 2010).

Secondly, most metabarcoding studies of environmental samples have found discrepancies between estimates of biodiversity obtained from DNA metabarcoding methods and those obtained from traditional biodiversity sampling methods (Ji et al. 2013) because some species identified morphologically were not identified via DNA methods, and vice versa (Hajibabaei et al. 2011; Murray et al. 2013). These discrepancies arise from differences in the biomass and behavior of animals (Andersen et al. 2012), as well as sampling effort, differential preservation of both fossils and the aDNA within them, technical ‘noise’ (such as amplification bias, PCR and sequencing error; Fonseca et al. 2012), and deficiencies in reference genetic databases, such as GenBank (Pedersen et al. 2014). It is likely that BBM studies would be affected by similar biases (Murray et al. 2013); however, the extent to which bulk bone taxonomic identifications reflect those arising from the fossil record has yet to be examined.

In this paper we critically evaluate the BBM approach in terms of its repeatability, accuracy and overall utility. Repeatability was assessed by estimating the contribution to variance made by experimental error using a nested experimental design containing pooled bone subsamples (biological variability), bone powder (extraction) subsamples, and library preparation and sequencing run replicates (technical variability)—this allows us to determine where sampling effort and replication need to be concentrated in order to reduce variance in the detection of

families and operational taxonomic units (OTUs). Accuracy was assessed by comparing the family assemblages derived from morphological identification of fossil collections with those derived from a subset subjected to BBM analysis. Finally, overall utility was assessed by applying the approach to two independent Holocene fossil deposits, Bat Cave (BC; Kangaroo Island, Australia) and Finsch's Folly (FF; Canterbury, New Zealand). These methods enabled us to gauge the strengths, limitations, and biases of the BBM approach in order to assess how it complements traditional paleontological methods.

## 2.4 MATERIALS *and* METHODS

---

### 2.4.1 STUDY SYSTEMS *and* SAMPLE COLLECTION

**BAT CAVE.** Located in the Kelly Hill Caves Conservation Park, southwest Kangaroo Island, South Australia (Figure S2.8.1), Bat Cave (BC; 35° 59' S, 136° 54' E; Cave Exploration Group of South Australia no. 5K65) consists of a single chamber with a rock-pile entrance (S2.8.1). The taphonomic characteristics of the BC assemblage (maximum species body mass, presence of invertebrate remains, degree of digestive erosion, and inclusion of predatory species in the accumulation) are consistent with accumulation by the predatory activity of Boobook owls (*Ninox novaeseelandiae*; Fusco 2014). Gloves were worn to collect fossils from shallow unstratified, calcareous sandy sediments found among rocks on the entrance slope. Samples were dry sieved using 1.5-mm mesh to concentrate bones. Diagnostic elements were removed for identification and analysis using a small brush and forceps. Morphologically non-diagnostic bones were stored in an airtight bag at -20°C until needed for molecular analyses.

**FINSCH'S FOLLY.** FF (44 2'S, 170 43.5' E) is a shallow (< 8 m) pitfall cave located in an isolated outcrop of limestone in southwestern Canterbury (Figure S2.8.1). Fossils were collected from muddy, unstratified, calcareous sediments that had accumulated at the bottom of the pitfall and also from sediments that have been concreted to the walls during a previous period of infill. Samples were wet sieved using 1-mm mesh to concentrate bones. Diagnostic elements were removed for

identification and analysis. The remaining morphologically non-diagnostic bones were dried and stored in airtight bags at  $-20^{\circ}\text{C}$  until needed for molecular analyses.

#### **2.4.2 DATING**

One bone sample from BC (the only bone suitable for radiocarbon dating, a left humerus of *N. novaeseelandiae*, Wk-36239), and three samples from FF were Accelerator Mass Spectrometry (AMS) radiocarbon dated by Waikato Radiocarbon Dating Laboratory. The conventional ages of the samples were calibrated with OxCal v.4.1.7 (Bronk Ramsey 2010) using the SHCal13 Southern Hemisphere calibration curve (Hogg et al. 2013). The *N. novaeseelandiae* bone yielded a conventional age of  $2862 \pm 25$  BP and calibrated age of 2750-3000 cal BP (Fusco 2014). While it is likely that this bone does not represent the age of all the bones in the BC accumulation, the presence of introduced species in the assemblage bolsters our confidence that it accumulated in the late Holocene, which provides adequate temporal resolution for comparison with FF. The samples from FF yielded conventional ages of  $1344 \pm 25$  BP,  $1646 \pm 25$  BP, and  $1645 \pm 24$  BP, and calibrated ages of 1150-1300 cal BP, 1400-1600 cal BP and 1400-1600 cal BP, respectively.

#### **2.4.3 MORPHOLOGICAL ASSESSMENT of FOSSIL ASSEMBLAGE**

Diagnostic bones from both sites (whole and part skulls, maxillae, dentaries and/or teeth, and post-cranial bones) were identified using published descriptions and comparative specimens (cf. Fusco 2014). All specimens were identified to the lowest taxonomic level possible (usually species). The minimum number of individuals (MNI) was determined by counting the number of the most common diagnostic element of each species in each assemblage. MNI was converted to relative abundance ( $R_i\%$ ), an expression of the MNI of a given species as a proportion of the total MNI for that collection (McDowell et al. 2012; Fusco 2014).

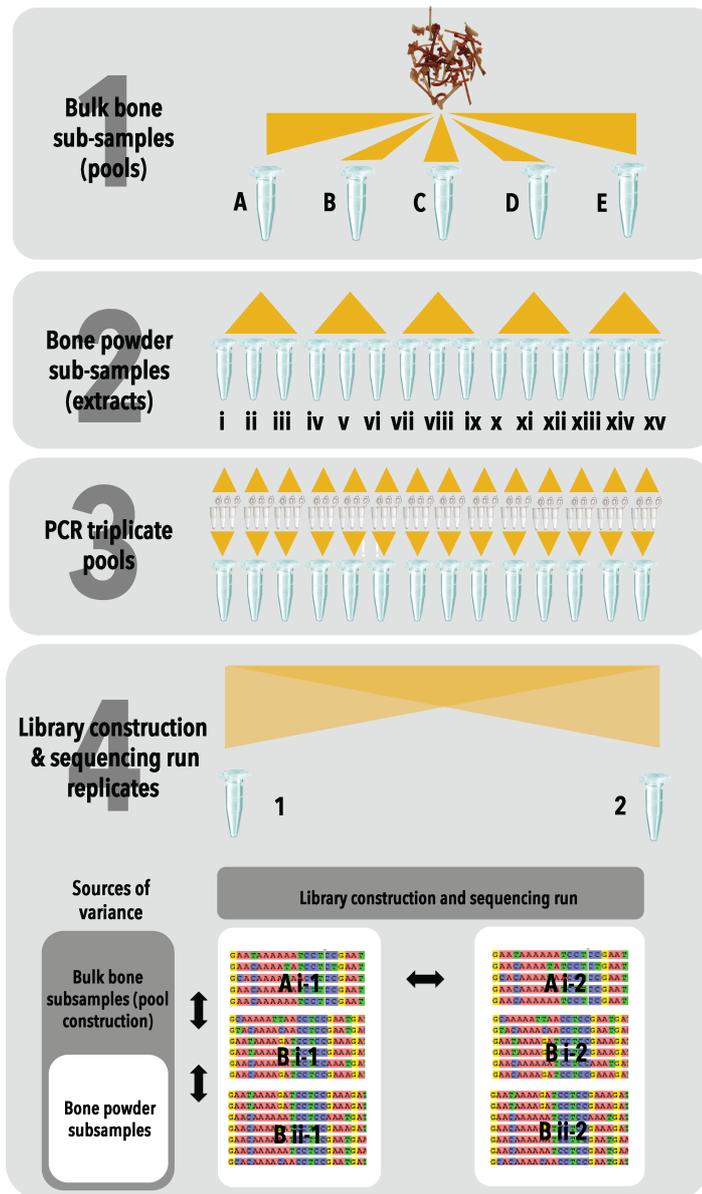
#### **2.4.4 EXPERIMENTAL DESIGN**

In order to partition the treatment effect from the technical ‘noise’ (Kitchen et al. 2010), we need to either control extraneous factors such that they do not contribute

to variance, or quantify the contribution that each of these factors makes to the total variance in the data (Kuehl 2000). To partition the components of experimental variance, differences in family biodiversity between and within subsamples and replicates at different levels of the experimental protocol was compared using a nested experimental design (Figure 2.1; S2.8.2). Differences in taxonomic assemblages between replicate sequencing runs within an extract allows us to determine the amount of technical variance that can be attributed to library preparation and sequencing run; comparing differences between extraction replicates (i.e., randomly drawn bulk bone powder subsamples) within a bulk bone subsample allows us to determine the amount of variance that can be attributed to subsampling bulk bone powder during the extraction process; and comparing differences between bulk bone subsamples (randomly drawn) allows us to determine the amount of biological variability within the greater bulk bone sample. Other aspects of the experimental protocol that could contribute to variance were controlled: all bulk bone powder subsamples were extracted together such that the extraction process does not contribute to variance, and PCR amplification was replicated, but replicates were then combined—in this way the contribution of the PCR process to variance was accounted for albeit not quantified.

#### **2.4.5 DNA EXTRACTION *of* BULK BONES**

The sieved and sorted bulk bone was subsampled, then ground into powder in a designated clean facility at Murdoch University, WA, Australia (Figure S2.8.2). All preparation surfaces and tools were cleaned with a solution of 10% bleach followed by 70% ethanol, and personal protective equipment (S2.8.3) was worn, in order to minimise contamination of samples with exogenous DNA. DNA from three, 100 mg aliquots of bone powder from each 50-bone subsample were extracted by incubating the powder in 1.5 mL of digest buffer containing final concentrations of 0.1% Triton-X-100, 0.02 M Tris-HCl, 1 mg/ml Proteinase K powder, 0.01 M DTT in 0.5 M EDTA for 24 hr at 55°C. The supernatant was collected, concentrated to a volume of 50 µl in a 30,000 MWCO Vivaspin-500 column, and purified using a modified *QIAGEN* protocol (S2.8.3), eluting in 50 µl EB buffer (*QIAGEN*, cat. No. 19066). One extraction control was included per 10 extractions. Extractions and



**FIGURE 2.1** | NESTED EXPERIMENTAL DESIGN WITH REPLICATION AT THREE STAGES: bulk bone subsampling or pool construction (experimental replicates), bone powder subsampling (extraction), and library construction and sequencing run (technical replicates). Five different bone-powder pools were constructed by randomly drawing 50 bones from the same bulk bone sample. Three extractions were performed on each pool by subsampling 100 mg of bone powder from each pool. For each site, a sequencing library was constructed containing a blend of the amplicons generated from each extract, and this library was sequenced twice. Sources of error can be partitioned by examining the differences within and between sequencing replicates (e.g., sequencing run 1 and 2), extraction replicates (e.g., Bi and Bii), and bone powder pools (e.g., A and B).

subsequent qPCR reactions were prepared in a separate ultra-clean facility to avoid contamination.

#### 2.4.6 NEXT-GENERATION SEQUENCING *of* DNA AMPICONS

Following the protocol outlined in Murray et al. (2013), extracts were diluted and screened for inhibition and amplification efficiency via qPCR (S2.8.4). Dilutions exhibiting the least inhibition and greatest amplification efficiency were then amplified in triplicate via qPCR using indexed primers specific for a diagnostic barcoding region of the *16S rRNA* gene in mammals for the BC samples (131 bp or 91 bp *sans* primer), and *12S rRNA gene* (274 bp or 232 bp *sans* primer) in birds for the FF samples (S2.8.4, Table S2.8.1). These metabarcodes were chosen according to the predominant class of organism identified morphologically at each site (aves at FF, mammals at BC) but without prior knowledge of the species within, and are preferred over standard barcoding genes such as *COI* or *cytb* due to their universality, the taxonomic resolution of the amplified insert, and small size (Deagle et al. 2014; Pedersen et al. 2014; Thomsen and Willerslev, 2015). 25 µl qPCR reactions consisted of final concentrations of 0.4 mg/ml Bovine Serum Albumin, 1X *ABI* GeneAmp PCR Buffer, 2 mM MgCl<sub>2</sub>, 0.4 uM forward and reverse primers, 0.25 mM dNTPs, as well as 0.25 µl of AmpliTaqGold DNA polymerase and 0.6 µl of 1:2000 SYBR green, in HPLC-grade water (S2.8.4). Thermocycling conditions included heat denaturation at 95°C for five min, followed by 50 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 45 sec, and a +1°C melt curve at 95°C for 15 sec, 60°C for 1 min, 95°C for 15 sec, with a final 72°C extension for 10 min (S2.8.4). Triplicates were combined and purified using an *Agencourt* AMPure XP PCR purification kit according the manufacturer's instructions, and samples were pooled in approximately equimolar concentrations twice (S2.8.5) to create two sequencing libraries per primer set (Figure 2.1). Extraction controls, three negative controls (water), and a positive control were included for each qPCR. All post-PCR methods were performed in a physically separated laboratory in keeping with standard aDNA practice (Willerslev and Cooper 2005; Shapiro and Hofreiter 2012). The absolute concentration of sequencing libraries was quantified via qPCR as per Murray et al. (2012) (S2.8.5) to determine how much to add to the sequencing reaction. Emulsion

PCR and enrichment was performed by following the manufacturer's instructions for the Ion PGM Template OT2 200 kit (for both BC libraries), and the Ion PGM Template OT2 400 kit (for both FF libraries). Each library was sequenced separately using Ion PGM Sequencing 200 v2 kits and 400 kits as per the manufacturer's instructions.

#### **2.4.7 SEQUENCE IDENTIFICATION *and* BIOINFORMATIC ANALYSIS**

After sequencing, the FastQ file was downloaded and imported into Geneious v.7.0.6 (<http://www.geneious.com>, Kears e et al. 2012), sorted by barcode and primer sequence (with only exact matches accepted), and trimmed of all primers using methods in Murray et al. (2013) with minor changes (S2.8.6). Quality control, chimera filtering, and abundance filtering were performed using several packages within Galaxy ([usegalaxy.org](http://usegalaxy.org); Giardine et al. 2005; Blankenberg et al. 2010; Goecks et al. 2010) and USEARCH v.6.1 (Edgar 2010; Edgar et al. 2011) as implemented in QIIME v.1.8.0 ([qiime.org](http://qiime.org); Caporaso et al. 2010b) using methods in Murray et al. (2013) with minor changes (S2.8.6); namely, reads with an average quality score less than 25, as well as chimeric reads, and low-abundant reads (<0.1% of the total number of reads) were discarded. Taxonomic identification was achieved by using YABI ([ccg.murdoch.edu.au/yabi](http://ccg.murdoch.edu.au/yabi); Hunter et al. 2012) to compare the sequences to NCBI's GenBank (Benson et al. 2006) nucleotide reference database via BLASTn (Altschul et al. 1990; no low complexity filter, gap penalties existence of 5 and extension 2 (default), e-value <1e-10, word size 7). BLAST results were imported into MEGAN v.4.70.4 ([ab.inf.uni-tuebingen.de/data/software/megan4](http://ab.inf.uni-tuebingen.de/data/software/megan4); Huson et al. 2007) for taxonomic assignment (S2.8.6): families were considered present if the at least 95% of the query aligned to a known member of the family with a similarity of 90% or more. In order to estimate overall genetic diversity at the 3% level within each family, sequences within each identified family were then subjected to taxonomy-independent (OTU; operational taxonomic unit) analyses by clustering sequences at 97% identity using USEARCH v.6.1 (Edgar 2010; Edgar et al. 2011) as implemented in QIIME v.1.8.0 ([qiime.org](http://qiime.org); Caporaso et al. 2010b) (S2.8.6). 97% identity was used to account for sequencing errors.

#### 2.4.8 STATISTICAL ANALYSIS

We restricted the scope of our analyses to quantifying the variance around the detection (presence/absence) of families and OTUs, and comparing their detection between morphological and molecular datasets generated from our study sites. As morphological and molecular techniques suffer different biases (Figure 2.4; S2.8.13), fossil abundance does not correlate with read abundance (Bohmann et al. 2014; S2.8.8, Table S2.8.3). Therefore, it is not valid to compare abundance between morphological and molecular datasets. Similarly, compared to morphological analysis, it is difficult to make highly credible species- and genus-level molecular identifications. Therefore the morphological and molecular datasets are comparable only at the family level.

A nested non-parametric multivariate analysis of variance (nested NPMANOVA or PERMANOVA) was used (<https://www.stat.auckland.ac.nz/~mja/Programs.htm>; Anderson 2001) to analyse the variance in the detection of families and OTUs from sequence data attributable to biological subsamples and experimental replicates (S2.8.7, Table S2.8.2). Sørensen's index of similarity (Sørensen 1948) was used to compare the diversity of families (presence/absence) between morphological and molecular data sets for both sites (calculated by hand; S2.8.7). We also calculated the percentage of families we would have *expected* to identify that were actually identified through BBM: these were families that represent >1% of the fossil record (i.e., had a 95% chance of one representative bone being sampled in 250), that have at least one of the species identified within that family represented in GenBank for the metabarcoding gene used. In addition, EstimateS v.9.1.0 ([purl.oclc.org/estimates](http://purl.oclc.org/estimates); Colwell 2013) was used to generate rarefaction curves in order to examine how differences in sampling depth between morphological and molecular samples influenced the number of families that could be identified.

#### 2.5 RESULTS *and* DISCUSSION

---

Five subsamples of 50 bones were extracted three times, independently amplified for one of two mitochondrial gene regions, and sequenced twice to yield, on average, a

**TABLE 2.1** | STATISTICS TABLE, including: nested PERMANOVA (4999 permutations, Jaccard dissimilarity or distance) showing the contributions to variance in the molecular family assemblages (presence/absence; bold), and the OTU assemblages (presence/absence) of FF and BC (asterisks indicate significance or  $\alpha < 0.05$ ); Standard Deviation (St Dev., Jaccard dissimilarity); Sørensen's index of similarity between morphological and molecular family assemblages; the percentage of expected ( $> 1\%$  of the fossil record with reference sequence present in GenBank) and unexpected ( $< 1\%$  of the fossil record or with no reference sequence present in GenBank) families detected molecularly; the number of families detected morphologically had only 250 bones been sampled.

Site	Source of variance	df	Contribution to variance (%)	St Dev (Jaccard dissimilarity)	Sørensen's Index of similarity	Expected families detected molecularly (%)	Unexpected families detected molecularly (%)	Families detected morphologically if 250 bones were sampled ( $\pm 95\%$ CI)
Finsch's Folly	Pool	4	<b>33.0*</b> 18.4*	0.36	0.48	100 (6 out of 6)	10.5 (2 out of 19)	10.40 $\pm$ 4.76 (23-61%)
	Extract (Pool)	10	<b>58.9*</b> 61.0*					
	Library construction and sequencing run (residual)	15	<b>8.1</b> 20.6*					
Bat Cave	Pool	4	<b>95.0*</b> 72.0*	0.36	1.00	100 (3 out of 3)	100 (2 out of 2)	3.31 $\pm$ 1.15 (43-90%)
	Extract (Pool)	10	<b>2.5*</b> < 0.1					
	Library construction and sequencing run (residual)	15	<b>2.5</b> 28.0					

total of 27722 and 72639 unique sequence reads per run, with an average of 1848 and 4842 unique sequence reads (Table S2.8.4) per sample for FF and BC, respectively. Eight and five families were identified molecularly, with on average four and two OTUs per family, for FF and BC, respectively (Table 2.1; Table S2.8.5). These sequence data were used to investigate the repeatability, accuracy, and overall utility of the BBM method.

### **2.5.1 REPEATABILITY *of* THE BBM METHOD**

We found that technical errors in the BBM method contribute similarly to the variance between widely different samples at both the family level and OTU level. Library construction and sequencing run contribute little to the variance at the family level, for both sites (8.1% FF, 2.5% BC; Table 2.1). Therefore, if family-level resolution is required, library construction and sequencing run may not need to be replicated, particularly on less error-prone sequencing platforms: this occurs despite the differences in sequencing depth between run replicates (Table S2.8.4), indicating that the amount of coverage has little impact on repeatability as we capture the same taxonomic assemblages with a small amount of coverage as we do with high coverage (Caporaso et al. 2010a). However, library construction and sequencing run contributed more to the variance at the OTU-level for both sites (20.2% FF, 24.3% BC; Table 2.1). The higher variance observed at the OTU-level compared with variance at the family-level may simply reflect how nested taxonomic levels (e.g., species or OTUs) are inherently more diverse than the level they are nested within (e.g., families). Consequently, we will more consistently sample families than the taxa within them (Foote and Miller 2007), resulting in a lower variance at the family level. In addition, OTU diversity is also notoriously sensitive to aDNA damage, PCR and sequencing errors, as well as sequencing depth, both of which can unpredictably inflate diversity (Bragg 2013; Murray et al. 2015). Therefore, sequencing run replication and coverage become extremely important, albeit costly, at the OTU level as it contributes to about a quarter of the variance in OTU diversity. The implementation of stringent quality control protocols during sequence data filtering, as well as subsampling (i.e., rarefying) sequences to account for differences in coverage may also help eliminate variance in OTU diversity (Murray et al. 2013, 2015). These results also suggest that high diversity samples will require

significantly more replication at all levels in order to consistently recover that diversity.

Although library preparation and sequencing run appears to make similar contributions to variance, each sample may require specific optimisation to determine the best sampling strategy to account for the variability within bulk bone samples and extract subsamples: for FF, the greatest contribution to variance was from extract subsample nested within bulk bone subsample pool (58.9% for family, 61.0% for OTU; Table 2.1; S2.8.9, Figure S2.8.3), while for BC, most of the variability is introduced when subsampling the bulk bone sample to form pools (98.7% for family, 75.8% for OTU; Table 2.1; S2.8.9, Figure S2.8.3). Similarly, pools contributed less to the variance for FF (33.0% for family, 18.4 % for OTU; Table 2.1; S2.8.9, Figure S2.8.3), while extract subsample contributed less to the variance at BC (2.5% for family, <0.1% for OTU; Table 2.1; S2.8.9, Figure S2.8.3). Thus, subsampling needs to be concentrated at different levels in different samples, and this is likely to be largely dictated by the diversity of the sample. Finally, total variance within FF and BC at the family-level is small and is about the same (standard deviation = 0.36 Jaccard dissimilarity; Table 2.1); however, to test whether the overall repeatability, or reproducibility, of the BBM method is truly universal would require a survey of many more sites.

Nevertheless, subsampling and replication at various steps in the bulk bone method will help reduce error variance and increase the power to detect significant temporal and spatial changes in biodiversity. That is, multiple subsampling and replication is essential to be able to distinguish whether the differences we observe in biodiversity between samples are real, or simply a result of insufficient sampling, aDNA damage, and technical error (S2.9.13). However, the cost-efficacy of the BBM method would be lost by sampling beyond what is necessary. Knowledge of the contributions to variance quantified by a pilot study can help determine how much sampling and replication would be required to obtain enough statistical power to detect a predetermined effect size significantly (if one exists). Resource allocation can be further optimised by factoring in cost per replicate and total budget into sample size considerations (Kitchen et al. 2010).



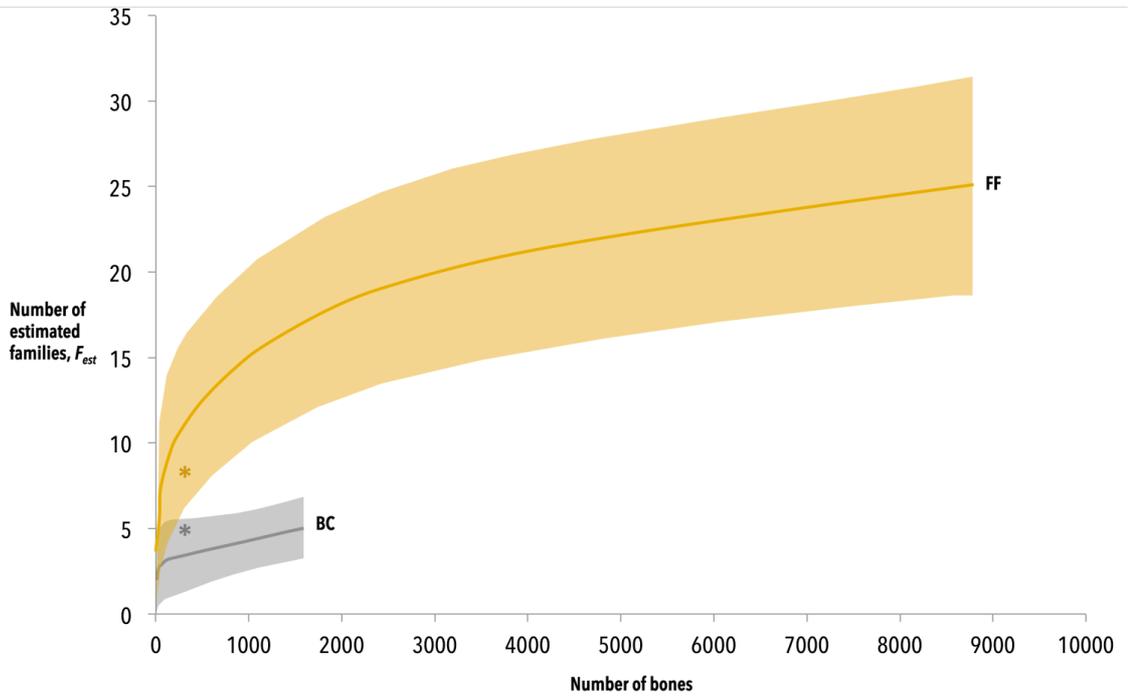
### 2.5.2 ACCURACY of THE BBM METHOD

In order to assess how accurately the taxa in the bulk bone assemblage that were identified genetically reflect the taxa in the fossil assemblage that were identified morphologically, we compared family biodiversity from FF and BC detected by the two methods. We found that the families identified via the BBM method are largely consistent with those identified through morphological analysis of the fossil record (Figure 2.2; Table S2.8.5). For both sites, the families identified molecularly are a subset of the families identified morphologically: for FF, 25 avian families were identified through morphology, with eight of these (32%) also identified through BBM, while for BC, five mammalian families were identified morphologically with all (100%) also identified through BBM. This is likely to be a consequence of sampling bias and sampling effort, DNA preservation, and the presence (or absence) of a reference sequence in GenBank (Figure 2.4). Several studies on plant family diversity, whether assessed using pollen and macrofossils, or sedimentary ancient DNA (*sedaDNA*) and metabarcoding, reported similar results (Jørgensen et al. 2011; Pedersen et al. 2013), but also suggest *sedaDNA* tends to reveal different and less diversity than these alternative methods (Boessenkool et al. 2013; Parducci et al. 2014). In this way the BBM method may be more reflective of the fossil record than *sedaDNA*, but this then raises the question: why weren't all the families found the fossil record detected by BBM?

Assuming that the abundance of each family in bulk bone sample is similar to the morphological sample, we would not expect rare families (families that make up less than 1% of the total number of fossils) to have been represented in a subsample total of 250 bone fragments. When we look at the percentage of families we *would* have expected to sample, we detected 100% of expected families in FF and BC (Table 2.1, Figure 2.2). As such, the BBM method is robust for detecting the most abundant families. Indeed, for both sites, the most abundant families (Anatidae and Muridae, from FF and BC, respectively) are the ones most consistently recovered. On the other hand, we detected families that were not expected to be detected due to their low abundance in the fossil record (Table 2.1); 10.5% of unexpected families were detected at FF (i.e., Dinornithidae and Accipitridae; Table S2.8.5), and 100% of

unexpected families detected at BC (i.e., Peramelidae, Potoroidae). This shows that the method is potentially sensitive enough to detect at least some rare taxa, but that this ability is sporadic. Furthermore, the diversity of the sample will influence the number of rare taxa present and the frequency of the rarest taxa, which may explain why half the families are shared between the morphological and molecular assemblages for FF (Sørensen index of similarity = 0.48; Table 2.1), while the morphological and molecular family assemblages are identical for BC (Sørensen index of similarity = 1; Table 2.1): the number of rare families in the more diverse FF (that also have a GenBank reference) is nine with the abundance of the rarest families (five) being 0.01%, whereas the number of rare families (that also have a GenBank reference) in the less diverse BC is two, with the abundance of the rarest families (two) being 0.06% (Table S2.8.5). This may also explain why a smaller percentage of unexpected families were detected at FF compared with BC.

In reality, the abundance of each taxon in a bulk bone sample may not be the same as the morphological sample, and the extent of this difference will differ between sites. For instance, bulk bone samples are subsamples of all the fossils collected that are in turn a subsample of the total fossils—as such, rare taxa in the fossil record can become rarer still in bulk bone sample, and with each successive subsampling the likelihood of rare species being lost increases. In addition, rare diagnostic bones may be removed from the bulk bone collection prior to DNA analysis, further biasing the sample towards abundant taxa (S2.8.13). Alternatively, some bones may be broken into several fragments, increasing their chance of being sampled, while others may not. On top of this, both DNA preservation that differs between bones, PCR bias (Figure 2.4), and deficiencies in reference databases (S2.8.13) will further affect the ability of taxa to be detected genetically, regardless of abundance. For example, post-mortem DNA degradation results in highly fragmented sequences that may not be long enough for both primers of the pair to bind and amplify, resulting in taxa with poor preservation not being detected. Again, this may explain why some rare families were detected molecularly and others were not, especially in FF where the amplified region was considerably longer than that of BC. Amplifying shorter metabarcoding regions, using multiple metabarcoding primers, using species-specific



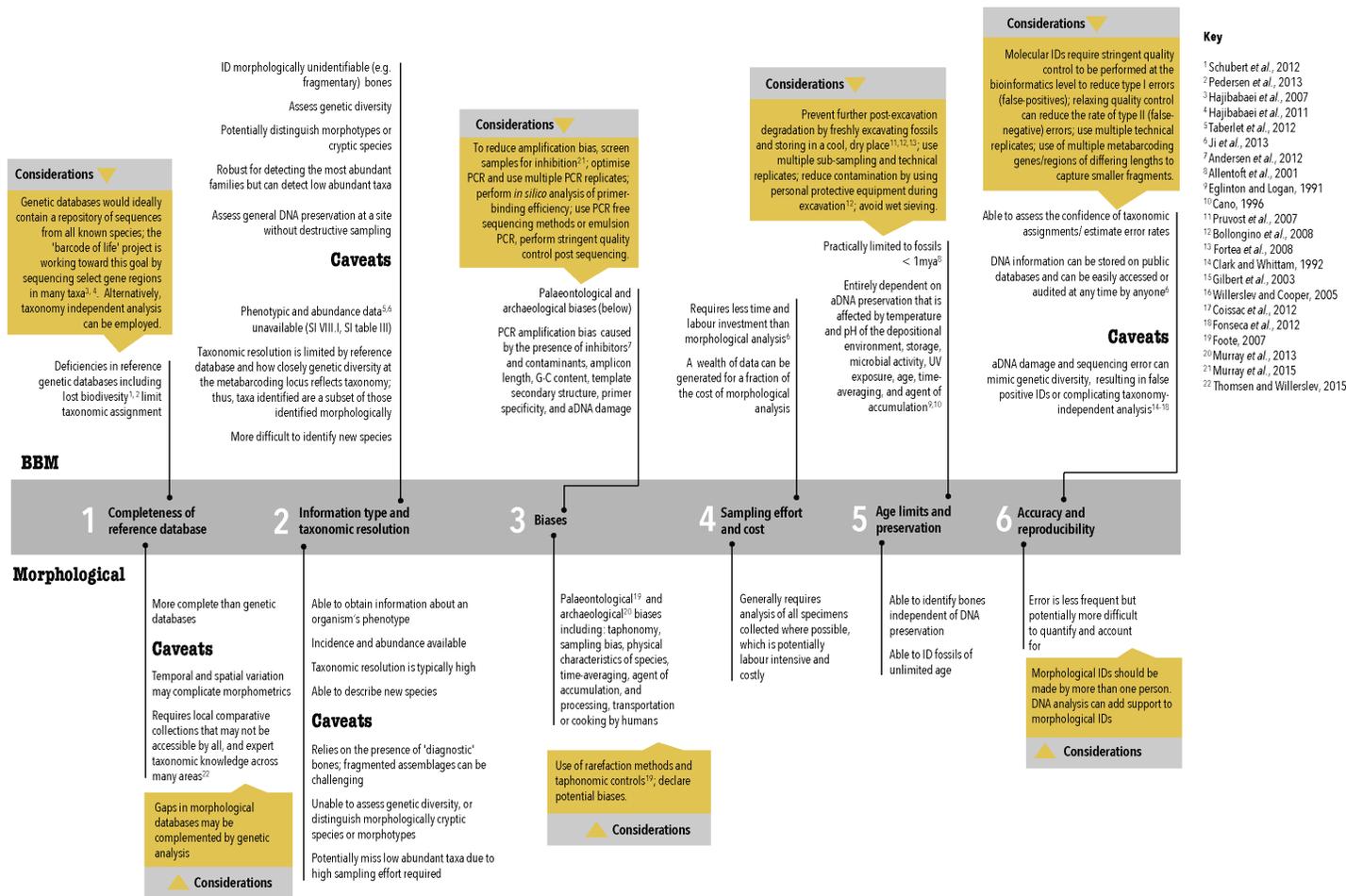
**FIGURE 2.3** | RAREFACTION CURVES SHOWING HOW MANY FAMILIES WOULD HAVE BEEN DETECTED IF FEWER BONES HAD BEEN SAMPLED MORPHOLOGICALLY FOR BOTH FF (YELLOW) AND BC (GREY). Shading represents 95% confidence intervals (*EstimateS*; Colwell, 2013). Asterisks represent how many families were detected through BBM of 250 bones from FF (yellow) and BC (grey).

primers and increasing the sampling effort may also help mitigate some of the biases in order to recover more taxa, more consistently (S2.8.13). However, there remains a trade-off between amplicon length and taxonomic discrimination, and the use of multiple primer sets and cost. Furthermore, although increasing the sampling effort will improve the repeatability and may result in the detection of more families (especially samples of high diversity), this will plateau as a consequence of the principle of diminishing returns; the point of sampling ‘saturation’ is best determined experimentally. In comparison to morphological methods, the same amount of sampling or less is required to detect the same number of families using bulk bone analysis. Rarefaction can be used to estimate the number of families that would have been identified morphologically had fewer bones (250) been examined morphologically (Figure 2.3; Foote and Miller 2007; Wooley et al. 2010). This was compared with the number of families detected by DNA analysis of the same number of bones. For FF,  $10.40 \pm 4.76$  (95% CI) families (23-61%) would have been detected if 250 bones were sampled as opposed to 25 families from a sample of 8771 bones—this is not significantly different from the percentage of families that were detected using the bulk bone method at the same level of sampling (32%, or eight out of 25 families; Figure 2.2a). For BC,  $3.31 \pm 1.15$  (95% CI) families would have been detected if 250 bones were sampled as opposed to five families from 1600 bones (43-90%); in comparison, we detected all of the families from BC via DNA analysis of 250 bulk bones (100%, or five out of five families; Figure 2.2b). In this way, the BBM method can generate as much data as morphological approaches with the same amount of sampling, potentially in a fraction of the time.

### **2.5.3 OVERALL UTILITY *and* FURTHER APPLICATIONS *of* THE BBM METHOD**

Although morphological identifications can typically achieve higher taxonomic resolution (species) than bulk bone metabarcoding can (family/genus level) at present (especially given the incomplete nature of the reference genetic databases (S2.8.13; Figure 2.4), we were able to confidently make some species-level assignments. Here, we define a species-level match as >98% similarity to a reference sequence across 100% of the amplicon length, with all other species within the genus present in GenBank, and equal similarity to no other species. For FF, we detected the extinct species *Dinornis robustus* (Giant moa), and the highly endangered species

*Strigops habroptilus* (Kākāpō), while for BC, we detected the vulnerable and locally extirpated species, *Phascogale tapoatafa* (Brush-tailed phascogale; Table S2.8.5). This shows that, like morphological methods and environmental DNA (eDNA) metabarcoding methods, the BBM approach has the potential for monitoring changes in biodiversity over time, which may be important for the management of already critically endangered species, or for identifying species that may have future risk of extinction (Boessenkool et al. 2013; Bohman et al. 2014; Pansu et al. 2015). Conservation can potentially also be aided by estimates of genetic diversity within species (McDowell 2014) that can be obtained using BBM. For example, although there is only one species of Kākāpō within the genus *Strigops*, three distinct genetic variants (i.e., OTUs) (that cannot be accounted for by aDNA damage or error) at the locus studied were detected (Table S2.8.5), and it is possible that this indicates the presence of higher historical genetic diversity, or even several unknown sub-species of *Strigops*. However, because metabarcodes are designed to minimise intra-specific variability, high-resolution estimates of genetic diversity within species (such as haplotype diversity) need to be obtained through the amplification of several specific molecular markers. Nevertheless, it would not be possible to glean this type of genetic information using morphological approaches, and shows how the BBM method, like eDNA metabarcoding methods, could also potentially help detect taxa that have not been formerly described in the fossil record (Krause et al. 2010; Murray et al. 2013; Parducci et al. 2014). This could assist in more accurately determining the timing extinction events (Haile et al. 2009; Pedersen et al. 2013; Willerslev et al. 2015) or species' former distributions. Similarly, for BC, we detected a taxon with 98% similarity to several species of bandicoot belonging to the genus *Isoodon*; however, this genus was not identified in the morphological analyses, which may indicate additional diversity within Peramelidae. Alternatively, the *Isoodon* taxon may represent the indeterminate bandicoot species identified morphologically but assigned to the genus *Perameles* (Table S2.8.5). This interpretation could be supported by the fact that several *Perameles* voucher specimens have references in GenBank but were not the closest match, and only one OTU was identified for this family. In this way, genetic analysis may add weight to or challenge morphological identifications, particularly if specimens are morphologically ambiguous (e.g., *Macropus fuliginosus* and *M. giganteus*, or



**FIGURE 2.4 | A COMPARISON OF THE STRENGTHS AND LIMITATIONS OF BOTH MORPHOLOGICAL AND MOLECULAR TAXONOMIC IDENTIFICATION OF FOSSILS. Detailed discussion of these caveats can be found in the Supplementary Information (S2.8.13).**

*Pseudochirus peregrinus* and *P. occidentalis*), cryptic species (using high resolution molecular markers), or tiny post-cranial bones that are typically difficult to distinguish at the species level (e.g., fish, reptiles, amphibians). Finally, the BBM method (as well as the use of species-specific primers on bulk bone samples) may offer a more sensitive tool for detection of invasive species (Bohmann et al. 2014), which may improve estimates of the timing of major historical events that have impacted biodiversity, such as human arrival, or the presence or absence of land bridges (e.g., land bridges connecting Kangaroo Island to the Australian mainland) (Pansu et al. 2015).

## 2.6 CONCLUSION

---

As with morphological methods of identifying fossil bones, BBM has intrinsic biases that make it difficult to identify every taxon present in the fossil assemblage (Figure 2.4; S2.8.13). The biases and limitations of the BBM method need to be considered when designing experiments that seek to apply it (Figure 2.4; S2.8.13). Weighing up the advantages and disadvantages of the BBM method for the taxonomic identification of fossils relative to traditional morphological taxonomic identification, we propose some practical recommendations for palaeontologists, archaeologists, and geneticists who are eager to apply the BBM method (Figure 2.4; S2.8.14, S2.8.15). Essentially, it is important to collect and subsample bulk bone material following a predetermined, statistically-optimised experimental design with clearly-defined questions. Many limitations of the method can be overcome through greater subsampling and replication at the most error-prone stages in order to quantify and minimise the collective contribution of these biases to variance in the results. Therefore, it is advisable that, when possible, a small pilot study with a hierarchical experimental design be employed to experimentally determine to what extent, and where, such effort should be concentrated. Ultimately, the way the method is applied will depend on the desired level of resolution and the type of questions being addressed. There also remains room for development in shotgun sequencing (Taberlet et al. 2012; Smith et al. 2015; S2.8.16, Figure S2.8.5), targeted (including enrichment capture or the use of species-specific primers) (Shokralla et al. 2012; Taberlet et al. 2012; Thomson and Willerslev 2015; S2.8.16), and PCR-free approaches (Ji et al. 2013), as well as sample preparation protocols (S2.8.16; Table

S2.8.6). Although genetic analysis cannot currently achieve the taxonomic resolution achievable using morphological analysis, this will improve with access to more comprehensive DNA reference material in the future. Metabarcoded samples also allow many diverse taxa to be identified simultaneously, and in this way, BBM can be rapid, and less reliant on regional taxonomic expertise than morphological analysis. For these reasons, the BBM method holds great potential as cost-effective technique for identifying taxonomic assemblages, estimating genetic diversity, and assessing general aDNA preservation at a range of sites from bones that were once considered disposable. Furthermore, the BBM has applications for the detection of cryptic or morphologically undiscoverable taxa, historical faunal turnover, and population fragmentation, which may provide a powerful tool for species conservation (McDowell 2014). Although both morphological and molecular methods have limitations, a combination of these approaches gives greater insight into past biodiversity, as the limitations of one are complemented by the other. We believe this study will help bridge the gap between disciplines, and hope that future development of the aDNA methods continues to integrate paleontological, ecological, and genetic principles, in order to overcome the limitations that remain.

## 2.7 REFERENCES

---

- Allentoft ME, Heller R, Oskam CL, Lorenzen ED, Hale ML, et al. 2014. Extinct New Zealand megafauna were not in decline before human colonization. *Proceedings of the National Academy of Sciences* **111**: 4922-4927.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, et al. 2012. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* **21**: 1966-1979.
- Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**: 32-46.

- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**.
- Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, et al. 2007. The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLOS One* **2**.
- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, et al. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology* / edited by Frederick M. Ausubel et al. 0 19, Unit 19.1021.
- Boessenkool S, McGlynn G, Epp LS, Taylor D, Pimentel M, et al. 2013. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conservation Biology* **28**: 446-455.
- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Ecology and Evolution* **29**: 358-367.
- Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLOS Computational Biology* **9**.
- Bronk Ramsey C. 2010. OxCal, v. 4.1.7.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, et al. 2010a. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108**: 4516-4522.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

Colwell RK. 2013. EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. Available: [purl.oclc.org/estimates](http://purl.oclc.org/estimates).

Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. 2014. DNA metabarcoding and the cytochrome oxidase subunit 1 marker: not a perfect match. *Biology Letters* **10**.

Deméré TA, Berta A, McGowen MR. 2005. The taxonomic and evolutionary history of fossil and modern balaenopteroid mysticetes. *Journal of Mammalian Evolution* **12**: 99-143.

Donoghue MJ, Doyle JA, Gauthier J, Kluge AG, Rowe T. 1989. The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics* **20**: 431-460.

Earp MA, Rahmani M, Chew K, Brooks-Wilson A. 2011. Estimates of array and pool-construction variance for planning efficient DNA-pooling genome wide association studies. *BMC Medical Genomics* **4**.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194-2200.

Epp LS, Boessenkool S, Bellemain EP, Haile J, Esposito A, et al. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology* **21**: 1821-1833.

Epp LS, Gussarova G, Boessenkool S, Olsen J, Haile J, et al. 2015. Lake sediment multi-taxon DNA from north Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews* **117**: 152-163.

- Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, et al. 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research* **40**.
- Foote M and Miller AI. 2007. Principles of palaeontology, Third Edition ed. W. H. Freeman and Company, New York City, NY.
- Fusco DA. 2014. The biodiversity of the Holocene terrestrial mammal faunas of the Murray Mallee, South Australia. Biological Sciences (Honours) Unpublished Honours Thesis, Flinders University.
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451-5.
- Goecks J, Nekrutenko A, Taylor J, and The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**.
- Haile J, Froese DG, MacPhee RDE, Roberts RG, Arnold LJ, et al. 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 22352-22357.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLOS One* **6**.
- Haouchar D, Haile J, McDowell MC, Murray DC, White NE, et al. 2014. Thorough assessment of DNA preservation from fossil bone and sediments exavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews* **84**: 56-64.

- Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG. 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Curá, Neuquén Province, southwest Argentina. *Quaternary Research* **59**: 364-378.
- Hogg AG, Quan H, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Hunter AA, Macgregor AB, Szabo TO, Wellington CA, Bellgard MI. 2012. Yabi: An online research environment for grid, high performance and cloud computing. *Source code for biology and medicine* **7**.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* **16**: 1245-1257.
- Jørgensen T, Haile J, Möller P, Andreev A, Boessenkool S, et al. 2011. A comparative study of ancient sedimentary DNA, pollen and microfossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* **21**: 1989-2003.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40**.
- Kitchen RR, Kubista M, Tichopad A. 2010. Statistical aspects of quantitative real-time PCR experiment design. *Methods* **50**: 231-236.

- Krause J, Fu Q, Good JM, Viola B, Shunkov MV, et al. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**: 894-897.
- Kuehl RO. 2000. Design of Experiments: Statistical Principles of Research Design and Analysis 2nd Edition. Brooks/Cole Publishing Company, Pacific Grove, CA.
- Larson S, Jameson R, Etnier M, Fleming M, Bentzen P. 2002. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Molecular Ecology* **11**: 1899-1903.
- Macgregor S. 2007. Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *European Journal of Human Genetics* **15**: 501-504.
- Manos PS, Soltis PS, Soltis DE, Manchester SR, Oh S, et al. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Systematic Biology* **56**: 412-430.
- McDowell MC, Baynes A, Medlin G, Prideaux G. 2012. The impact of European colonization on the late-Holocene non-volant mammals of Yorke Peninsula, South Australia. *The Holocene* **22**: 1441-1450.
- McDowell M. 2014. Holocene vertebrate fossils aid the management and restoration of Australian ecosystems. *Ecological Management and Restoration* **15**: 58-63.
- Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M. 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research* **35**.
- Minamoto T, Yamanaka H, Takahara T, Honjo MN, Kawabata ZI. 2012. Surveillance of fish species composition using environmental DNA. *Limnology* **13**: 193-197.

- Murray DC, Pearson SG, Fullagar R, Chase BM, Houston J, et al. 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews* **58**: 135-145.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. 2013. Scrapheap Challenge: A novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific reports* **3**.
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS One* **10**.
- Pansu J, Giguet-Covex C, Ficetola GF, Gielly L, Boyer F, et al. 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Molecular Ecology* **24**: 1485-1498.
- Parducci L, Väiliranta M, Salonen JS, Ronkainen T, Matetovici I, et al. 2014. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philosophical Transactions Royal Society B* **370**.
- Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, et al. 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Reviews* **75**: 161-168.
- Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, et al. 2014. Ancient and modern environmental DNA. *Philosophical Transactions Royal Society B* **370**: 20130383.
- Porter TM, Golding GB, King C, Froese D, Zazula G, et al. 2013. Amplicon pyrosequencing late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale reproducibility. *Molecular Ecology Resources* **5**: 798-810.

- Rodríguez-Aranda JP and Calvo JP. 1998. Trace fossils and rhizoliths as a tool for sedimentological and palaeoenvironmental analysis of ancient continental evaporite successions. *Palaeogeography Palaeoclimatology Palaeoecology* **140**: 383-399.
- Rohland N, Reich D, Mallick S, Meyer M, Green RE, et al. 2010. Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLOS Biology* **8**: 12.
- Shapiro B and Hofreiter M. 2012. Ancient DNA: methods and protocols. Humana Press (Springer Science), New York.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**: 1794-1805.
- Smith O, Momber G, Bates R, Garwood P, Fitch S, et al. 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**: 998-1001.
- Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* **5**: 1-34.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**, 2045-2050.
- Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, et al. 2012. Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLOS One* **7**.

- Thomsen PF and Willerslev E. 2015. Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**: 4-18.
- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Willerslev E, Davison J, Moora M, Zobel M, Coissac E, et al. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**: 47-51.
- Wooley JC, Godzik A, Friedberg I. 2010. A primer on metagenomics. *PLOS Computational Biology* **6**.
- Zhang X, Shi GR, Gong Y. 2008. Middle jurassic trace fossils from the ridang formation in saija county, south Tibet, and their palaeoenvironmental significance. *Facies* **54**: 45-60.

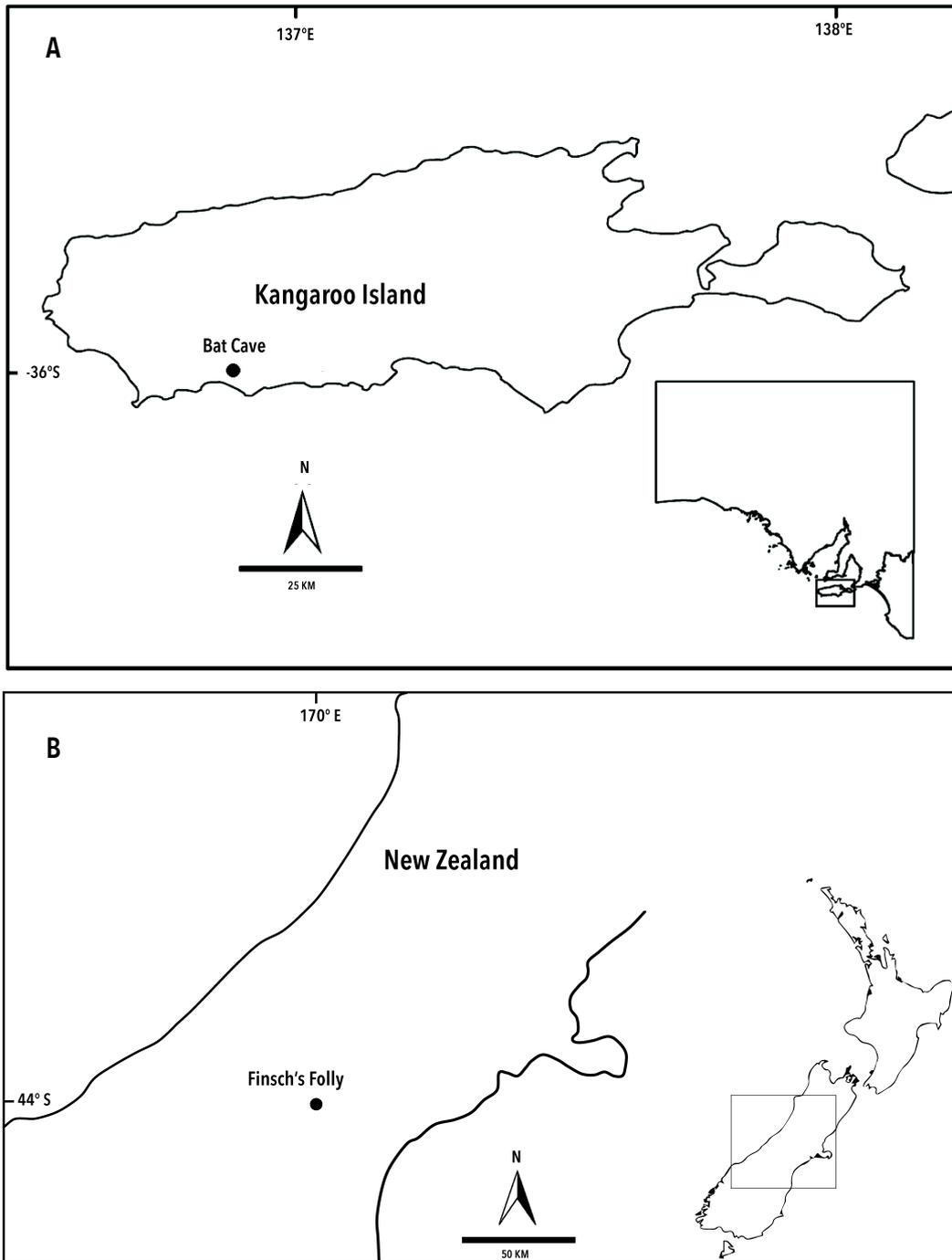
## 2.8 SUPPLEMENTARY INFORMATION

---

### S2.8.1 STUDY SYSTEMS *and* DATING

**BAT CAVE.** Bat Cave (BC; 35° 59' S, 136° 54' E; CEGSA No. 5K65) is a single chamber cave in the Kelly Hill Caves Conservation Park (Figure S2.8.1a) on the South West of Kangaroo Island (KI), South Australia. Located 13 km offshore to the Fleurieu Peninsula in Southern Australia, KI was connected to the mainland by a land bridge until rising sea levels caused separation around 8.9 thousand years ago (kyr) (Belperio and Flint, 1999). With 47% of its native bush land remaining, and being devoid of introduced foxes and rabbits, KI may be the best preserved native ecosystem in Southern Australia and carries a high conservation value (Robinson and Armstrong, 1999). The Kelly Hill Caves system is the result of solution cavities formed by horizontal groundwater movement in Pleistocene calcarenite dunes (Davies et al. 2002). A left humerus of *Ninox novaeseelandiae* weighing 0.96 g was selected and submitted to the Waikato radiocarbon laboratory for accelerator mass spectrometry (AMS) <sup>14</sup>C radio carbon dating (lab sample no Wk-36239; Fusco 2014). This was the only bone in the assemblage that met the weight requirement of AMS radio carbon dating. The sample was cleaned and ground, then decalcified in 2% HCl, rinsed and dried then gelatinised at pH = 3 with HCl at 90°C for four hr before being ultrafiltered and freeze-dried. Radiocarbon dating of a *Ninox novaeseelandiae* humerus and sternum returned an uncalibrated age of 2862 ± 25 BP. The age was calibrated to 3000-2750 cal BP (rounded to nearest 50 years) using the Southern Hemisphere calibration curve SHCal04 (Hogg 2013) with OxCal v.4.1.7 (Bronk Ramsey 2010).

**FINSCH'S FOLLY.** Finsch's Folly is a shallow (< 8 m) pitfall cave located in an isolated outcrop of limestone in southwestern Canterbury (Figure S2.8.1b).



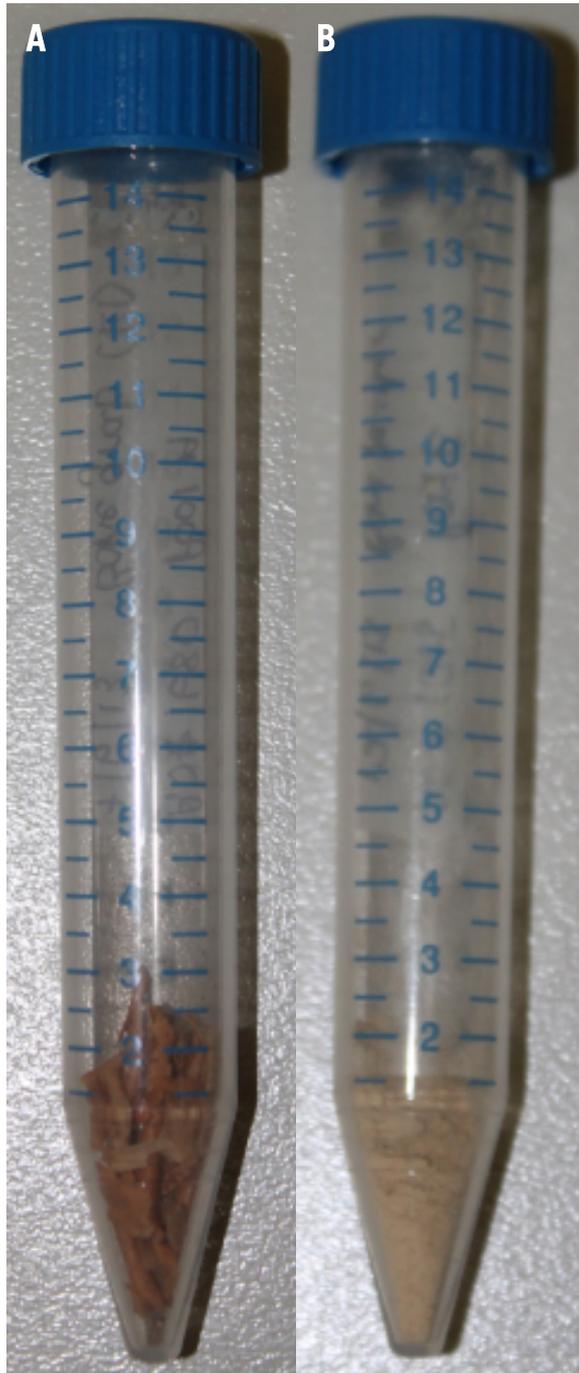
**FIGURE S2.8.1** | LOCATION AND SITE INFORMATION. **A** Bat Cave within Kelly Hill Conservation Park, Kangaroo Island, Australia (Fusco 2014), and **B** Finsch's Folly, Canterbury, New Zealand.

### **S2.8.2 EXPERIMENTAL DESIGN**

This type of experiment has been employed to determine the optimal design for genome-wide association studies that require DNA, or even whole blood, from many individuals to be pooled in order to reduce costs (Macgregor 2007; Earp et al. 2011). Like the BBM method, there are several levels in these protocols where experimental error can be introduced, namely, during pool construction and during genotyping on SNP arrays. Quantifying which steps contribute the most to variance successfully helped determine how many replicates at each step would be necessary to minimize variance; replication at steps that contribute significantly to variance is more important than replication at steps that contribute little to variance (Macgregor 2007; Earp et al. 2011). Macgregor (2007) found that for genome-wide association studies involving array-based DNA pooling, over 80% of variance in allele frequency was due to array variation while only 12% was due to pool construction. They suggest that “using larger numbers of arrays on smaller numbers of pools (with more individuals per pool) will be more effective than smaller numbers of arrays on larger numbers of pools” (Macgregor 2007). Like association studies, the optimal amount of replication required in the bulk bone method needs to be determined so that we have the power to detect statistically significant differences in biodiversity, but more than that, so that we may report how confident we are in those estimates.

### **S2.8.3 DNA EXTRACTION *of* BULK BONES**

Bulk bone subsamples were prepared in a designated clean facility at Murdoch University, WA, Australia. Coveralls, double gloves, mask, eyewear, and boots were worn as personal protective equipment, and to minimise contamination with exogenous DNA. Before beginning, all surfaces (including tray, table, chair, Dremel tool, drill bits, electronic balance, mortar and pestles, and gloves) were cleaned with a solution of 10% bleach, followed by a solution of 70% ethanol. All extractions and subsequent qPCR reactions were prepared in a separate ultra-clean room to avoid contamination. All post-PCR methods were performed in a physically separated laboratory in keeping with standard aDNA practice (Willerslev and Cooper 2005; Shapiro and Hofreiter 2012).



---

**FIGURE S2.8.2** | A SUBSAMPLE OF 50 BONES FROM THE BULK BONE SAMPLE OF MORPHOLOGICALLY UNIDENTIFIABLE BONES FROM BAT CAVE. **A** before grinding into bone powder, and **B**, after grinding into bone powder.

50 bones were selected at random from each bulk bone sample, and placed in a clean 15-50 ml Falcon tube (Figure S2.8.2a), after drilling off excess sediment. The mass of these bones were measured using an electronic balance (Scout Pro, *DHaus*), and were recorded. This was repeated for a total of five subsamples. Using an electronic balance, for each subsample, approximately 20 mg of each bone was placed into a mortar and pestle, or 20 mg of bone powder was drilled off into the mortar and pestle. These bones were then crushed into a fine bone powder and placed into a clean 15 ml Falcon tube for a final mass of 1 g bone powder (Figure S2.8.2b). A clean or new drill bit and mortar and pestle were used when beginning a new pool to avoid cross-contamination between subsamples. DNA from three, 100 mg aliquots of bone powder from each subsample was extracted. An extraction control was included per 10 extractions.

Bone digest buffer was prepared in a dedicated aDNA clean room, and contained final concentrations of 0.1% Triton-X-100 (14.85  $\mu$ l 10% Triton-X-100), 0.02 M Tris-HCl (29.7  $\mu$ l 1 M Tris-HCl), 1 mg/ml Proteinase K powder (1.49 mg), 0.01 M DTT (14.85  $\mu$ l 1M DTT), topped up to a final volume of 1.5 ml with 0.5 M EDTA. EDTA, Tris, DTT and Proteinase K were combined first and mixed, followed by the Triton-X-100. The buffer was incubated at 55°C with rotation for five min to activate the enzyme.

1.5 ml of digestion buffer was then added to 100 mg of bone powder. The Falcon tube was sealed with parafilm, and incubated in a rotating oven at 55°C for 24 hr. The following day, samples were removed from the oven and were mixed by vortexing. Samples were then centrifuged for two min at maximum speed in a bench top centrifuge. The supernatant was then transferred to a Vivaspin column, 500  $\mu$ l at a time, and centrifuged with the membrane facing outward for 10 min, or until concentrated to 50  $\mu$ l, at maximum speed in a bench top centrifuge. The concentrate was then transferred to a clean 1.5 ml microcentrifuge tube. Five volumes (250  $\mu$ l) of Buffer PB (*QIAGEN*, cat. No. 19066) was then added to the concentrate, and mixed by pipetting up and down. The 300  $\mu$ l mixture was then transferred to a QIAquick spin column (*QIAGEN* QIAquick PCR purification kit, cat. No. 28106), was allowed to sit at room temperature for two min, then centrifuged for one min at 13,000 rpm in a bench top centrifuge. After discarding the flow-through, 500  $\mu$ l of Buffer AW1

(*QIAGEN*, cat. No. 19081) was added to the column, allowed to incubate at room temperature for two min, and then column was centrifuged for one min at 13,000 rpm to wash the membrane. After discarding the flow-through, 500 µl of Buffer AW2 (*QIAGEN*, cat. No. 19072) was added to the column, let incubate at room temperature for two min, and then column was centrifuged for one min at 13,000 rpm to further wash the membrane. After discarding the flow-through, the column was centrifuged for a further one min at maximum speed to ensure all buffer had passed through the membrane. The column was then placed into a clean 1.5 ml microcentrifuge tube, and 50 µl of Buffer EB (*QIAGEN*, cat. No. 19066) was added to the centre of the membrane. This was allowed to sit at room temperature for five min, before being centrifuged for one min at 10,000 rpm to elute the DNA. Eluted DNA was then passed through the column again after letting incubate for an additional five min at room temperature. Eluted DNA was then transferred to a clean, labelled 1.5 ml microcentrifuge tube, and stored at -20°C.

#### **S2.8.4 qPCR QUANTIFICATION**

Serial dilutions of each extract (undiluted, 1/10 and 1/50) were amplified via qPCR using primers specific for a suitable barcoding region of the *16S rRNA* gene in mammals (Taylor 1996; 132 bp or 91 bp without primers) for the BC samples, and *12S rRNA* gene in birds (Cooper 1994; 274 bp or 232 bp without primers) for the FF samples (Table S2.8.1). The qPCR reactions consisted of final concentrations of: 0.4 mg/ml Fisher BSA (1 µl of 10 mg/ml), 1X ABI GeneAmp PCR Buffer (2.5 µl of 10X), 2 mM ABI MgCl<sub>2</sub> (2 µl of 25 mM), 0.4 uM IDT forward and reverse primers (1 µl of 10 uM, each), 0.25 mM Boline dNTPs (0.25 µl of 25 mM), 1-2 U of AmpliTaqGold DNA polymerase (0.25 µl), and 0.6 µl of 1:2000 SYBR green, made up to a final volume of 25 µl with GIBCO HPLC-grade water. qPCR reaction conditions (*Applied Biosystems*) were performed as follows: 95°C for five min, followed by 50 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 45 sec, followed by a +1 melt curve 95°C for 15 sec, 60°C for one min, 95°C for 15 sec, and finally a 72°C hold for 10 min. C<sub>T</sub>-values were recorded for each sample, and fluorescence and melt curves were analysed to give a rough indication of PCR efficiency. The dilutions exhibiting the least PCR inhibition were selected for subsequent fusion-tagging (S2.8.5). Cycle-threshold (C<sub>T</sub>) values were typically between 23-28 cycles

for the 16SMam1/2 primer set, and 30-35 cycles for the 12SAH primer set (data available from the authors upon request).

### **S2.8.5 NEXT-GENERATION SEQUENCING of DNA AMPLICONS**

For each subsample, three fusion-tag qPCRs were performed as above (S2.8.4) to generate sequencing amplicons of gene specific primers (above) flanked by unique indexing and Ion Torrent sequencing adapter (IT\_A and IT\_P1) sequences (Table S2.8.1) A positive control, extraction control, and three negative controls (water) were included, and if amplified, were carried through to sequencing.

Structure of fusion-tag primers:

Forward: IT\_A – Unique Index (7 bp) – Forward gene-specific primer (e.g. 12SA)

Reverse: IT\_P1 – Unique Index (7 bp) – Reverse gene-specific primer (e.g. 12SH)

The three replicate reactions per sample were combined. Based on the average amplification and melt curves of the replicates, amplicons for each sample were combined according to primer set (i.e., 16SMam, 12SAH) in approximately equimolar concentrations. Amplicon pools were purified twice by using an *Agencourt* AMPure XP PCR purification kit, and run on an agarose gel electrophoresis. Based on the relative band intensity of the products, the amplicon pools were further combined in approximately equimolar concentrations to create two sequencing libraries, one containing mammal *16S* amplicons for all BC subsamples, and the other containing avian *12S* amplicons for all FF subsamples.

Each sequencing library was purified two more times by using an *Agencourt* AMPure XP PCR purification kit. A standard dilution series, three negative controls, and serial dilutions of the sequencing library (1/10, 1/100, 1/1000, 1/5000, 1/10000, 1/25000, 1/50000, and 1/100000) were amplified in duplicate via qPCR using Ion Torrent primers corresponding to IT\_A and IT\_P1 (Table S2.8.1) to determine the number of template copies per  $\mu\text{l}$  in the sequencing library. The final concentrations of the qPCR reaction mix were: 1X *ABI* master mix (12.5  $\mu\text{l}$  of 10X), 0.4  $\mu\text{M}$  of

**TABLE S2.8.1** | METABARCODING PRIMER PAIRS USED, AND ION TORRENT ADAPTER SEQUENCES THAT FLANK THE METABARCODING PRIMERS FOR FUSION TAGGING. NNNNNNNN refers to the index sequence (any bases).

Name	Sequence (5'-3')	Target Taxa	Gene	Reference	Annealing Temp (°C)
12SA (Forward)	CTGGGATTAGATACCCCACTAT	Bird (uni) with 12SH	12S	(Cooper 1994)	57
12SH (Reverse)	CCTTGACCTGTCTTGTAGC	Bird (uni) with 12SA	12S	(Cooper 1994)	57
16Smam1 (Forward)	CGGTGGGGTGACCTCGGA	Mammal (uni)	16S	(Taylor 1996)	55
16Smam2 (Reverse)	GCTGTTATCCCTAGGGTAACT	Mammal (uni)	16S	(Taylor 1996)	55
IT_A-Index (Forward)	CCATCTCATCCCTGCGTGTCTC CGACTCAGNNNNNNN	-	-	-	-
IT_P1-Index (Reverse)	CCTCTATGGGCAGTCGGTG ATNNNNNNN	-	-	-	-

IT\_A primer (1  $\mu$ l of 10  $\mu$ M), 0.4  $\mu$ M of IT\_P1 (1  $\mu$ l of 10  $\mu$ M), and 2  $\mu$ l of library DNA in a total reaction volume of 25  $\mu$ l (8.5  $\mu$ l *GIBCO* HPLC-grade water). qPCR reaction conditions were as follows: 95°C for five min, 40 cycles of 95°C for 30 sec, 60°C for 40 sec, followed by a melt curve of 95°C for one min, 55°C for 30 sec, and 95°C for 15 sec.

Emulsion PCR and enrichment was performed by following the manufacturer's instructions for the Ion PGM Template OT2 200 kit (for the BC library) and the Ion PGM Template OT2 400 kit (for FF library). Sequencing was performed using the Ion PGM Sequencing 200 kit v2 and 400 kit as per the manufacturer's instructions.

### **S2.8.6 SEQUENCE IDENTIFICATION *and* BIOINFORMATICS ANALYSIS**

Sequence data in FastQ format that had passed the default filtering parameters was downloaded directly from the Ion Torrent server, and imported into Geneious v.7.0.6 (<http://www.geneious.com>, Kearse et al. 2012). Sequences were sorted by barcode by selecting the "separate reads by barcode" option from the drop-down menu "Sequence". Next, a file was created for each of the primer sequences used by selecting the "new sequence" option from the drop-down menu "Sequence" and pasting in the sequence and name. Primer sequences were annotated on each read by selecting the "find annotations" option from the drop-down menu "Annotate and Predict". Reads containing both primer sequences allowing no mis-match were extracted by selecting "extract annotations" from the drop-down menu "Tools". The primer name was inserted into the space "annotation name" "contains", and selecting "OK". Adapter sequences, barcodes, and primers sequences were removed from the reads by selecting the "trim ends" option under the drop-down menu "Annotate and Predict". Next, reads were sorted by length by right-clicking and selecting "sort by length"; sequences under 25 bp were deleted. This file was re-named by barcode and the word "trimmed", and sequences were batch-renamed to include the file name by selecting "batch rename" from the drop-down menu "Edit". Each file was then exported in FASTQ format.

The trimmed FASTQ files were then imported into the open-source, web-based bioinformatics platform Galaxy ([usegalaxy.org](http://usegalaxy.org); Giardine et al. 2005; Blankenberg et

al. 2010; Goecks et al. 2010). “FastQ groomer” was used to convert the quality score encoding from Sanger to ASCII. Sequences with an average quality score less than 25 were removed using “Generic FastQ manipulation” -> “Filter FastQ by quality score and length” by selecting “add new quality filter on a range of bases”, then “define base offsets as a percentage of read length”, followed by “aggregate read score for specified range” as “mean of scores”, and finally “keep read length when aggregate score is greater than or equal to 25”. Next, sequences were removed by applying quality cut-offs for a certain percentage of bases in the sequence using “FASTX-toolkit for FastQ data”: first, the “quality cut-off value” was set to 10 under “filter by quality”, followed by “percent of bases in sequence that must have quality equal to/higher than cut-off value” as 100 to keep only those sequences that contained all bases with a quality score of higher than 10; next, the cut-off value was set to 15 and the percentage of bases to 98 to keep only those sequences where at least 98% of bases had a quality higher than 15; and finally, the cut-off value was set to 20 and the percentage of bases to 90 to keep only those sequences where at least 90% of bases had a quality higher than 20 .

These filtered sequences were exported from Galaxy and imported in Geneious. From Geneious, they were batch exported in FASTA format (.fa). Using the open-source software QIIME v.1.8.0 (qiime.org; Caporaso et al. 2010b) run in VirtualBox, chimeric sequences and low-abundant sequences (<0.1% of the total number of sequences) were filtered out of each file using USEARCH v.6.1 (Edgar 2010; Edgar et al. 2011) in QIIME. In a terminal window under QIIME, the total number of sequences in a file were counted by typing the command `count_seqs.py -i [input file pathway/ input file name.fna]`. To identify chimeric sequences, the following command was used: `identify_chimeric_seqs.py -i [input file name.fna] -m usearch61 -suppress_usearch61_ref -o [output directory name]_directory/`. To remove chimeric sequences, the following command was used: `filter_fasta.py -f [original input fasta file.fna] -o [output file e.g., chimera_filtered.fna] -s chimera.txt -n`. To dereplicate, the following command was used: `pick_otus.py -m usearch61 -i [chimera_filtered.fna] -o [output2 directory name]_directory/ -s 1`. The output was then opened in *MatLab* R2011b v.7.13.0.564 (2011), and all clusters present in an abundance of less than 0.1% of the total number of sequences were identified. The file containing low abundant sequences was opened again in QIIME. To remove low

abundant sequences, the following command was used: `filter_fasta.py -f [chimera_filtered.fna] -o [output file name e.g., tets_ndr.fasta] -s [name of the MatLab output e.g., otu_filtered14.txt]`. Any sequences left in the PCR no-template control samples were removed from all other samples if present; however, these were entirely human.

The final output FASTA file was uploaded to the online bioinformatics workflow YABI ([ccg.murdoch.edu.au/yabi](http://ccg.murdoch.edu.au/yabi); Hunter et al. 2012) in order to obtain taxonomic assignments for the sequences by comparison to NCBI's GenBank. Files were uploaded to YABI by selecting the file inside the option "select file" in the drop-down menu "Design". Under "blast", the out-put file name was changed to something identifiable. The "run name" was changed, and the BLAST was run, using the search database `blastn` with the parameters: no low complexity filter, gap penalties existence of 5 and extension 2 (default), e-value  $<1e-10$ , word size 7 (default is 11). After the job was complete, the blast file was downloaded from the "blast" option under the drop-down menu "Jobs". Once downloaded, the file extension was changed from `.bls` to `.blast` and stored in a designated folder.

Next, blast files and FASTA files were imported into the software MEGAN v.4.70.4 ([ab.inf.uni-tuebingen.de/data/software/megan4](http://ab.inf.uni-tuebingen.de/data/software/megan4); Huson et al. 2007) in order to more easily visualise the BLAST output files and judge the credibility of taxomic assignments. Files were imported by selecting "import from BLAST" in the drop-down menu "File". From the browser, the blast file created above was selected, and under the tab "LCA parameters", 2 was chosen for minimum support, 35 was chosen for minimum score, and 9 was chosen for top percent. No minimum complexity filter was selected, or percent identity filter. Once the tree was created, the tree was collapsed at the family taxonomic level by selecting the option "collapse at taxonomic level" under the drop-down menu "Tree".

"Sequences with percentage similarity scores below 90% were discarded" (Murray et al. 2013). Families were considered present if the at least 95% of the query aligned to a known member of the family with a similarity of 90% or more. Highly credible species assignments were made only if alignments were  $> 98\%$  identity to the reference, equal in similarity to no other species, all other species of the genus

contain a reference sequence in GenBank, and the alignment was made across 100% of the query. A family assemblage matrix (presence 1, absence 0) was constructed in Microsoft's *Excel* for statistical analysis (S2.8.7). The sequences belonging to each family in a sample were extracted from the node of the tree as a FASTA file. For each sample, the FASTA files were imported into Geneious, where sequences from different samples that belonged to the same family were grouped (e.g., Muridae sequences from all BC samples were grouped). OTU analysis was carried out on each family separately. The file for each family was exported as FASTA file and imported into QIIME. To identify OTUs, sequences were clustered at 97% using USEARCH: `pick_otus.py -m usearch61 -i [chimera_filtered.fna] -o [output2 directory name]_directory/ -s 0.97`. This file was opened in Microsoft's *Excel* and used to create an OTU assemblage (presence 1, absence 0) matrix for further statistical analysis (S2.8.7).

### **S2.8.7 STATISTICAL ANALYSIS**

Sørensen's index of similarity was calculated by hand using the formula:

$$SI = 2C / (A + B) \quad \text{Sørensen (1948)}$$

Where C is the number of families shared by the morphological and molecular assemblages, A is the number of families in the morphological data set, and B is the number of families in the molecular data set.

To analyse the factors that contribute to variance in the presence/absence of families and OTUs detected using the BBM method, a nested non-parametric multivariate analysis of variance (nested NPMANOVA or PERMANOVA) was used. This method commonly employed for comparing community composition (or 'assemblage' or 'species diversity', etc.) in ecology (Anderson 2001). In this method, data is collected on the species present within treatments (here, this refers to biological and experimental replicates). The test then compares the variance in community composition within treatments to the variance in community composition between treatments (Anderson 2001). A non-parametric test is preferable over traditional parametric tests (e.g., ANOVA, MANOVA) because parametric tests

“rely on assumptions that are not generally met by ecological data sets” (Anderson 2001). NPMANOVA “allows direct partitioning of variance in complex models [such as nested designs]...while maintaining the flexibility and lack of formal assumptions of other non-parametric methods” (Anderson 2001).

First, a dissimilarity matrix (e.g., Bray-Curtis/Sørensen, Adjusted Jaccard index, Morisita index; Chao et al. 2005) or distance matrix (e.g., Euclidean, ChiSq) is constructed (there are advantages and disadvantages to using each; Chao et al. 2006): each treatment is listed across the first row and down the first column, with dissimilarity indices between each treatment forming the matrix below the diagonal. Dissimilarity indices can be calculated from the raw data counts or presence/absence data. Here, we used the Jaccard dissimilarity index (also known as Jaccard distance), best for presence/absence data.

Next, the sums of squares ( $SS_{\text{within treatment}}$  and  $SS_{\text{between treatments}}$ ) is calculated from the dissimilarity indices, and the mean sums of squares is calculated as SS divided by degrees of freedom. An F-statistic is calculated based on the mean sums of squares ( $MS_{\text{between}}/MS_{\text{within}}$ ). Next, a distribution of F-statistics ( $F^*$ ) is generated by using a program that performs thousands of permutations of the data by shuffling around groups. The idea is that if there is no difference in community composition between treatments, the F-statistic obtained from the data would fall within a distribution of randomly generated F-statistics. A *p*-value is calculated as the proportion of  $F^* \geq F$ . This analysis can be done in R ([www.R-project.org](http://www.R-project.org); R Core Team, 2014) using the package ‘vegan’ (Oksanen et al. 2015) within BiodiversityR (Kindt and Coe 2005). Other software can also perform this analysis. Here, NPMANOVA was performed using the PERMANOVA program available: <https://www.stat.auckland.ac.nz/~mja/Programs.htm> (Anderson 2001). The variance components for each factor can be calculated using the mean sums of squares in the ANOVA table (Table S2.8.2).

**TABLE S2.8.2** | PARTITIONING VARIANCE COMPONENTS.  $\hat{\sigma}_{c(b)}^2$  is the variance between library preparations and sequencing runs within extract replicate,  $\hat{\sigma}_{b(a)}^2$  is the variance from extract replicate within pool, and  $\hat{\sigma}_a^2$  is the variance between pools in a nested ANOVA (can this be done the same way for NPMANOVA).

Source of variation	Degrees of freedom	Mean Square	Expected mean square
a (pools)	$x-1$	MSA	$\sigma_{c(b)}^2 + c\sigma_{b(a)}^2 + bc\sigma_a^2$
b within a (extracts within pools)	$x(y-1)$	MS(B/A)	$\sigma_{c(b)}^2 + c\sigma_{b(a)}^2$
c within b (sequencing run within extract)	$xy(z-1)$	MS(C/B)	$\sigma_{c(b)}^2$
Total	$xyz-1$		

The variance components are estimated by (using the MS from the ANOVA table):

$$\hat{\sigma}_{c(b)}^2 = MS(C/B)$$

$$\hat{\sigma}_{b(a)}^2 = \frac{MS(B/A) - MS(C/B)}{z}$$

$$\hat{\sigma}_a^2 = \frac{MSA - MS(B/A)}{yz}$$

Contribution of variance components to total variance:

$$\hat{\sigma}_{Total}^2 = \hat{\sigma}_a^2 + \hat{\sigma}_{b(a)}^2 + \hat{\sigma}_{c(b)}^2$$

$$\hat{\sigma}_a^2 / \hat{\sigma}_{Total}^2 \times 100 = \% \text{ contribution of pool construction to total variance}$$

$$\hat{\sigma}_{b(a)}^2 / \hat{\sigma}_{Total}^2 \times 100 = \% \text{ contribution of extract to total variance}$$

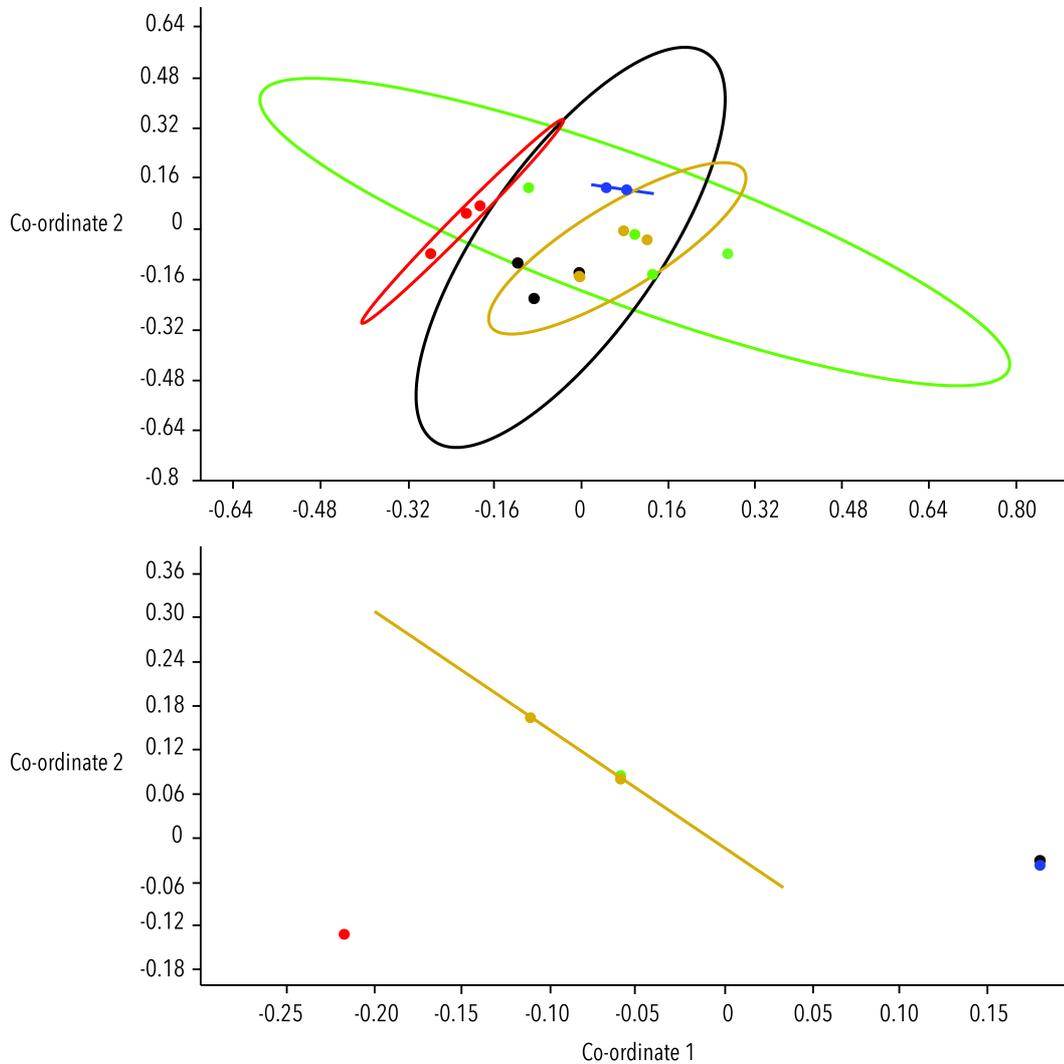
$$\hat{\sigma}_{c(b)}^2 / \hat{\sigma}_{Total}^2 \times 100 = \% \text{ contribution of library preparation and sequencing run to total variance}$$

### **S2.8.8 DOES FOSSIL ABUNDANCE CORRELATE WITH READ ABUNDANCE?**

Spearman's rank correlation coefficient was calculated (in *Microsoft's* Excel) to assess whether there was a correlation between morphological family abundance (the relative % of bones obtained per family), and molecular family 'abundance' (the relative % of reads per family) for both Bat Cave and Finsch's Folly (Table S2.8.3). No correlation was observed between family fossil abundance and read abundance for either site ( $p > 0.15$ ,  $\alpha = 0.05$ ). This shows that read abundance does not reflect fossil abundance. Although this assumes that the abundances of the taxa in the bulk bone sample are similar to those in the morphologically identifiable collection, we can never know what the abundances of each taxa are in the bulk bone sample; thus, this is the best approximation.

**TABLE S2.8.3** | SPEARMAN’S RANK CORRELATION COEFFICIENT,  $\rho$ , BETWEEN FAMILY FOSSIL ABUNDANCE (% OF THE FOSSIL RECORD) AND READ ABUNDANCE (% OF READS) FOR BOTH BAT CAVE AND FINSCH’S FOLLY.  $n$  is the number of observations.

Site	$n$	$\rho$	T statistic	$p$ -value
Bat Cave	5	0.60	1.30	0.28
Finsch's Folly	8	0.55	1.60	0.16



**FIGURE S2.8.3** | NON-METRIC MULTIDIMENSIONAL SCALING (NMDS) PLOT SHOWING THE MAIN CONTRIBUTIONS TO VARIANCE IN FAMILY RICHNESS. **A** Finsch's Folly, and **B** Bat Cave. Generated in PAST (Hammer et al. 2001). Different colours represent different pools, and dots correspond to different extracts nested within pools. Ellipses represent 95% Confidence Intervals.

### **S2.8.9 NMDS ANALYSIS OF THE MAIN CONTRIBUTIONS TO VARIANCE IN FAMILY RICHNESS**

Non-metric multidimensional scaling (NMDS) in PAST (Hammer et al. 2001) was used to visualize the components of variance (Wooley et al. 2010; Table S2.8.3).

For FF, the 95% CI for all pools overlap, indicating there that there is little difference family richness between pools (i.e., pool variance contributes little to the total variance; Table S2.8.3). However, there is more of a spread within each pool, indicating that the variance within pools (extract replicates) contributes more to the total variance. Conversely, for BC, the 95% CIs for pools don't overlap, indicating that there is a difference in family richness between pools and pool variance contributes to the total variance. However, there is no spread within each pool, indicating that variance within pools (extract replicates) do not contribute very little the total variance.

## S2.8.10 SEQUENCING COVERAGE

**TABLE S2.8.4** | SEQUENCING COVERAGE (NUMBER OF AMPLICON SEQUENCING READS) FOR EACH POOL, EXTRACTION, AND SEQUENCING REPLICATE. **A** families from FF, **B** families from BC, **C** OTUs from FF, and **D** OTUs from BC.

Gene	AD#	MB#	Pool	Extract	Number of reads/ family								Total
					Anatidae	Rallidae	Apterygidae	Dinornithidae	Emeidae	Strigopidae	Aptornithidae	Accipitridae	
12SAH	1545	2125	A	i	51	96	277	290	0	0	0	0	714
12SAH	1545	2125	A	i	649	347	870	704	0	0	0	0	2570
12SAH	1545	2126	A	ii	36	18	1118	93	0	0	0	0	1265
12SAH	1545	2126	A	ii	21	87	2381	0	0	0	0	0	2489
12SAH	1545	2127	A	iii	31	38	166	0	5	0	0	0	240
12SAH	1545	2127	A	iii	799	167	596	0	7	0	0	0	1569
12SAH	1545	2128	B	iv	69	0	931	268	11	7	0	0	1286
12SAH	1545	2128	B	iv	107	0	2501	0	43	84	0	0	2735
12SAH	1545	2129	B	v	77	0	1007	39	0	0	18	0	1141
12SAH	1545	2129	B	v	708	0	1338	32	0	0	19	0	2097
12SAH	1545	2130	B	vi	110	0	2426	0	0	8	0	0	2544
12SAH	1545	2130	B	vi	58	0	2256	0	0	134	0	0	2448
12SAH	1545	2131	C	vii	416	0	0	0	0	70	85	0	571
12SAH	1545	2131	C	vii	1607	0	0	0	0	369	419	0	2395
12SAH	1545	2132	C	viii	289	0	268	314	0	8	32	8	919
12SAH	1545	2132	C	viii	859	15	993	0	0	193	165	126	2351

12SAH	1545	2133	C	ix	208	0	14	1291	0	0	0	0	1513
12SAH	1545	2133	C	ix	1882	0	50	499	0	0	0	0	2431
12SAH	1545	2149	D	x	196	0	173	24	0	26	9	0	428
12SAH	1545	2149	D	x	1571	0	431	0	0	182	48	0	2232
12SAH	1545	2150	D	xi	47	0	982	0	0	288	0	0	1317
12SAH	1545	2150	D	xi	249	0	786	0	0	1506	0	0	2541
12SAH	1545	2151	D	xii	45	0	370	303	0	116	49	0	883
12SAH	1545	2151	D	xii	142	0	891	0	0	1037	505	0	2575
12SAH	1545	2152	E	xiii	41	0	2352	0	0	0	9	0	2402
12SAH	1545	2152	E	xiii	86	0	2157	0	0	0	84	0	2327
12SAH	1545	2153	E	xiv	59	0	2381	0	0	0	11	0	2451
12SAH	1545	2153	E	xiv	479	0	1800	4	0	0	14	0	2297
12SAH	1545	2154	E	xv	66	0	2078	0	0	0	89	0	2233
12SAH	1545	2154	E	xv	378	0	1961	0	0	0	141	0	2480

## B

Gene	AD#	MB#	Pool	Extract	Number reads/ family					Total
					Burramyidae	Dasyuridae	Potoroidae	Peramelidae	Muridae	
16SMam	1680	2134	A	i	166	0	22	91	4468	4747
16SMam	1680	2134	A	i	2	0	80	3	4658	4743
16SMam	1680	2135	A	ii	69	0	20	208	4445	4742
16SMam	1680	2135	A	ii	5	0	52	18	4666	4741
16SMam	1680	2136	A	iii	88	0	31	94	4479	4692
16SMam	1680	2136	A	iii	4	0	173	6	4561	4744
16SMam	1680	2137	B	iv	0	0	0	0	5251	5251
16SMam	1680	2137	B	iv	0	0	0	0	5231	5231
16SMam	1680	2138	B	v	0	0	0	0	4815	4815
16SMam	1680	2138	B	v	0	0	0	0	4767	4767
16SMam	1680	2139	B	vi	0	0	0	0	4625	4625
16SMam	1680	2139	B	vi	0	0	0	0	4745	4745
16SMam	1680	2140	C	vii	16	0	0	0	4785	4801
16SMam	1680	2140	C	vii	8	0	0	0	5232	5240
16SMam	1680	2141	C	viii	19	0	0	0	4776	4795
16SMam	1680	2141	C	viii	18	0	0	0	4778	4796
16SMam	1680	2142	C	ix	5	0	0	0	4785	4790
16SMam	1680	2142	C	ix	5	0	0	0	4786	4791
16SMam	1680	2143	D	x	35	0	0	0	4784	4819
16SMam	1680	2143	D	x	28	0	0	0	5228	5256
16SMam	1680	2144	D	xi	33	30	0	0	4754	4817
16SMam	1680	2144	D	xi	22	0	0	0	5234	5256
16SMam	1680	2145	D	xii	47	55	0	0	4715	4817
16SMam	1680	2145	D	xii	36	52	0	0	5166	5254
16SMam	1680	2146	E	xiii	0	0	0	0	4307	4307

16SMam	1680	2146	E	xiii	0	0	0	0	4739	4739
16SMam	1680	2147	E	xiv	0	0	0	0	4740	4740
16SMam	1680	2147	E	xiv	0	0	0	0	4739	4739
16SMam	1680	2148	E	xv	0	0	0	0	4740	4740
16SMam	1680	2148	E	xv	0	0	0	0	4739	4739



## D

Gene	AD#	MB#	Pool	Extract	Number of reads/ OTU/ family											Total
					Burramyidae	Dasyuridae	Potoroidae	Peramelidae	Muridae							
					1	1	1	1	1	2	3	4	5	6		
16SMam	1680	2134	A	i	166	0	22	91	1234	3215	12	7	0	0	4747	
16SMam	1680	2134	A	i	2	0	80	3	4430	192	36	0	0	0	4743	
16SMam	1680	2135	A	ii	69	0	20	208	1179	3248	9	0	9	0	4742	
16SMam	1680	2135	A	ii	5	0	52	18	4097	519	50	0	0	0	4741	
16SMam	1680	2136	A	iii	88	0	31	94	955	3499	7	6	12	0	4692	
16SMam	1680	2136	A	iii	4	0	173	6	3921	590	50	0	0	0	4744	
16SMam	1680	2137	B	iv	0	0	0	0	468	4783	0	0	0	0	5251	
16SMam	1680	2137	B	iv	0	0	0	0	1944	3211	76	0	0	0	5231	
16SMam	1680	2138	B	v	0	0	0	0	520	4276	19	0	0	0	4815	
16SMam	1680	2138	B	v	0	0	0	0	3061	1618	88	0	0	0	4767	
16SMam	1680	2139	B	vi	0	0	0	0	416	4198	11	0	0	0	4625	
16SMam	1680	2139	B	vi	0	0	0	0	4207	428	110	0	0	0	4745	
16SMam	1680	2140	C	vii	16	0	0	0	1927	2807	51	0	0	0	4801	
16SMam	1680	2140	C	vii	8	0	0	0	1724	3392	116	0	0	0	5240	
16SMam	1680	2141	C	viii	19	0	0	0	2398	2344	34	0	0	0	4795	
16SMam	1680	2141	C	viii	18	0	0	0	2004	2724	50	0	0	0	4796	
16SMam	1680	2142	C	ix	5	0	0	0	2475	2250	60	0	0	0	4790	
16SMam	1680	2142	C	ix	5	0	0	0	2119	2594	73	0	0	0	4791	
16SMam	1680	2143	D	x	35	0	0	0	924	3814	46	0	0	0	4819	
16SMam	1680	2143	D	x	28	0	0	0	911	4228	89	0	0	0	5256	
16SMam	1680	2144	D	xi	33	30	0	0	1041	3667	46	0	0	0	4817	
16SMam	1680	2144	D	xi	22	0	0	0	941	4203	86	0	0	4	5256	
16SMam	1680	2145	D	xii	47	55	0	0	1042	3629	44	0	0	0	4817	
16SMam	1680	2145	D	xii	36	52	0	0	1002	4066	95	0	0	3	5254	
16SMam	1680	2146	E	xiii	0	0	0	0	4210	0	0	97	0	0	4307	

16SMam	1680	2146	E	xiii	0	0	0	0	4730	9	0	0	0	0	4739
16SMam	1680	2147	E	xiv	0	0	0	0	4740	0	0	0	0	0	4740
16SMam	1680	2147	E	xiv	0	0	0	0	4737	2	0	0	0	0	4739
16SMam	1680	2148	E	xv	0	0	0	0	4740	0	0	0	0	0	4740
16SMam	1680	2148	E	xv	0	0	0	0	4737	2	0	0	0	0	4739

### S2.8.11 MORPHOLOGICAL TAXONOMIC IDENTIFICATION *and* ABUNDANCE DATA

**TABLE S2.8.5** | A COMPARISON OF THE FAMILY AND SPECIES-LEVEL TAXA DETECTED IN THE FOSSIL RECORD (MORPHOLOGICAL) AND BY THE “BULK BONE” METHOD FOR BOTH FF AND BC, RESPECTIVELY. ✗ indicates not detected/ not expected to be detected, ✓ indicates detected within similarity cut-off across at least 95% of the query, \* indicates a species with a highly credible taxonomic assignment (i.e., within the similarity cut-off, equal similarity to no other species, all other species have a reference sequence in GenBank, alignment made across at least 100% of the query). Similarity cut-offs for taxonomic IDs: species represent > 98% similarity, genus represents 96-97% similarity, and family represents 90-95% similarity. OTUs that were detected within families are listed. “Expected” refers to families that represent > 1% of the fossil record that are also represented in GenBank (by at least 1 of the species listed within) for the metabarcoding gene (either *12S* or *16S*) used.

A

Family	Genus	Species	Common name	Morphological	Relative abundance of identified specimens (%)	Molecular	# OTUs	12S region that we sequenced in GenBank	Family expected? (>1% of fossil record and in GenBank)
Anatidae				✓	47.5	✓	6		✓
	<i>Chenonetta</i>	<i>finschi</i>	Finsch's duck	✓	45.98	✗		✗	
	<i>Cnemiornis</i>	<i>calcitrans</i>	New Zealand goose	✓	1.34	✗		✗	
	<i>Anas</i>	<i>chlorotis</i>	Brown teal	✓	0.05	✗		✗	
		<i>superciliosa</i>	Pacific black duck	✓	0.05	✗		✗	
	<i>Hymenolaimus</i>	<i>malacorhyncos</i>	Blue duck	✓	0.05	✗		✓	
	<i>Tadorna</i>	<i>variegata</i>	Paradise shelduck	✓	0.03	✗		✗	
Strigopidae				✓	8.89	✓	3		✓
	<i>Strigops</i>	<i>habroptilus</i>	Kākāpō	✓	8.89	✓*		✓	
Apterygidae				✓	9.44	✓	13		✓

	<i>Apteryx</i>	spp.	Unknown kiwi	✓	9.44	✓		?	
Rallidae				✓	6.18	✓	5		✓
	<i>Gallirallus</i>			✓	4.55	✓		✓	
		<i>australis</i>	Weka	✓	4.54	X		✓	
		<i>phillipensis</i>	Buff-banded rail	✓	0.01	X		✓	
	<i>Gallinula</i>	<i>hodgenorum</i>	Hodgen's waterhen	✓	1.30	X		X	
	<i>Fulica</i>	<i>prisca</i>	New Zealand coot	✓	0.14	X		X	
	<i>Porphyrio</i>	<i>hochstetteri</i>	Takahe	✓	0.15	X		✓	
		<i>melanotus</i>	Pukeko	✓	0.04	X		X	
Callaeidae				✓	0.74	X			X
	<i>Callaeas</i>	<i>cinerea</i>	Kōkākō	✓	0.71	X		X	
	<i>Philesturnus</i>	<i>carunculatus</i>	Saddleback	✓	0.03	X		✓	
Sturnidae				✓	0.14	X			X
	<i>Sturnus</i>	<i>vulgaris</i>	Common starling	✓	0.14	X		✓	
Aptornithidae				✓	2.45	✓	4		✓
	<i>Aptornis</i>			✓	2.45	✓		✓	
		<i>defosser</i>	South Island adzebill	✓	2.45	X		✓	
Strigidae				✓	0.43	X			X
	<i>Sceleglaux</i>	<i>albifacies</i>	Laughing owl	✓	0.42	X		X	
	<i>Ninox</i>	<i>novaezeelandiae</i>	Southern boobook	✓	0.01	X		✓	
Aegothelidae				✓	0.06	X			X
	<i>Aegotheles</i>	<i>novaezealandiae</i>	Owelt-nightjar	✓	0.06	X		X	
Phasianidae				✓	0.06	X			X
	<i>Coturnix</i>	<i>novaezealandiae</i>	New Zealand quail	✓	0.06	X		X-	
Psittaculidae				✓	0.07	X			X
	<i>Cyanoramphus</i>	spp.	Parakeets	✓	0.07	X		X	
Dinornithidae				✓	0.54	✓	2		X
	<i>Dinornis</i>	<i>robustus</i>	Giant moa	✓	0.54	✓*		✓	
Emeidae				✓	2.54	✓	2		✓

	<i>Euryapteryx</i>	<i>curtus</i>	Coastal moa	✓	0.05	X	X	
	<i>Anomalopteryx</i>	<i>didiformis</i>	Bush moa	✓	0.03	X	✓	
	Other Emeid	spp.	?	✓	2.46	?	?	
Meliphagidae				✓	0.05	X		X
	<i>Prosthemadera</i>	<i>novaeseelandiae</i>	Tui	✓	0.04	X	X	
	<i>Anthornis</i>	<i>melanura</i>	New Zealand bell bird	✓	0.01	X	✓	
Acanthisittidae				✓	0.07	X		X
	<i>Traversia</i>	<i>lyalli</i>	Stephen's Island wren	✓	0.04	X	X	
	<i>Pachyptichas</i>	<i>yaldwyni</i>	Stout-legged wren	✓	0.02	X	X	
	<i>Xenicus</i>	<i>gilivetrus</i>	New Zealand rock wren	✓	0.01	X	X	
Fringillidae				✓	0.01	X		X
	<i>Cardeulis</i>	<i>cardeulis</i>	European gold finch	✓	0.01	X	X	
Accipitridae				✓	0.08	✓	1	X
	<i>Circus</i>			✓	0.08	✓	✓	
		<i>teauteensis</i>	Eyle's harrier	✓	0.08	X	X	
Scolopacidae				✓	0.02	X		X
	<i>Coenocorypha</i>	<i>iredalei</i>	South Island snipe	✓	0.02	X	X	
Columbidae				✓	0.01	X		X
	<i>Hemiphaga</i>	<i>novaeseelandiae</i>	New Zealand pigeon	✓	0.01	X	✓	
Mohouidae				✓	0.01	X		X
	<i>Mohoua</i>	<i>ochrocephala</i>	Yellowhead	✓	0.01	X	✓	
Petroicidae				✓	0.03	X		X
	<i>Petrocia</i>	<i>macrocephala</i>	Tomtit	✓	0.03	X	✓	
		<i>australis</i>	New Zealand robin	✓	0.01	X	X	
Procellariidae				✓	0.01	X		X
	<i>Procellaria</i>	sp.	Petrel	✓	0.01	X	✓	
Rhipiduridae				✓	0.01	X		X
	<i>Rhipidura</i>	<i>fuliginosa</i>	Grey fantail	✓	0.01	X	✓	
Halcyonidae				✓	0.01	X		X
	<i>Todiramphus</i>	<i>sanctus</i>	Sacred kingfisher	✓	0.01	X	✓	

Oriolidae				✓	0.05	X		X
	<i>Turnagra</i>	<i>capensis</i>	South Island piopio	✓	0.05	X	X	
Other taxa				✓	20.62			

B

Family	Genus	Species	Common name	Morphological	Relative abundance of identified specimens (%)	Molecular	# OTUs	16S in GenBank	Family expected? (>1% of fossil record and in GenBank)	
Muridae				✓	72.49	✓	6		✓	
		<i>Pseudomys</i>		✓	26.4	X		✓		
			<i>apodemoides</i>	Silky mouse	✓	0.24	X		X	
			sp. cf. <i>australis</i>	Plains mouse	✓	0.12	X		X	
			<i>occidentalis</i>	Western mouse	✓	24.05	X		X	
			<i>shortridgei</i>	Heath mouse	✓	0.30	X		X	
			sp. indet.	<i>Unknown native mouse</i>	✓	1.69	X		?	
		<i>Rattus</i>			✓	43.98	✓		✓	
			<i>fuscipes</i>	Bush rat	✓	37.76	X		✓	
			<i>lutreolus</i>	Swamp rat	✓	0.18	X		✓	
			sp. cf. <i>rattus</i>	Black rat	✓	0.36	X		✓	
			sp. indet.	Unknown rat	✓	5.68	X		?	
		<i>Mus</i>			✓	2.11	X		✓	
	<i>musculus</i>	House mouse	✓	2.11	X		✓			
Dasyuridae				✓	2.47	✓	1		✓	
		<i>Phascogale</i>		✓	0.12	✓		✓		
			<i>tapoatafa</i>	Brush tailed phascogale	✓	0.12	✓*		✓	
		<i>Sminthopsis</i>			✓	2.35	X		✓	
				Kangaroo Island						
			<i>aitkeni</i>	dunnart	✓	1.63	X		✓	
			<i>crassicaudata</i>	Fat tailed dunnart	✓	0.54	X		✓	
	sp. indet.	Unknown dunnart	✓	0.18	X		?			
Burramyidae				✓	21.57	✓	1		✓	
		<i>Burramys</i>		X	NA	✓		✓		
		sp.	Pygmy possum	X	NA	X		?		
	<i>Cercartetus</i>		✓	21.57	X		✓			

	<i>concinnus</i>	Western pygmy possum	✓	17.40	X		✓	
	<i>lepidus</i>	Little pygmy possum	✓	1.75	X		X	
	<i>nanus</i>	Eastern pygmy possum	✓	1.21	X		X	
	sp. indet	Unknown pygmy possum	✓	1.21	X		?	
Peramelidae			✓	0.06	✓	1		X
	<i>Isoodon</i>		X	NA	✓		✓	
	sp.	Bandicoot	X	NA	✓		?	
	<i>Perameles</i>	sp.	Unknown bandicoot	0.06	X		✓	
Potoroidae			✓	0.06	✓	1		X
	<i>Potorous</i>		✓	0.06	X		✓	
	<i>platyops</i>	Broad-faced potoroo	✓	0.06	X		X	
Other taxa			✓	3.35	NA	NA	NA	NA

### **S2.8.12 COMPARISON OF FF and BC FAMILY ASSEMBLAGES BY BBM**

To illustrate how one might statistically test the difference between faunal assemblages between sites using the experimental design employed here, we tested for a statistical difference between the family assemblages of FF and BC generated via the BBM method, using nested PERMANOVA (Anderson 2001) that included “Site” as an additional factor with “Pool” now nested within “Site” (data available on request). “Site” contributed 72% to variance ( $p$ -value = 0.0072,  $df$  = 1,  $\alpha$  = 0.05), while the other 28% is attributed to the other three factors (Pool(Site)+Extract(Pool)+Library preparation and Sequencing run(Extract(Pool)));  $p$ -value < 0.001,  $df$  = 59,  $\alpha$  = 0.05). Though this result may be obvious because FF and BC share no families in common (especially because we only focused on avian families in FF and mammalian families in BC), this type of analysis could be used to detect, with confidence, more subtle differences in family assemblages over time within a site, or between sites that are geographically close.

We also used the program Piface (Lenth 2006) to evaluate the power to detect a significant difference in family assemblage between the two sites at the 5% level by inputting the standard deviation for each factor calculated by PERMANOVA. Power was estimated to be 0.947. This indicates that the probability of a false negative (failure to detect a difference when one truly exists) occurring is 0.053. Quantifying the contributions to variance in a pilot study permits *post-hoc* assessment of the power that the design had to detect significant effects, allowing the experimenter to add further replicates to improve power of a larger study in the future.

### **S2.8.13 BIASES and LIMITATIONS**

Note that many of the biases and limitations of the BBM method are similar to those of environmental DNA metabarcoding methods, which have been addressed by several recent reviews (i.e., Bohmann et al. 2014; Pedersen et al. 2014; Rawlence et al. 2014; Thomsen and Willerslev 2015).

**TAPHONOMIC BIAS and SAMPLING BIAS.** As in palaeontological or archaeological studies, the conditions of the site will influence the preservation of bones within an

accumulation, and therefore, which species are likely to be identified. In addition to standard taphonomic biases (e.g., physical characteristics of species, agent of accumulation, etc.), additional biases apply to bulk bone samples in a way that is unique from the rest of the accumulation. For instance, morphologically identifiable bones may not be included for bulk bone analysis, which could bias the bulk bone sample. For example, larger animals, such as moa, might not be included in bulk bone analysis because they can be identified by virtue of their size. Therefore, the taxa within a bulk bone sample may not be a random or representative sample of the taxa in the entire accumulation, let alone all the taxa that were present in the past.

Taphonomy will also influence the preservation of aDNA (both quantity and quality), and this may vary at random both between and within bones in a bulk sample. For example, bones that have suffered gastric corrosion may have poor aDNA preservation having undergone a period of high acidity, or alternatively, may contain the DNA of both the predator and the prey. Therefore, some species may not be detected genetically despite leaving remains, while other species that did not leave remains may be detected. Other factors such as burial sediment type, microbial digestion, pH, temperature fluctuations of the burial environment, and exposure to water/ humidity, also affect the preservation of DNA (Allentoft et al. 2012). Post-mortem DNA degradation will influence the ability of the DNA to be amplified by PCR (see below).

In addition, DNA sampling will be inherently more complete at higher taxonomic levels: “this is a necessary consequence of the nesting of taxonomic groups within one another. Therefore, there will tend to be more individuals in any genus than in one of its component species, and more individuals in any family than in one of its component genera” (Foote and Miller 2007). As such, detection of taxa will always be better at higher taxonomic levels.

**COLLECTION and STORAGE BIAS.** Collection and storage post-excavation also have an affect on DNA preservation. Handling bones without gloves and wet sieving can introduce contamination that may swap the detection of endogenous DNA, while fluctuations in storage temperature can cause DNA to rapidly degrade. These issues can be simply mitigated with adequate planning (S2.8.14).

**INTRA- VERSUS INTER- SPECIES VARIATION IN BARCODING GENES.** Taxa identification is also influenced by choice of metabarcoding gene region (Taberlet et al. 2012), and the ability of these regions to resolve taxa is influenced by how much they vary within and between species. As such, there is not “a universal classification threshold for all genes at all taxonomic levels” (Liu et al. 2011). This means that, often, taxonomic assignment thresholds are arbitrarily determined. For instance, an initial screen of one bulk bone subsample from Bat Cave with mammalian-specific *12S rRNA* primers detected the presence of *Rattus rattus*, *Cercartetus*, and *Trichosurus vulpecula* based on 98% sequence similarity to the reference sequences (pers. comm.); however, these taxa were not detected with mammalian-specific *16S rRNA* primers at the same threshold (Table S2.8.5). Indeed, “it is not possible to unambiguously classify a specimen using a single barcode marker in the presence of variation shared among taxa (due to either ancestral polymorphism and/or introgression)” (Little 2011). Furthermore, choice of barcoding region will depend on the length of DNA fragments, which may vary between species within a sample, and will vary with age and preservation. It is therefore advisable that several different metabarcoding gene regions be targeted in order to further interrogate the samples at the species level. In addition, taxonomic assignment thresholds should be empirically determined for each primer set (Murray et al. 2015).

**DEFICIENCIES IN REFERENCE DATABASES.** Sequences obtained from DNA analysis of bulk bones need to be aligned with a reference sequence to obtain taxonomic assignments based the percent similarity between the query and reference sequence; if a reference sequence is unavailable, taxonomic assignments may not be made. Many species identified morphologically (e.g., *Cnemiornis calcitrans*) are not represented in the genetic reference database GenBank (Benson et al. 2006), making the genetic identification of many species impossible, particularly extinct species. Ideally, genetic databases such as GenBank should contain a repository of sequences from all known species. The ‘barcode of life’ project as well as others (see Thomsen and Willerslev 2015), are working toward this goal by sequencing, in many organisms, regions of genes that are able to distinguish taxa, such as mitochondrial genes (Hajibabaei et al. 2007; Hajibabaei et al. 2011); however, many standard barcoding genes such as the *cytochrome c oxidase subunit 1* gene are not suitable for

many metabarcoding projects (Deagle et al. 2014). As such, the absence of so many species in GenBank “would bias assignments at this taxonomic level” (Pedersen et al. 2013), making the incompleteness of GenBank a major limiting factor (Schubert et al. 2012). An example of this is the identification of *Emeus* in the family Emeidae at FF. This genus was detected genetically despite not being found in the fossil record, while its close relative that *was* found in the fossil record, *Euryapteryx*, was not detected; this may be because *Euryapteryx* is not represented in GenBank for *12S rRNA*. Such deficiencies in GenBank will only be overcome by concerted efforts to sequence a range of extinct and extant taxa several barcoding loci (Taberlet et al. 2012).

Alternatively, the use of ‘taxonomy independent’ analyses, such as OTU (operational taxonomic unit) or DTU (distance-based taxonomic unit) analysis can be used to assess species richness and diversity (Thomsen and Willerslev 2015). These methods involve assigning sequence variants to groups based on their identity. This allows us to overcome some of the constraints of GenBank. For instance, we estimated two OTUs within the family Emeidae; this may correspond to the two species identified in the fossil record, *Euryapteryx* and *Anomaloptyerx*, but we cannot be sure. Nevertheless, OTU richness often approximates the relative species richness within families; however, this depends on the taxonomic resolution of the barcode within families, and this may vary between families.

**ADNA DAMAGE, SEQUENCING, and POST-MORTEM DEGRADATION.** aDNA is generally riddled with damage, and sequencing introduces error, both of which will confound signals of evolutionary change (Clark and Whittam 1992; Gilbert et al. 2003; Willerslev and Cooper 2005; Rawlence et al. 2014). It is difficult to distinguish between damage (such as cytosine to thymine deamination), sequencing errors, and true polymorphism, especially when a reference genome is lacking. Adequate sequencing coverage and stringent quality control at the bioinformatics stage (especially chimera and abundance filtering) becomes crucial because damage, contamination, sequencing error, and the formation of chimeras during PCR amplification can result in erroneous taxonomic assignments or inflate estimates of diversity, particularly when taking a taxonomy-independent approach (Coissac et al. 2012; Fonseca et al. 2012; Pedersen et al. 2014; Murray et al. 2015; Thomsen and

Willerslev 2015). Using multiple metabarcoding genes may help make more accurate taxonomic assignments, estimate genetic diversity and determine the true number of OTUs present. In addition, post-mortem DNA degradation results in fragmented DNA that may be too short to amplify, which could result in the amplification of only the longest fragments. However, there is a trade-off between capturing short sequences, and being able to assign taxonomy to them. The use of multiple metabarcoding primer sets that target different fragment lengths will therefore also help uncover more taxa.

**PCR BIAS.** The presence of inhibitors (such as humic acids; Andersen et al. 2012), contaminants, and other PCR amplification biases (such as primer binding bias) can prevent the genetic detection of taxa in a bulk bone sample, as with environmental DNA samples (Pedersen et al. 2014; Bohmann et al. 2014; Murray et al. 2015; Thomsen and Willerslev 2015). The impact of these biases depends on the quality and quantity of the target DNA, which is typically low and poor in ancient samples, as well as the specificity of primers used. For instance PCR will preferentially amplify high-quality (undamaged) DNA over low-quality DNA, and is sensitive enough to amplify even low-level contamination. Thus, the presence modern contaminants, or species with well-preserved DNA, may swamp the detection of ancient targets. Blocking primers may be used to reduce or inhibit the amplification of known contaminants (Rawlence et al. 2014). In addition, “the number of primer-template mismatches of each species [has] a disproportionate effect on the amplification efficiency” (Piñol et al. 2014); thus, there can be differential amplification, and therefore detection, of taxa within the sample. PCR reaction efficiency can also vary with amplicon length, G-C content, and template secondary structure meaning that the amount of product from independent PCR reactions may not be comparable. The combined effect of PCR bias has been investigated experimentally: D. Murray (*pers. comm.*) has detected taxonomic ‘dropout’ after sequencing a synthetic blend of DNA from known taxa; amplification biases resulted in the output not entirely reflecting the input. Initial screening of samples at different concentrations with unindexed primers via qPCR is essential for assessing inhibition and amplification efficiency in order to overcome these (Murray et al. 2015). The contribution of PCR bias to experimental variance can be investigated by tagging PCR replicates with their own unique indexes; however, this will significantly

increase the cost of the experiment. Although PCR bias is difficult to control, every effort should be made to assess and correct for these biases, including *in silico* analysis of primer-binding efficiency, experimental PCR optimization and replication, and post-sequencing quality control (Pruvost et al. 2008; Fonseca et al. 2012; Shokralla et al. 2012; Bohmann et al. 2014; Murray et al. 2015).

Furthermore, many studies have demonstrated that the number of sequences per taxa identified in molecular datasets do not correlate with species abundances estimated from standard biodiversity censuses (Ji et al. 2013). Similarly, for BBM, the PCR biases discussed above will result in some taxa being preferentially amplified over others; thus, sequence read abundance cannot be used as an accurate proxy for the proportion of each taxa within a DNA extract (Taberlet et al. 2012), let alone their relative or absolute abundance in the fossil record. For example, low-abundant taxa with good aDNA preservation may be preferentially amplified, resulting in higher read counts than more abundant taxa with poor aDNA preservation. Even small differences in the starting amount of DNA for each taxon in the pool will be amplified exponentially by PCR and result in large and disproportionate differences in read abundances that may not reflect fossil abundance. PCR-free sequencing methods, such as single-molecule real-time sequencing, or shotgun sequencing may also be able to overcome this; however, these methods remain to be tested with bulk bone material and carry with it their own limitations (S2.8.15).

#### **S2.8.14 PRACTICAL GUIDELINES FOR PALAEOLOGISTS and ARCHAEOLOGISTS**

The bulk bone metabarcoding method is complementary to morphological analysis, not an alternative. This method may be used to corroborate morphological taxonomic assignments with genetic evidence, help identify morphologically indistinguishable taxa, or detect taxa that may be absent from the fossil record. However, it has been found that handling and storage methods can cause as much DNA degradation after excavation as occurred during the entire length of burial; as such, freshly excavated fossils yield more DNA than those that have been stored in less than ideal conditions (Pruvost et al. 2007). Guidelines have also been proposed for the excavation and subsequent storage of specimens for aDNA analysis (Pruvost et al. 2007; Bollongino

et al. 2008; Fortea et al. 2008; Rawlence et al. 2014); the most important considerations being the prevention of further DNA degradation by storage in a cool, dry place without major fluctuations in ambient temperature, and the prevention of contamination by excavating with personal protective equipment such as gloves and face masks (Bollongino et al. 2008). Therefore, to get the best results from bulk bone analysis, bulk bone samples should be collected ‘fresh’ (depending on the application, preferably as a separate subsample of the excavation), should not be handled without proper personal protective equipment (including disposable gloves, face mask, and coveralls), or wet sieved in the field (dry sieved samples should be placed directly into sterile airtight plastic bags or containers), and should be stored in a place of low humidity and temperature that will mimic the burial conditions. Sieving should then be carried out in a designated clean room facility with decontaminated sieves and ultrapure water. Morphologically identifiable fossils can be documented here; however, it is recommended that they be included for DNA analysis to reduce sampling bias. Although it may not be feasible to follow these in all cases, every effort should be taken to prevent further aDNA degradation post collection. The remainder of the excavated material can be treated more practically.

#### **S2.8.15 PRACTICAL GUIDELINES FOR GENETICISTS**

Bulk bone samples should be sieved and sorted in a designated clean room facility using decontaminated reagents and equipment. All downstream methods (such as DNA extraction) should be conducted in physically separated clean rooms with sterile reagents and equipment.

The key recommendation is to conduct a small hierarchical pilot study with minimal replication at each level of your chosen protocol to estimate the contributions to error variance. If this can be done, the number of bones or the amount of each bone pooled becomes less important because the variance will be accounted for, rather than controlled. A typical workflow may include: (1) extracting in triplicate DNA from three bulk bone subsamples and PCR amplifying a mitochondrial metabarcoding gene region to screen for overall aDNA preservation and PCR inhibition, as well as to optimise the PCR; (2) if amplification is successful, DNA from each extract may be indexed (fusion-tagged) in triplicate, triplicates combined (effectively ‘averaging

out' PCR variance), and sequenced in parallel to identify the taxa present, twice (note that PCR replicates may be individually tagged if the variance attributable to the PCR is to be quantified; however, this will triple the number of unique indexes required, and increase costs); (3) the differences within and between replicates can be used to evaluate sources of error, construct rarefaction curves, and estimate power in order to determine the level and extent to which replication needs to be concentrated in a larger study; (4) subsample bulk bone samples across time or space, extract, and sequence at the desired level of replication—we don't recommend extracting more than 10-20 subsamples (but probably no less than five) because if extraction replicates are also required the method quickly loses its cost-efficacy. Additionally, it is probably not necessary to replicate sequencing run; our results show that sequencing run on the Ion Torrent PGM contributes little variance to the data, and we expect this contribution will be even less on higher fidelity sequencing platforms such as *Illumina*; however, this will depend on the questions being asked, and as always, it is best to quantify the error of your own system. Independent replication of the entire protocol in another laboratory can be a good way to identify potential contaminants (Rawlence et al. 2014), as well as serve as additional subsamples of the bulk bone. Technical replicates at additional levels can be included if you are particularly concerned that those levels contribute significantly to the variance.

Our working recommendation is that no more than 50-100 bones are pooled because the probability of subsampling each bone from the powder will decrease as more bones are pooled, unless the number of subsamples are increased; however, this will likely also depend on the complexity (number of different taxa and their frequency) of the assemblage. Although pooling whole bones may not affect the detection of taxa, if bones are small (< 50 mg), we suggest powdering whole bones, whereas only 20 mg of powder from larger bones (> 50 mg) should be sampled. This is because larger bones may contain more DNA than smaller bones simply by virtue of their size. A good rule of thumb to follow would be that all bones should be of roughly the same mass if pooling whole bones. If possible, the surface of the bones should be cleaned with as much as possible by drilling and discarding the outer surface prior to powdering. The effect of bleaching the surface of bones for BBM is discussed in section S2.8.16. We recommend extracting between 100-200 mg of bone powder per

extraction (Caputo et al. 2013), based on our DNA extraction method. This may be optimised as desired.

We also recommend using several different metabarcoding genes to target different taxonomic groups, including broader groups (e.g., mammals) and more specific groups (e.g., marsupials), and primers that target longer and shorter fragments; this will increase the probability of detecting the full range of taxa present in the sample. Multiplexing PCRs in conjunction with NGS may be a cost-effective option that will allow you to prepare and sequence multiple amplicons in parallel; however, this may also introduce additional PCR bias and has not been tested thus far.

In addition, the use of both taxonomy and taxonomy-independent analyses (e.g., OTU analysis), with stringent quality control, will allow a better assessment of the biodiversity because it takes into consideration the sequences that don't have an orthologue in GenBank. This is particularly important when working with older samples that may contain extinct species. Another way the deficiencies in GenBank could be overcome is by creating a custom database of "voucher" sequences/OTUs for comparison between samples, or by sequencing metabarcoding genes in known single-source specimens to help "fill in" GenBank.

**TABLE S2.8.6** | ONE-WAY ANOVA TESTING WHETHER THERE IS A STATISTICALLY SIGNIFICANT EFFECT OF BLEACH TREATMENT ON  $C_T$ -VALUES FROM A QPCR. The test compares  $C_T$ -values of the neat (undiluted) DNA extract of bulk bone subsamples that have undergone no treatment prior to grinding with  $C_T$ -values from bleach treated bones (10% bleach for 10 min followed by two washes of ultrapure water) prior to grinding, and bleach treated bones after grinding. The asterisk indicates a significant treatment effect at the 5% level.  $n=3$  subsamples per treatment.

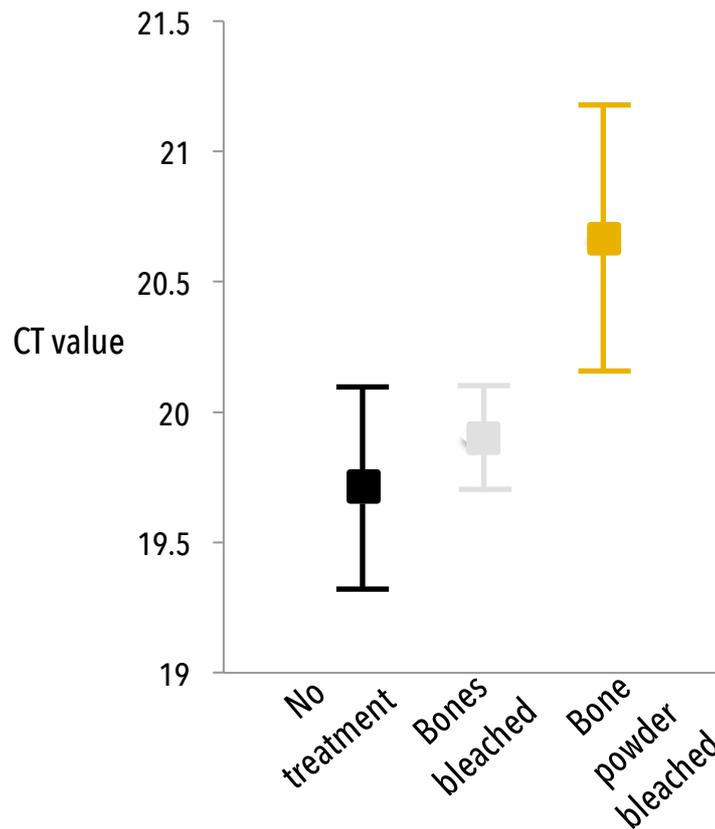
<b>Source of variance</b>	<b>SS</b>	<b>df</b>	<b>MS</b>	<b>F</b>	<b>p-value</b>
<b>Treatment</b>	1.52	2	0.76	6.233	0.034*
<b>Error</b>	0.732	6	0.122		
<b>Total</b>	2.252	8			

## S2.8.16 FUTURE DIRECTIONS

**SAMPLE PREPARATION.** Improvements in sample preparation protocols may decrease the variance in taxonomic assemblages identified using BBM. For example, bleach treating bones or bone powder prior to DNA extraction to eliminate exogenous contamination and to potentially improve endogenous DNA recovery; however, while it has been shown that bleach treatment can remove most (but not all) surface contaminants, it remains unknown whether it also destroys endogenous DNA, and it can also damage contaminating DNA such that it appears to be authentic aDNA (Barta et al. 2013). In a pilot study, we tested the effect of bleach treatment (soaking for 10 min in 10% bleach followed by three washes with 70% ethanol to remove the bleach and dry) of bones and bone powder on  $C_T$ -values in the initial qPCR screen and, at three dilutions of the DNA (neat, 1/10, and 1/50), found a statistically significant differences between treatments ( $p$ -value = 0.034,  $\alpha$  = 0.05; Table S2.8.6, Figure S2.8.4), but only for the neat dilution. Because bleach treatment took extra time to conduct without improving yield (and often decreasing yield), we opted not to bleach treat bones; however, bleach treatment may have more of an impact if shotgun libraries are to be prepared as opposed to amplicon libraries, and this should be tested. Furthermore, these samples were not carried through to sequencing, so in the future, tests could be conducted on the effect that bleach treatment has on the final taxonomic assemblages detected—the  $C_T$ -value may have increased by bleaching bone powder, but this may be because it was eliminating exogenous DNA that was being amplified in the other two treatments (e.g., human).

Using a planetary ball mill or other grinding apparatus to improve the consistency and fineness of bone powder, or slicing bone as opposed to powdering for DNA extraction may improve DNA yield (however, slicing may not be feasible for very small bones; Caputo et al. 2013).

The possible presence of covariates such as bone mass (e.g., does average bone mass influence the number of taxa detected?) or average body size may have an effect on which taxa can be detected and could be tested in the future. Finally, testing whether the number of bones pooled or the volume of bones pool affects the ability to detect taxa could help optimise the BBM method. In a preliminary experiment to

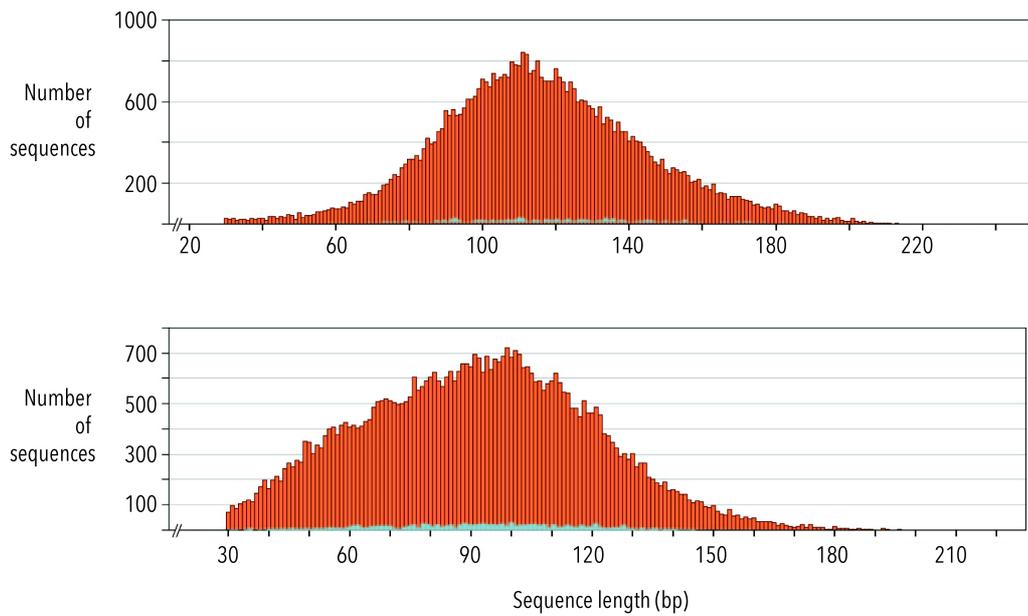


**FIGURE S2.8.4** | A COMPARISON OF AVERAGE  $C_T$ -VALUES FROM QPCRS OF NEAT (UNDILUTED) DNA EXTRACTED FROM BLEACH TREATED VERSUS NON-BLEACH TREATED BONES. Bulk bone subsamples that have undergone no treatment prior to grinding are shown in black, bleach treated bones (10% bleach for 10 min followed by three washes with 70% ethanol) prior to grinding are shown in grey, and bleach treated bones after grinding are shown in yellow. Vertical lines represent 95% confidence intervals.

investigate this (data available from the authors upon request), the family composition was compared across several different bone pooling and powdering strategies: (1) three pools of 50 bones, 20 mg sampled from each bone; (2) three pools of 50 bones, ground whole bones; and (3) three pools of 100 bones, 10 mg sampled from each bone. DNA was extracted from 100 mg of bone powder from each of these samples using the methods described in this paper. The family composition tended to vary in the same way no matter which strategy was used (results available upon request): that is, not every family was detected in every subsample, but across all subsamples of the treatment every family was detected. Pooling 100 bones as opposed to 50 did not result in the detection of more families, and pooling whole bones as opposed to 20 mg from each did not result in the detection, or loss, of families. This suggests that having more subsamples is more important than having larger pools but fewer subsamples. However, this is, again, likely to be dependent on the diversity of the sample and possibly the size of the bones.

**SHOTGUN SEQUENCING *and* OTHER PCR FREE METHODS.** Shotgun sequencing as opposed to amplicon sequencing can be used to potentially recover more taxa rapidly, as well as to evaluate levels of bacterial contamination relative to endogenous DNA, assess the DNA fragment-length distribution, recover nuclear loci in addition to mitochondrial loci, reduce amplification bias, and map aDNA error (such as the deamination of cytosine to uracil) (Taberlet et al. 2012). Shotgun sequencing has been used on environmental DNA samples (Smith et al. 2015), but is not without its drawbacks (Pedersen et al. 2014). One shotgun library was prepared as per Meyer and Kircher (2010) and sequenced on the Ion Torrent platform (method details available upon request) for one DNA extract from each site to briefly explore the potential of this method.

The degradation and contamination of DNA at both FF and BC was examined by the fragment length distribution reads sequenced from the shotgun libraries built on the DNA samples (Figure S.2.8.5). While the 16SMam primer set would have targeted a large section of the BC distribution, the 12SAH primer set would have targeted only a small section in the upper tail of the FF distribution; this suggests that the use of



**FIGURE S2.8.5** | FRAGMENT LENGTH DISTRIBUTION OF 50,000 SUBSAMPLED READS FROM SEQUENCED SHOTGUN LIBRARIES BUILT ON ADNA EXTRACTS. **A** FF and **B** BC (orange). Blue distributions within the orange distribution show the fragment lengths and relative proportion of endogenous reads (reads that aligned to **A** avian sequences (0.6%), and **B** marsupial sequences (3.55%) in GenBank).

smaller primer set may have resulted in the detection of more taxa. In addition, as can be seen from the fragment length distribution of the endogenous reads (blue) relative to the total number of reads (orange), the DNA contains a high proportion of contaminating sequences that are the same length as the aDNA. These results lead us to observe several major limitations of the shotgun method, at least for its use in identifying taxonomic assemblages from bulk bone. These include: (1) the cost of library preparation as well as time and labor required to prepare libraries; (2) greater depth of coverage is required because the majority of reads sequenced are of bacterial origin, resulting in a very small proportion of reads assigning to target or endogenous taxa; (3) an even more deficient reference genomic database for nuclear loci, of which many reads are expected to be (for example, many reads map to nuclear loci of South American marsupials such as *Monodelphis domestica*; the most likely explanation for this is that there is no closer match for those loci in the reference database so it maps to the next closest thing); (4) an inability to distinguish exogenous sequences from endogenous ones (should sequences mapping to, for example, *Rattus rattus* be considered contamination or not when they could potentially belong to a native rat that just has no representation in GenBank or could *Rattus rattus* really be present?); (5) prior knowledge of the taxa expected to be within the sample would be required in order to assemble the reads—this would result in significant ascertainment bias (for example, many reads can be mapped with equal probability to different genomes, or alternatively, reads may not map to any species identified by metabarcoding, giving the false impression that they are absent (for example, no shotgun reads from FF mapped to the *Dinornis* or *Ninox* mitochondrial genome, yet these were detected via metabarcoding); and (6) for single-stranded library build methods (cf., Gansauge and Meyer, 2013), there is a bias towards the capture of shorter fragments (Bennett et al. 2014)—while in most cases this is preferable since aDNA is expected to be highly degraded, longer fragments (> ca. 200 bp) of better preserved samples may not be captured, giving a false sense of the fragment length; and (7) the variation between multiple shotgun libraries from the same extract has not been investigated but is presumably quite high—therefore many shotgun library replicates may need to be performed in order to capture the full diversity of the sample, which would be very costly.

Other PCR-free methods such as ‘third generation’ single molecule real-time sequencing could be used to “reduce read number biases and allow the detection of taxa that do not amplify well” (Ji et al. 2013; Bohmann et al. 2014).

**ENRICHMENT.** Hybridisation capture probes can be used to enrich for target DNA—this may allow DNA from multiple taxa to be captured at once and reduce amplification bias (Taberlet et al. 2012; Bohmann et al. 2014). Another way to enrich for endogenous aDNA could be to perform a size selection step (for example, using an electronic gel system such as the *Pippin Prep*) on genomic extracts after the qPCR screen but prior to fusion-tag qPCR to remove high-molecular weight DNA that is not of interest and is likely to be bacterial in origin; however, this may reduce the total overall yield of DNA and how much this will improve the results remains to be tested. Highly specific primers (e.g., species-specific primers) may also be used to target species of interest if these are known *a priori*.

#### **S2.8.17 SUPPLEMENTARY REFERENCES**

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society of B-Biological Sciences* **279**: 4724-4733.

Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, et al. 2012. Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* **21**: 1966-1979.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**: 32-46.

Barta JL, Monroe C, Kemp BM. 2013. Further evaluation of the efficacy of contamination removal from bone surfaces. *Forensic Science International* **231**: 340-348.

Belperio A and Flint R. 1999. Geomorphology and geology. In: Robinson, A., and Armstrong, D. (eds.) *A Biological Survey of Kangaroo Island, South Australia*

1989 & 1990. *South Australia, Heritage and Biodiversity Section, Department for Environment, Heritage and Aboriginal Affairs.*

Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl E, et al. 2014. Library construction for ancient genomics: single strand or double strand? *BioTechniques* **56**: 289-300.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, et al. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology* / edited by Frederick M. Ausubel et al. 0 19, Unit 19.1021.

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Ecology and Evolution* **29**: 358-367.

Bollongino R, Tresset A, Vigne JD. 2008. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* **7**: 91-98.

Bronk Ramsey C. 2010. OxCal, v. 4.1.7

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

Caputo M, Irisarri M, Alechine E, Corach D. 2013. A DNA extraction method of small quantities of bone for high-quality genotyping. *Forensic Science International Genetics* **7**: 448-493.

Chao A, Chazdon RL, Colwell RK, Shen TJ. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* **8**: 148-159.

- Chao A, Chazdon RL, Colwell RK, Shen TJ. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62**: 361-371.
- Clark AG and Whittam TS. 1992. Sequencing errors and molecular evolutionary analysis. *Molecular Biology and Evolution* **9**: 744-752.
- Coissac E, Riaz T, Puillandre N. 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology* **21**: 1834-1847.
- Cooper A. 1994. Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthisittidae). *Experientia* **50**: 558-563.
- Davies M, Tyler M, Twidale C. 2002. Natural history of Kangaroo Island, Adelaide, *Royal Society of South Australia*.
- Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P. 2014. DNA metabarcoding and the cytochrome oxidase subunit 1 marker: not a perfect match. *Biology Letters* **10**.
- Earp MA, Rahmani M, Chew K, Brooks-Wilson A. 2011. Estimates of array and pool-construction variance for planning efficient DNA-pooling genome wide association studies. *BMC Medical Genomics* **4**.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
- Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, et al. 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research* **40**.

- Foote M and Miller AI. 2007. Principles of palaeontology, Third Edition ed. W. H. Freeman and Company, New York City, NY.
- Fortea J, de la Rasilla M, García-Tabernero A, Gigli E, Rosas A, et al. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidrón Cave (Asturias, Spain). *Journal of Human Evolution* **55**: 353-357.
- Fusco DA. 2014. The biodiversity of the Holocene terrestrial mammal faunas of the Murray Mallee, South Australia. Biological Sciences (Honours) Unpublished Honours Thesis, Flinders University.
- Gansauge M and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451-5.
- Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, et al. 2003. Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics* **72**: 48-61.
- Goecks J, Nekrutenko A, Taylor J, and The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**.
- Hajibabaei M, Singer GAC, Clare EL, Hebert PDN. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology* **5**.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011. Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLOS One* **6**.

- Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 9.
- Hogg AG, Quan H, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Hunter AA, Macgregor AB, Szabo TO, Wellington CA, Bellgard MI. 2012. Yabi: An online research environment for grid, high performance and cloud computing. *Source code for biology and medicine* **7**: 1.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, et al. 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters* **16**: 1245-1257.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Kindt R and Coe R. 2000. Tree diversity analysis, In: A manual and software for common statistical methods for ecological and biodiversity studies. *World Agroforestry Centre (ICRAF)*; Nairobi. ISBN 92-9059-179-X.
- Kuehl RO. 2000. Design of Experiments: Statistical Principles of Research Design and Analysis 2nd Edition. Brooks/Cole Publishing Company, Pacific Grove, CA.
- Lenth RV. Java Applets for Power and Sample Size (PiFace) 2006-2009. Retrieved 2013 from <http://www.stat.uiowa.edu/~rlenth/Power>.
- Little DP. 2011. DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability. *PLoS One* **6**.

- Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M. 2011. Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* **12**: 1741-2164.
- Macgregor S. 2007. Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *European Journal of Human Genetics* **15**: 501-504.
- Meyer M and Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbour Protocols* **6**.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. 2013. Scrapheap Challenge: A novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific reports* **3**.
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS One* **10**.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. 2015. Vegan: Community Ecology Package, R package version 2.2-1, <http://CRAN.R-project.org/package=vegan>.
- Pedersen MW, Ginolhac A, Orlando L, Olsen J, Andersen K, et al. 2013. A comparative study of ancient environmental DNA to pollen and microfossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quaternary Science Reviews* **75**: 161-168.
- Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, et al. 2014. Ancient and modern environmental DNA. *Philosophical Transactions Royal Society B* **370**.
- Piñol J, Mir G, Gomez-Polo P, Agustí N. 2014. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the

quantitative metabarcoding of arthropods. *Molecular Ecology Resources* **15**: 819-830.

Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, et al. 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences USA* **104**: 739-744.

Pruvost M, Schwarz R, Correia VB, Champlot S, Grange T, et al. 2008. DNA diagenesis and palaeogenetic analysis: Critical assessment and methodological progress. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **266**: 211-219.

Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, et al. 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science* **29**: 610-626.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria.

Robinson A and Armstrong DM. 1999. A Biological Survey of Kangaroo Island, South Australia in November 1989 and 1990. Department for Environment, Heritage and Aboriginal Affairs; South Australia.

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KAS, et al. 2012. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**: 178.

Shapiro B and Hofreiter M. 2012. Ancient DNA: methods and protocols. Humana Press (Springer Science), New York.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**: 1794-1805.

- Smith O, Momber G, Bates R, Garwood P, Fitch S, et al. 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**: 998-1001.
- Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* **5**: 1-34.
- Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* **21**: 2045-2050.
- Taylor PG. 1996. Reproducibility of ancient DNA sequences from extinct pleistocene fauna. *Molecular Biology and Evolution* **13**: 283-285.
- Thomsen PF and Willerslev E. 2015. Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**: 4-18.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Wooley JC, Godzik A, Friedberg I. 2010. A primer on metagenomics. *PLOS Computational Biology* **6**.

## 2.9 EPILOGUE

---

In this chapter, a new method of examining fossils that complements traditional morphological methods was further developed: we critically evaluated the utility of the novel bulk bone metabarcoding method and found that several steps in the method significantly contribute to variance in the results. Following the experiments described here, we might append the phrase “a man with multiple watches knows what time it is *not*” to the epigraph adage to illustrate that a good understanding of the uncertainty in a system leads us to a better estimate of the truth.

Through the use of aDNA and next-generation sequencing, we also found that families identified molecularly are a subset of those identified morphologically, and that the bulk bone metabarcoding method requires significantly less sampling than traditional methods. When coupled with careful experimental design, the bulk bone metabarcoding method has been shown to be a highly useful tool for identifying the countless morphologically indistinguishable bone fragments from palaeontological and archaeological excavations that are otherwise unidentifiable, and complements traditional morphological methods. We have highlighted the strengths and limitations of the method and emphasised ways in which the biases can be mitigated, which will be of great importance as the bulk bone metabarcoding method is being increasingly used among natural historians, including a broad range of scientists within the fields of ancient DNA, palaeontology, archaeology, molecular ecology, and conservation biology.

In addition to the experiments outlined in this chapter, several other pilot experiments were conducted after this study, in order to: (1) compare pooling strategies, including testing the effect of the number of bones and the amount of bone to subsample on the amount of recoverable DNA; (2) compare the effect of different bone digest buffers (e.g., TritonX-100 versus SDS as a detergent versus no detergent) and temperatures (e.g., digestion at 37°C versus 55°C) on the amount of recoverable DNA; and (3) compare the effect of bleach treatment of bones and bone powder on recoverable DNA. Although the sample size of the experiments were small (typically  $n=3$  or 4 per treatment), the results of these experiments helped

inform the experimental design employed future BBM experiments (such as those described in Chapters 3 and 4).

Nevertheless, there remains room for future improvement of the method, including the need to test its utility as a tool for assessing aDNA preservation at a site, as well as for describing past biodiversity where no morphological fossil record exists. Chapters 3 and 4 hereafter comprise an exploration of these avenues. However, it is the ideas examined in this chapter that formed the foundation upon which these subsequent studies were built.

— CHAPTER 3 —

---

**AN ASSESSMENT OF ANCIENT DNA PRESERVATION *in* HOLOCENE-  
PLEISTOCENE FOSSIL BONE EXCAVATED FROM THE WORLD  
HERITAGE NARACOORTE CAVES, SOUTH AUSTRALIA**

---

*There is still much to learn from the fossil record, as scientists increasingly look to the past to better understand how to conserve biodiversity today.*

- Ian Hunter

Sustainability, Environment  
and Conservation Minister,  
South Australia

### 3.1 PROLOGUE

---

Chapter 2 evaluated the strengths and limitations of the bulk bone metabarcoding method. The improvements to the method that were explored therein were a prerequisite for the study described in this chapter. As discussed in the last chapter, it remains necessary to examine how the method could be used to assess general aDNA preservation at a site. This is particularly important when the destruction of rare fossils is potentially at stake for little reward, as is often the case when attempting to retrieve aDNA from fossil deposits located in warm environments, such as Australia. The following chapter addresses this topic using bulk bone from a site within a hot and arid Australian climate that is suboptimal for the preservation of aDNA.

The Naracoorte Caves in South Australia are world renowned for their exceptional preservation of vertebrate fossils, but despite sustained palaeontological interest in the Naracoorte Caves, aDNA has not been integrated into any palaeontological study from this locality, thus far. The use of novel substrates, such as bulk bone, offers a new opportunity to retrieve aDNA from this site without the destruction of significant fossil specimens. Using the bulk bone metabarcoding method, we successfully retrieved and characterised ancient DNA in fossil bone from this World Heritage site for the first time, and provide new information regarding the past faunal biodiversity of Robertson Cave within the Naracoorte Caves World Heritage Area (NCWHA). A shotgun next-generation sequencing approach also allowed us to characterise the quality and quantity of ancient DNA preservation over the past ca. 20,000 years. This information was then used to discuss the site's potential for addressing more sophisticated palaeoecological questions regarding past extinction processes and genetic responses to climate change, which has implications for the conservation and management of today's biodiversity.

The study presented in this chapter resulted in a manuscript published in a 2016 issue of *Journal of Quaternary Science* (Grealy et al., Vol. 31, Pg. 33-45), a facsimile of which can be found in Appendix I. This chapter is a reproduction of the aforementioned manuscript (formatting, including in-text referencing, cross referencing and headings, excepted).

### **3.1.1 ACKNOWLEDGEMENTS**

The authors acknowledge funding for this project via Friends of Naracoorte Caves (ER), Sir Mark Mitchell Research Foundation (AM), and Caring for our Country Grant 0C11-00487. MB was supported in this research by an Australian Research Council (ARC) future fellowship FT0991741 and DP120104435. Computing resources at the University of Copenhagen as well as the Pawsey Centre, WA, were used to perform BLAST searches. We thank S Bourne for assistance to ER with the 2011 excavations of RCEC. The authors thank D Haouchar for permitting the use of the 12SMarsMini primers.

### **3.1.2 AUTHOR CONTRIBUTIONS**

ER excavated and collected materials from RCEC, provided dates, stratigraphy and faunal data from morphological analyses. AM provided fauna data from morphological analyses and palaeoecological input. AG designed, performed, and analysed experiments with input from MA, MB and NR. AG wrote the manuscript with contributions and edits from all co-authors.

### **3.1.3 AUTHOR DECLARATIONS**

All necessary permits were obtained for the described study, which complied with all relevant regulations.

The authors declare no competing interests.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work (e.g., journal articles) in which the copyright is held by another party (e.g. publisher, co-author). Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged. Reprinted from Publication *Journal of Quaternary Science*, Vol. 31, Grealy AC et al., An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone

excavated from the world heritage Naracoorte Caves, South Australia, Pg. 33-45, Copyright (2016), with permission from Elsevier.

All supplementary data related to this article can be found at in the file “Supplementary Information” published with the online version of the article, as well as section 3.8.

Sequencing data is deposited on the online data repository Data Dryad, and is available at: <http://dx.doi.org/10.5061/dryad.vf345>.

**AN ASSESSMENT OF ANCIENT DNA PRESERVATION IN HOLOCENE-PLEISTOCENE  
FOSSIL BONE EXCAVATED FROM THE WORLD HERITAGE NARACOORTE CAVES,  
SOUTH AUSTRALIA**

—*in*—

***JOURNAL OF QUATERNARY SCIENCE (2016) | VOL. 31 | PG. 33-45***

Alicia Grealy<sup>1</sup>, Amy Macken<sup>2,3</sup>, Morten E. Allentoft<sup>4</sup>, Nicolas J. Rawlence<sup>5</sup>,  
Elizabeth Reed<sup>2,6</sup>, and Michael Bunce<sup>1\*</sup>

<sup>1</sup> Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment  
and Agriculture, Curtin University, Perth, WA 6102, Australia.

<sup>2</sup> School of Biological Sciences, Flinders University, Bedford Park, SA 5042,  
Australia.

<sup>3</sup> Department of Environment, Water and Natural Resources, Naracoorte Caves, SA  
5271, Australia.

<sup>4</sup> Centre for GeoGenetics, Natural History Museum, University of Copenhagen,  
1350, Denmark.

<sup>5</sup> Allan Wilson Centre, Department of Zoology, University of Otago, Dunedin 9016,  
New Zealand.

<sup>6</sup> School of Physical Sciences, The University of Adelaide, Adelaide, SA 5000,  
Australia.

\* Corresponding author. Trace and Environmental DNA Laboratory, Department of  
Environment and Agriculture, Curtin University, Perth, WA 6102, Australia. *E-mail*  
*address*: michael.bunce@curtin.edu.au

**Key words:** ancient DNA, bulk bone metabarcoding, Naracoorte Caves,  
palaeontology, DNA preservation

### 3.2 ABSTRACT

---

Although there is a long history of research into the fossil deposits of the Naracoorte Caves (South Australia), ancient DNA (aDNA) has not been integrated into any palaeontological study from this World Heritage site. Here, we provide the first evidence of aDNA preservation in Holocene- and Pleistocene-aged fossil bone from a deposit inside Robertson Cave. Using a combination of metabarcoding and shotgun next-generation sequencing approaches, we demonstrate that aDNA from diverse taxa can be retrieved from bulk bone as old as 18,600 cal a BP. However, the DNA is highly degraded and contains a lower relative proportion of endogenous sequences in bone older than 8,400 cal a BP. Furthermore, modeling of DNA degradation suggests that the decay rate is rapid, and predicts a very low probability of obtaining informative aDNA sequences from extinct megafaunal bones from Naracoorte (ca. 50,000 cal a BP). We also provide new information regarding the past faunal biodiversity of Robertson Cave, including families that have not been formerly described in the fossil record from here before. Collectively, these data demonstrate the potential for future aDNA studies to be conducted on material from Naracoorte, which will aid in the understanding of faunal turnover in southern Australia.

### 3.3 INTRODUCTION

---

**V**ERTEBRATE fossil assemblages are an important source of information about the timing and extent of past biodiversity change. The Naracoorte Caves in south-eastern South Australia are world renowned for their preservation of vertebrate fossils, which includes mammals, birds, reptiles and frogs. Some deposits across the site date from approximately 1000 to 500,000 years in age, while others still actively trap animals today (Prideaux et al. 2007; Macken and Reed, 2013). The natural and scientific value of the caves is reflected in their World Heritage listing, as well as in on-going palaeontological research that has contributed to our knowledge about taxonomy, the past biodiversity of southern Australia, and the effects of climate and associated habitat changes on the distribution of both extinct and living species (e.g., Prideaux et al. 2007; Macken et al. 2012; Macken and Reed, 2013; Macken and Reed, 2014). Research into the fossil deposits within the Naracoorte Caves World Heritage Area (NCWHA) has involved examination of the geochemical and physical properties of fossil bones themselves, as well as additional materials that preserve information about past ecosystems (i.e., faunal diet, local climatic conditions, etc.), such as sediments and cave formations (e.g., Ayliffe et al. 1998; Forbes and Bestland, 2007; Macken, 2009). Newly discovered organic plant macro- and microfossils preserved in the Naracoorte caves (e.g., seeds, leaves, pollen and phytoliths) also provide direct evidence about past habitats and how these have changed over time in the south-east region of South Australia (Darrénougué et al. 2009; Reed, 2012), and also suggest the possibility of biomolecule preservation at Naracoorte.

Ancient DNA (aDNA) isolated from fossil assemblages represents another powerful source of information about past biodiversity. Ancient DNA refers to degraded, fragmented, and chemically modified DNA that exists in trace amounts in material such as bones (Kuhn et al. 2010), leaves (Jaenicke-Despres et al. 2003), feathers (Rawlence et al. 2009), eggshell (Oskam et al. 2010), coprolites (Wood et al. 2012), and sediment (Haile, 2012; Rawlence et al. 2014; Willerslev et al. 2003). Ancient DNA can be used, for example, to estimate the effective population size and past genetic diversity of animal communities that is not available from the morphological

study of fossil bones (Hadly et al. 2004; O'Keefe et al. 2009). Studies using aDNA have also addressed a wide range of research questions regarding taxonomy (e.g., Bunce et al. 2003), phylogeny (e.g., Mitchell et al. 2014), palaeodiet (e.g., Hofreiter et al. 2003), palaeoclimate (e.g., Jorgensen et al. 2012), climate change (e.g., Hadly et al. 2004; Magyari et al. 2011), population dynamics (e.g., Leonard et al. 2002; Bunce et al. 2009; Allentoft et al. 2014), and interspecies relationships (e.g., Wood et al. 2013a; Wood et al. 2013b). Furthermore, the use of aDNA has had implications for the conservation and management of endangered animal populations (Shepherd and Lambert, 2008). With advances in DNA sequencing techniques, aDNA is being extracted from increasingly older fossil material in increasingly poor preservation conditions (Orlando et al. 2011; Murray et al. 2012), including the relatively warm, moist microclimates of many Australian cave systems (Murray et al. 2013; Haouchar et al. 2014; Llamas et al. 2014; Grealy et al. 2015).

Despite sustained palaeontological interest in the Naracoorte Caves (e.g., Wells et al. 1984; Prideaux et al. 2007; Macken et al. 2012), aDNA has not been integrated into any palaeontological study from this locality because past attempts to extract aDNA from fossil materials were unsuccessful. To address this deficit, we provide the first systematic study to explore the preservation of aDNA in fossil bones from the Naracoorte Caves and its potential as a tool for future palaeontological research. Our study focused on a fossil record spanning the last ca. 18,500 years from the Robertson Cave Entrance Chamber (RCEC), located within the NCWHA. RCEC preserves the youngest dated *in situ* fossil remains for the Naracoorte Caves, which offers the best opportunity to retrieve aDNA and assess its preservation. It contains a diverse small mammal fauna (body mass < 2.5 kg) of 37 species identified from ca. 2000 specimens (Macken and Reed, 2013); however, large mammal, reptile, frog, bird and plant fossils have also been recovered from the site. Following examples from other vertebrate fossil localities within Australia (Murray et al. 2013; Haouchar et al. 2014, Grealy et al. 2015), we applied a bulk bone metabarcoding (BBM) method to test for the presence aDNA in the vertebrate fossil record from six layers of the RCEC deposit. The BBM method has been demonstrated to be an accurate, rapid, cost-effective way of assessing aDNA preservation and investigating past biodiversity that does not require the destruction of whole fossil specimens (Murray et al. 2013). This method makes use of non-diagnostic bone fragments that are

numerous in fossil deposits like RCEC. By extracting, amplifying, next-generation sequencing (NGS), and comparing mitochondrial DNA from many morphologically unidentifiable bone fragments in parallel, it is possible to identify the taxa present in the assemblage (Murray et al. 2013). We also aimed to assess the quality (degradation and contamination) and limits of aDNA preservation at RCEC using a ‘shotgun’ sequencing approach, where total genomic DNA from three layers (youngest, middle-aged, and oldest) was sequenced and used to model DNA decay (Allentoft et al. 2012). This information allowed us to assess the site’s potential for addressing more sophisticated palaeoecological questions relating to population ecology, genetic diversity, extinction processes, and genetic responses to climate change.

### **3.4 MATERIALS *and* METHODS**

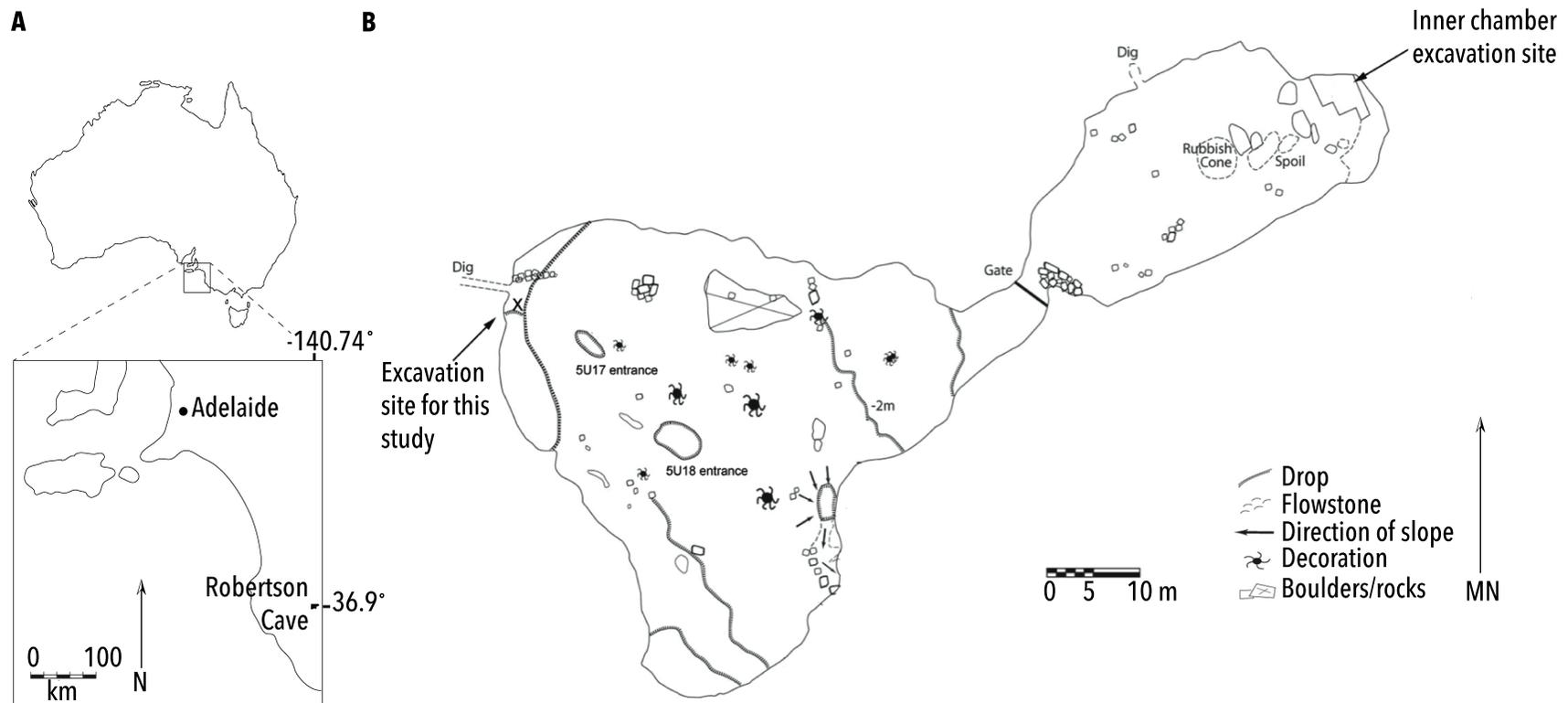
---

#### **3.4.1 STUDY SITE *and* DATING**

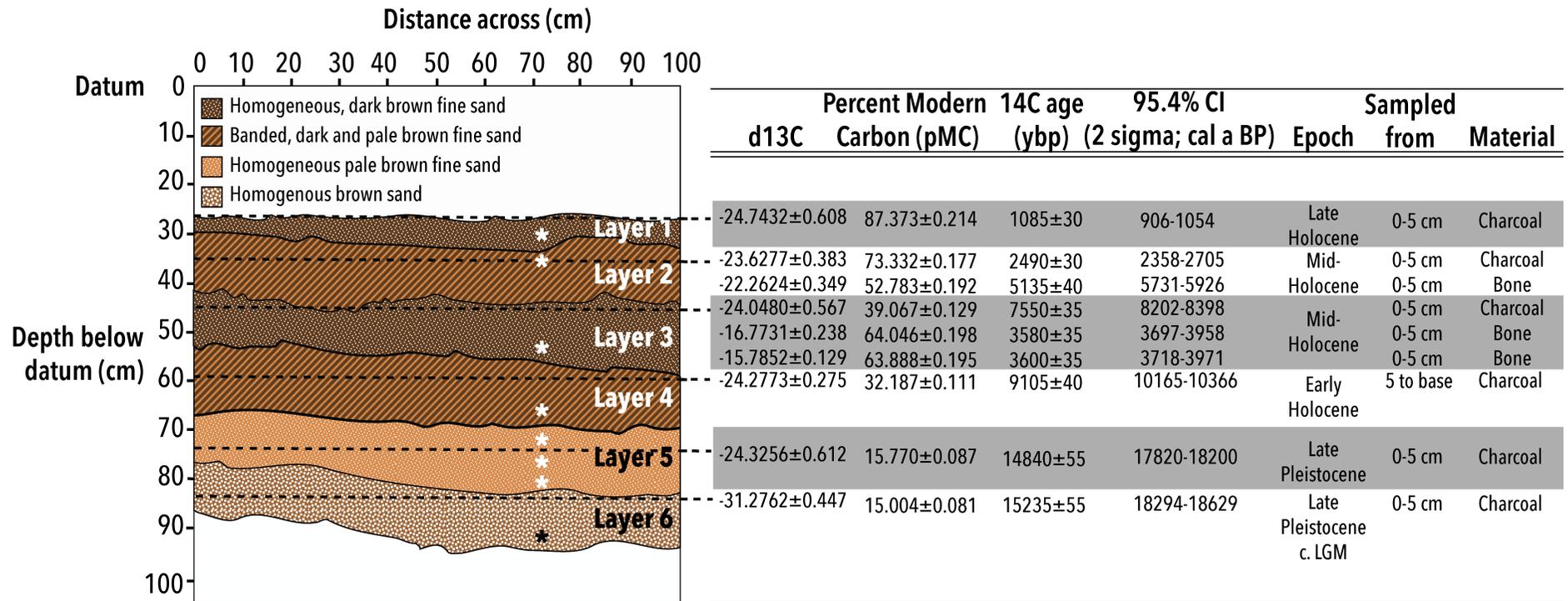
Robertson Cave (37° 5.789’ S, 140° 50.125’ E) is located south-east of Naracoorte, in South Australia (refer to Macken and Reed (2013) for details; Figure 3.1). The site was excavated to a depth of 83 cm below datum, covering six distinct sedimentary strata (Figure 3.2) that were radiocarbon dated to between 906 and 18,629 cal a BP (Figure 3.2, S3.8.1, Table S3.8.1). Charcoal for Accelerator Mass Spectrometry (AMS) <sup>14</sup>C dating was selected from each of the sedimentary layers, with multiple samples (spits) collected from some layers. AMS <sup>14</sup>C dating was completed by the ANU Radiocarbon Dating Centre (Canberra, Australia) using the techniques described in Fallon et al. (2010). Dates were calibrated using the ShCal13 calibration curve to 2 St Dev. (95.4%; Hogg, 2013) using OxCal (Ramsey, 2009).

#### **3.4.2 SAMPLE COLLECTION *and* BULK BONE SAMPLING**

Vertebrate fossil material was sorted from wet-screened samples excavated from RCEC. All fossil material for analysis was collected from excavations conducted in 2011 (see section 3.4.1). The samples were excavated following standard palaeoecological methods with sediments and associated fossil material collected



**FIGURE 3.1** | MAPS OF THE NARACOORTE CAVES WORLD HERITAGE AREA IN SOUTHEAST SOUTH AUSTRALIA AND ROBERTSON CAVE ENTRANCE CHAMBER. **A** Location of the Naracoorte Caves World Heritage Area, indicating the relative location of Robertson Cave (adapted from Macken and Reed, 2013), **B** map of Robertson Cave Entrance Chamber (5U17), Naracoorte, SA (surveyed by L. Reed and M. Koukлина 2004, drawn by D. Grindley, adapted from CEGSA Map No. 4014, September 2007).



**FIGURE 3.2 |** THE STRATIGRAPHY AND DATING FOR LAYERS 1-6 OF THE EXCAVATION. Dotted lines indicate where samples were taken for dating analysis. Asterisks indicate approximately where bulk bone samples were taken (5 cm spits).

from individual sedimentary layers. Respirators were worn during the excavation and facemasks were worn during sorting and wet sieving. Other DNA contamination reduction methodologies (Allentoft et al. 2013) were not employed during the excavation of this material. An assortment of 105 bone fragments were randomly collected from screened material from each of Layers 1–6 and placed into labelled vials.

### **3.4.3 SAMPLE PREPARATION**

In a designated ultra-clean aDNA facility (TRACE) at Curtin University, WA, Australia, bulk bones from each layer were divided into three subsamples of 35 bones each (average mass 28 mg, range 5–250 mg), following standard aDNA practice (Figure S3.8.1, S3.8.2; S3.8.3). For each subsample, approximately 20 mg of each bone was placed together into a clean stainless steel grinding pot and ground into a fine powder using a planetary ball mill (*Retsch* PM200; 200 rpm for three min), which was then transferred to a clean 15 ml tube.

### **3.4.4 DNA EXTRACTION *and* TOTAL DNA QUANTIFICATION**

DNA from two 100 mg aliquots of bone powder from each subsample was extracted (i.e., 36 extracts in total) using the method described in Dabney et al. (2013) with minor changes (S3.8.4). Two blank (DNA free) extraction controls were included per 10 extractions. This DNA was then used for both amplicon and shotgun sequencing (Figure S3.8.1). All extractions and downstream qPCR reactions were prepared in a physically isolated, pre-PCR ultra-clean environment following standard aDNA practice (Willerslev and Cooper 2005; Knapp et al. 2012; Shapiro and Hofreiter 2012). One additional extraction per layer was performed and used for subsequent quantification via spectrophotometry, gel electrophoresis, and relative qPCR (Figure S3.8.1). The total DNA concentration was quantified using a Nanodrop spectrophotometer (*Thermo Scientific*), following the manufacturer's instructions (S3.8.5). 5 µL of genomic DNA from each layer was run on a 3% agarose gel electrophoresis (S3.8.6) in order to visualise the relative concentrations and total fragment lengths of the genomic DNA, cognisant of the fact that much of the visualised DNA was likely to be of microbial origin. Quantitative PCR was

performed to determine the efficiency of amplification and assess the relative quantities of template DNA extracted from each layer (S3.8.7).

### 3.4.5 AMPLICON SEQUENCING

Following the suggestions of Murray et al. (2013, 2015), the 36 extracts (above) were diluted and screened for inhibition using qPCR (*Applied Biosystems*; S3.8.7). Dilutions exhibiting the least inhibition were then amplified via qPCR (in triplicate) using uniquely indexed *Illumina* fusion primers specific for a small, diagnostic barcoding region of the mitochondrial *12S rRNA* gene in mammals and birds (12SAO, a 150 bp amplicon or 103 bp without primers; Cooper, 1994; Table S3.8.2), as well as an additional region specific for marsupials (12SMarsMini, an 85 bp amplicon or 41 bp without primers; Haouchar et al. (*unpublished*); Table S3.8.2). The purpose of targeting two different sized amplicons from different barcoding regions is two-fold: it allows us to gauge the fragment length of amplifiable target DNA, as well as to independently verify taxonomic IDs, potentially resolve taxonomic ambiguities, or identify more taxa than could be identified by using only one region (Greal et al. 2015). Extraction controls and three template-free PCR controls were included, and unique indexes were used to eliminate the possibility of sequence contamination arising from previously amplified DNA. Details of the qPCR reaction and thermocycling conditions used are described in S3.8.7. Amplicons were purified using an *Agencourt* AMPure XP PCR purification kit with minor changes (*Beckman-Coulter*; S3.8.8), run on a 2% agarose gel electrophoresis (S3.8.9), and pooled in approximately equimolar concentrations to create sequencing libraries (Murray et al. 2013). The absolute concentration of the sequencing library was quantified via qPCR (Murray et al. 2012; S3.8.10, Table S3.8.3) to determine how much to add to the sequencing reaction. Unidirectional sequencing was performed on *Illumina's* MiSeq platform by following the manufacturer's instructions for the MiSeq 300 v.2 Nano kit using a custom sequencing primer (S3.8.11).

### 3.4.6 SHOTGUN SEQUENCING

Shotgun sequencing allows us to more accurately estimate the proportion of the total DNA that is endogenous vs. exogenous (non-target), as well as visualise the fragment length distribution of endogenous aDNA (Allentoft et al. 2012; Heintzman et al. 2014), estimate the DNA decay rate (Allentoft et al. 2012), and examine damage (cytosine deamination) patterns (Briggs et al. 2007). Shotgun libraries were prepared on extracts from Layers 1, 3 and 6, a DNA free control, and a positive oligonucleotide control by following the protocol outlined in Gansauge and Meyer (2013) with minor modifications (S3.8.12, Table S3.8.4). After amplifying libraries with uniquely indexed fusion primers (S3.8.13), each library was purified using an *Agencourt* AMPure XP PCR purification kit (*Beckman-Coulter*) following the manufacturer's instructions with minor modifications (S3.8.8). The absolute concentration of each shotgun library was quantified via qPCR using a synthetic standard of known molarity as per Murray et al. (2012; S3.8.10, Table S3.8.3) and the libraries were then pooled in approximately equimolar concentrations to create the final sequencing library. The absolute concentration of the sequencing library was quantified via qPCR as above to determine how much to add to the sequencing reaction (S3.8.10). Unidirectional sequencing was performed on *Illumina's* MiSeq platform by following the manufacturer's instructions for the MiSeq 150 v.3 kit, with minor modifications (S3.8.14).

### 3.4.7 BIOINFORMATICS *and* DATA ANALYSIS

After sequencing, raw FastQ files were downloaded and imported into Geneious v.7.1.4 (<http://www.geneious.com>; Kears e et al. 2012) for index separation and trimming (Murray et al. 2013). Only reads flanked either side by sequences matching indexes and primers with 100% identity were accepted. Quality control was performed using the web-based workflow Galaxy ([usegalaxy.org](http://usegalaxy.org); Giardine et al. 2005; Blankenberg et al. 2010; Goecks et al. 2010), and chimera filtering and abundance filtering were performed using USEARCH v.6.1 (Edgar 2010; Edgar et al. 2011) in QIIME v.1.8.0 (Caporaso et al. 2010), as per Grealy et al. (2015): reads were required to have an average quality score above 25 (i.e.,  $\geq Q25$ ), 100% of bases over Q10, 98% of bases over Q15, and 90% of bases of Q20 to be accepted; chimeric

reads and low-abundant reads (comprising less than 0.1% of the total number of reads) were discarded. Abundance filtering was not performed on shotgun data, as PCR amplification was stopped during the linear phase resulting in few clonal copies. For both amplicon and shotgun datasets, sequences present in the extraction or PCR controls (typically microbial or human in origin) were mapped to each dataset and subtracted from them to obtain the final dataset. Taxonomic identification of amplicons was achieved by comparing unique sequences to NCBI's GenBank (Benson et al. 2006) nucleotide reference database via BLASTn (-F No, -e 0.01, -m Pairwise, -G 5, -E 2, -v 20, -b 20, -W 7, -reward 1) (Altschul et al. 1990), with searches executed in YABI (ccg.murdoch.edu.au/yabi; Hunter et al. 2012). BLAST results were imported into MEGAN v.4.70.4 (ab.inf.uni-tuebingen.de/data/software/megan4; Huson et al. 2007) to evaluate taxonomic assignments. The LCA parameters used in MEGAN were: min support of 1, min score of 35, top per cent of 10, min complexity 0.44, win score 0. Identifications were based on the percent sequence similarity of the query to the reference across 100% of the query, in the BLAST hits of the terminal nodes of the MEGAN tree. Current and historic taxon distributions described refer to records housed in the online database Atlas of Living Australia ([www.ala.org.au](http://www.ala.org.au), accessed 30 October 2015).

### **3.4.8 MODELING of aDNA PRESERVATION and DECAY OVER TIME**

Each shotgun dataset (not dereplicated) was rarefied in MEGAN v.4.70.4 (Huson et al. 2007) to account for differences in sequencing depth between the different libraries: 60,000 sequences (approximately half the smallest dataset) were sufficient to capture almost all the diversity, and as such, were randomly subsampled from each layer. Fragment-length distributions of endogenous aDNA from the rarefied dataset (defined as the unique marsupial shotgun sequences obtained from mapping sequences against GenBank's nr nucleotide database using BLASTn as above) were constructed for Layers 1, 3, and 6. The summary statistics for each distribution (number of sequences, mean fragment length, standard deviation, mode fragment length, and maximum length retrieved) were calculated in Geneious v.7.1.4 (Kearse et al. 2012). Under an exponential decay model, fragment length should be inversely

proportional to the log of the copy number, where the probability of a bond in the DNA backbone being broken,  $\lambda$ , is equal to the exponential co-efficient (Allentoft et al. 2012). An exponential relationship was modelled using the declining part of the distribution, excluding biases in the distribution tails (Allentoft et al. 2012), and the exponential co-efficient ( $\lambda$ ) and fit ( $R^2$ ) of the relationship was calculated in *Microsoft's* Excel. The per nucleotide fragmentation rate per year ( $k$ , per site per year) was calculated for the average, minimum, and maximum possible calibrated ages of layers with > 500 marsupial sequences (i.e., Layers 1 and 3), as  $\lambda$  divided by age (cal a BP; Allentoft et al. 2012).

The estimations of  $k$  for Layers 1 and 3 were used to predict a number of relevant measures (Allentoft et al. 2012): the average fragment length of the DNA in the extract (bp) ( $1/\lambda$ ); the number of years until the DNA is completely degraded (i.e., average fragment length = 1 bp) ( $1/k$ ); the decay constant ( $k_{30} = 1 - e^{-k*30}$ ) and the molecular half-life (years) of the smallest informative fragment size of 30 bp ( $\ln(2)/k_{30}$ ), corresponding to the number of years it would take 50% of 30 bp fragments to be gone (i.e., an average of 0.5 strand breaks per 30 bp); and the expected proportion of broken bonds after 50,000 years ( $\lambda_{50,000} = k*50,000$ ). Next, the expected proportion of 30 bp fragments left after time  $t$  was modeled by plotting  $e^{-k_{30}*t}$ , and the proportion of 30 bp fragments left after 50,000 years was calculated using  $e^{-k_{30}*50,000}$ . Finally, the probability of a fragment of size  $L$  bp surviving after 50,000 years was modeled by plotting  $e^{-\lambda_{50,000}*L}$ , and the probability of a 30 bp fragment surviving 50,000 years was calculated using  $e^{-\lambda_{50,000}*30}$ .

### 3.4.9 BASE COMPOSITION ANALYSIS

The GC content of the unique marsupial shotgun sequences for Layers 1, 3, and 6 was calculated in Geneious v.7.1.4 (Kearse et al. 2012) as the proportion of all the bases in the dataset that were either a G or C nucleotide. The T-content of the first and last base of these same sequences were also calculated as the proportion of first-base nucleotides in the dataset that were a T, or the last-base nucleotides in the dataset that were a T.

## 3.5 RESULTS *and* DISCUSSION

---

### 3.5.1 DNA PRESERVATION *at* RCEC

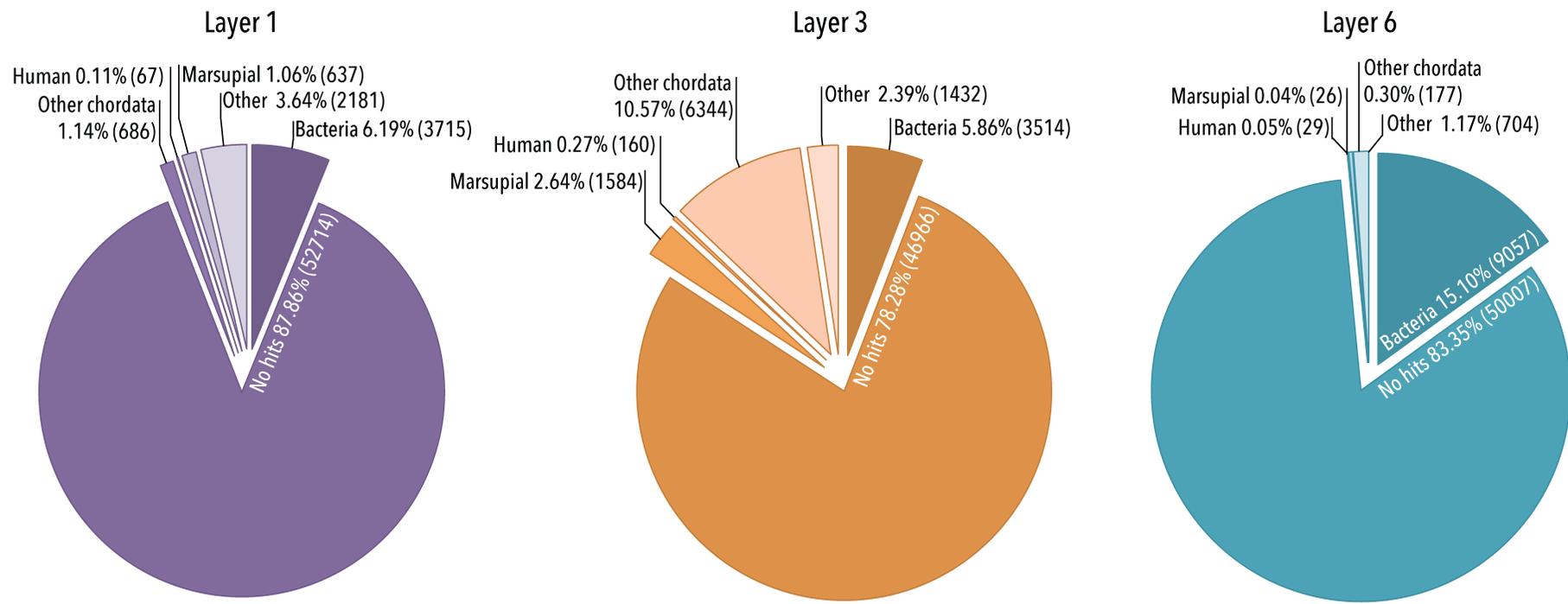
Fossils preserved in caves tend to exhibit less degraded aDNA in comparison to fossils from other non-frozen environments, as caves typically offer a cool, dry environment and can buffer temperature (Bollongino et al. 2008; Elsner et al. 2014; Gutiérrez-García et al. 2014; Olade et al. 2014). However, aDNA retrieval from caves within the Naracoorte Caves World Heritage Area (NCWHA) has never been published, despite the fact that they house a rich palaeontological record, displaying exceptional preservation of fossils dating as far back as 500,000 years. Here, NGS metabarcoding of two short mitochondrial gene regions amplified from DNA extracted from bulk bones present across six stratigraphic layers spanning approximately 18,000 cal a BP yielded a total of 411,717 sequences (Table S3.8.5, S3.8.15). NGS of shotgun libraries from Layers 1, 3 and 6 yielded 648,739, 117,428, and 240,293 sequences, respectively; subsampling 60,000 sequences from each layer yielded between 16 and 746 unique marsupial sequences per layer (Figure 3.4a). Examination of the base composition of these reads (S3.8.16, Table S3.8.6, Figure S3.8.2), shows that there is a bias towards thymine bases at both the 5' and 3' ends (> 35-75%, Table 3.2), which is to be expected in aDNA (Gansauge and Meyer 2013) as DNA hydrolysis and oxidation can result in the deamination of cytosine to uracil (sequenced as a T; Hoss et al. 1996; Hofreiter et al. 2001; Gilbert et al. 2003). Although we cannot be certain that this bias is caused by cytosine deamination, the composition of thymine in these positions is higher than would be expected from undamaged DNA (roughly 25%). The bias towards thymine bases at the ends of fragments is also higher in Layer 6 than Layers 1 and 3 (Table 3.2). This type of damage pattern is highly suggestive that the DNA obtained is genuinely ancient in nature.

While we were able to recover aDNA from all six layers, the aDNA retrieved from layers older than ca. 8,000 years (early Holocene) was considerably scarcer, more degraded, and more contaminated than that of the younger layers (note that the overlapping of ages for Layers 2 and 3 suggests that post depositional disturbance may have occurred, disrupting the temporal integrity of these layers; however, for the

**TABLE 3.1** | REPLICATE AND AVERAGE QPCR CYCLE-THRESHOLD ( $C_T$ ) FOR BOTH 12SAO AND 12SMARS MINI PRIMER SETS FOR LAYERS 1-6. The fold difference in the quantity of amplifiable DNA is calculated relative to Layer 6.

	12SAO						12SMarsMini					
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
<b><math>C_T</math> value replicate 1</b>	27.48	26.55	26.77	32.73	29.86	30.87	38.21	29.69	30.15	45.35	NA	NA
<b><math>C_T</math> value replicate 2</b>	27.56	26.01	27.58	30.33	30.83	33.44	33.72	31.56	29.77	NA	NA	44.11
<b>Average <math>C_T</math> value</b>	27.52	26.28	27.18	31.53	30.35	32.16	35.97	30.63	29.96	45.35	NA	44.11
<b>Fold difference in the quantity of amplifiable DNA (relative to Layer 6)</b>	24.85	58.69	31.56	1.54	3.51	1.00	283.07	11465.41	18179.19	0.42	NA	1.00

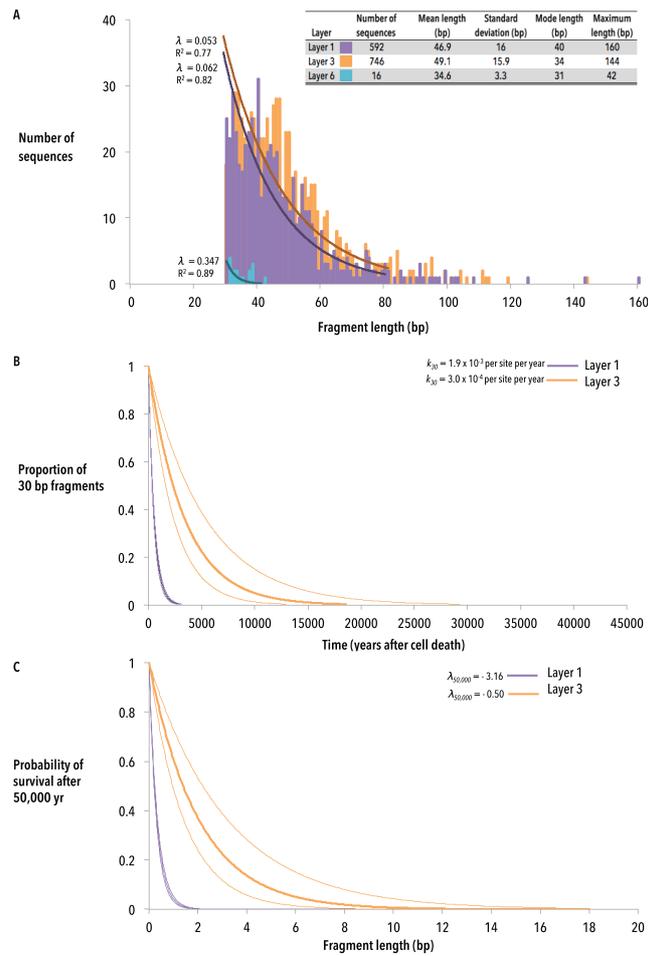
purposes of assessing aDNA preservation, and in the context of the whole site, Layers 2 and 3 provide a mid-Holocene sample for comparison against the late Holocene record of Layer 1, the early Holocene record of Layer 4, and the late Pleistocene record of Layers 5 and 6). Firstly, although much of the genomic DNA is likely to be of bacterial origin, the total amount of genomic DNA recovered from Layers 1–3 is similarly higher (on average 30.75 ng/μL) and more pure (average 260/280 = 1.66, where the 260/280 ratio of pure DNA = 1.8) than that of Layers 4–6 (25.57 ng/μL, 260/280 = 1.56) as measured by spectrophotometry (Table S3.8.5). Visualisation of genomic DNA by agarose gel electrophoresis (Table S3.8.7) shows brighter smears of similar intensity for Layers 1–3 compared with Layers 4–6, also indicating that while the DNA from all layers is highly degraded, DNA from Layers 1–3 is more concentrated than Layers 4–6. These results are also supported by the finding that the relative amount of amplifiable DNA (expressed by the ‘fold difference’) for both primer sets is higher extracts from Layers 1-3 compared with 4-6 as determined by qPCR (Table 3.1, Table S3.8.8); for 12SAO, Layers 1–3 yielded similar (the same order-of magnitude) amounts of template DNA, about 24–60 times as much as Layers 4–6, and similarly, Layers 1–3 contain over 100-fold more 12SMarsMini template DNA than Layers 4–6, which themselves contain very low quantities (i.e., higher  $C_T$  values) of template DNA (Table 3.1). Moreover, the poor efficiency (< 95%; Table S3.8.9, Figure S3.8.3, Figure S3.8.4) and the stochasticity (i.e., the lack of consistency between the  $C_T$  values of duplicate PCR reactions—Table S3.8.8—and sporadic amplification with both primer sets in extraction replicates—Table S3.8.10) of successful target DNA amplification in Layers 4–6, suggests that they are more inhibited and contain considerably less template DNA molecules than Layers 1–3 (Table 3.1). Finally, Layers 1 and 3 contain a lower proportion of bacterial DNA (5.86% and 6.19%, respectively) and a higher proportion of endogenous (marsupial) DNA (1.06% and 12.64%, respectively) compared with Layer 6 (15.10% bacteria, 0.04% marsupial) as determined by shotgun sequencing (Figure 3.3). That is, Layer 6 contains about 27 times less endogenous DNA than Layer 1 and about 316 times less than Layer 3. Thus, these findings suggest that the quantity of endogenous aDNA is greater in Layers 1-3 compared with Layers 4-6.



**FIGURE 3.3** | PIE CHARTS SHOWING THE PERCENTAGE OF ENDOGENOUS AND EXOGENOUS SHOTGUN SEQUENCE READS. Pie charts for Layer 1 (purple), 3 (orange), and 6 (blue) showing the percentage of shotgun sequence reads out of 60,000 subsampled sequences that map to bacterial, human, marsupial, other chordate, and other organisms, and nothing by BLASTn to GenBank’s nr nucleotide database.

In addition, the shorter mean and maximum fragment length of aDNA in Layers 4–6 compared with Layers 1–3, gauged by both amplicon and shotgun sequencing, indicate that endogenous DNA from Layers 4–6 is more degraded than Layers 1–3. Although some endogenous mtDNA was amplified with both primer sets in extracts from Layers 4–6, most taxa were only amplified with the short 12SMarsMini primer set (Table 3.3). While this suggests that there is some endogenous mtDNA that is at least 150 bp in length present in these layers, the majority of it is expected to be much shorter than this and would be in trace amounts that approach the limits of detection. In contrast, many of the same taxa were detected in Layers 1–3 with both primer sets, and therefore these extracts have a greater proportion of mtDNA that is at least 150 bp in length compared with Layers 4–6. These results are consistent with those of the shotgun sequencing of DNA from Layers 1, 3, and 6 (Figure 3.4a), which was used to determine the distribution of fragment lengths of informative (> 30 bp) endogenous (marsupial) DNA: the mean fragment length of sequences from Layers 1 and 3 was between 45–50 bp, with the maximum fragment retrievable being above 140 bp in length, whereas for Layer 6, the mean fragment length was around 35 bp, with the maximum fragment retrievable being 42 bp (Figure 3.4a).

Although these results show that DNA preservation largely appears to be decreasing through time (as would be expected by a rate model of DNA fragmentation), we find that the molecules recovered from Layer 3 appear to be less degraded than Layer 1, despite being approximately five times older. Using the observed fragment length distributions of marsupial shotgun sequences from Layers 1 and 3 (Figure 3.4a) to model the rate of DNA degradation by depurination (the generation of abasic sites, resulting in strand breaks) over time (Allentoft et al. 2012), the per nucleotide fragmentation rate,  $k$ , was found to be approximately six times faster for Layer 1 than Layer 3:  $k_{av}$  for Layer 1 was estimated to be  $6.33 \times 10^{-5}$  per site per year, while for Layer 3,  $k_{av}$  was estimated to be  $9.95 \times 10^{-6}$  per site per year (Table 3.2). This translates to an estimated true average fragment length in the DNA extract of 16 bp for Layer 1, and 19 bp for Layer 3. Layer 3 also contained more unique marsupial DNA than Layer 1 (Figure 3.3). This supports the finding that Layer 3 also had lower average  $C_T$  values in the qPCR assays than Layer 1 for both primer sets (Table 3.1), indicating the presence of more template copies of DNA, despite amplification being less efficient in Layer 3 (Table S3.8.9). Note that although the decay rate of mtDNA



**FIGURE 3.4** | FRAGMENT LENGTH DISTRIBUTIONS OF UNIQUE MARSUPIAL SHOTGUN READS, THE PROPORTION OF ENDOGENOUS DNA EXPECTED TO BE REMAINING GIVEN A CERTAIN TIME, AND THE PROBABILITY OF SURVIVAL OF 50,000 YEARS AFTER CELL DEATH FOR VARIOUS FRAGMENT LENGTHS. **A** Fragment length distributions of unique marsupial shotgun reads for Layer 1 (purple), 3 (orange), and 6 (blue) showing number of sequences (out of 60,000 subsampled sequences) for each read length (bp) that mapped to marsupial sequences by BLASTn to GenBank’s nr nucleotide database. Tables embedded within the histogram show summary statistics for the distribution. The slope ( $\lambda$ ) and fit ( $R^2$ ) of the exponential relationship modeled is also shown. **B** The proportion of unique marsupial 30 bp sequences expected to be remaining given a certain time (years after cell death), as predicted by the average (thick line), minimum and maximum (thin lines) decay constants calculated from both Layer 1 (purple) and 3 (orange). **C** The probability of survival of 50,000 years after cell death for various fragment lengths (bp), as predicted by the average (thick line), minimum and maximum (thin lines) decay constants calculated from both Layer 1 (purple) and 3 (orange).

is slower than that of nuDNA (Allentoft et al. 2012), the proportion of marsupial DNA that mapped to annotated mitochondrial reference genomes was only between 1-2%—too little to quantify the mtDNA decay rate. However, this may be possible to determine with increased sequencing depth in the future.

These results suggest that DNA preservation at RCEC is influenced by more than age alone (Haynes 2002; Hansen et al. 2006). The discrepancy between the age and the observed DNA preservation “likely derives from differences in taphonomy or bone diagenesis”, including microbial digestion, bone thickness (Allentoft et al. 2012), and other conditions within the cave, such as oxygenation, moisture, radiation (Campos et al. 2012; Sawyer et al. 2012), pH, salinity, redox potential, and temperature (Allentoft et al. 2012; Smith et al. 2003). Sediment type (Rawlence et al. 2014) and character, such as “grain size, mineralogical composition, organic matter load” (Corinaldesi et al. 2008), the presence of humics and humates (Allentoft et al. 2012), movement, compaction, and lithification (Eglinton and Logan 1991) can all further influence aDNA preservation. In addition, the physical aspects of caves such as entrance type and size, depth in limestone, surface topography, and relative depth of passage all influence how materials are accumulated, as well as the frequency, volume and direction of water flows into caves, and the stability or otherwise of within-cave environmental conditions (Reed 2003, 2009). Depurination may also not be the only mechanism of DNA damage in caves within the NCWHA: DNA can be damaged by other mechanisms, such as cross-linking, and blocking and miscoding lesions (Hoss et al. 1996; Hofreiter et al. 2001; Gilbert 2003). Although these mechanisms can play a greater role in DNA damage than depurination (Hansen et al. 2006), they have not been taken into account in this model of the DNA decay. Together, these factors are expected to play a large role in the likelihood of aDNA preservation, which may explain why aDNA has not been recovered from any other cave in the NCWHA before now. This lack of sole dependence of DNA degradation on age could explain why mtDNA preservation appeared to be largely consistent up to 8,000 cal a BP after which it ‘drops off’, but remains consistently low for the next 10,000 years, or why DNA decay in Layer 3 has proceeded more slowly than Layer 1. It is possible that something about the environment ca. 8,000 ya may have been more favourable for DNA survival. For example, specimens from Layer 3 may have been more rapidly buried than Layer 1. Another possibility is that Layer 3 has been

disturbed less than Layer 1, which may have been compacted and contaminated by recent human and animal activity in the cave.

The decay rates for Layers 1 and 3 were used to estimate the molecular half-life of a 30 bp DNA fragment (Table 3.2) and the likelihood of such a minimally informative DNA sequence surviving through time. Such models can be useful to predict the probability of recovering informative DNA from megafaunal species (e.g., sthenurine kangaroos) in the 40,000–50,000 cal a BP period over which they became extinct (Llamas et al. 2014). Assuming that the decay rate of DNA in older layers is comparable to those in Layers 1 or 3, modeling (Figure 3.4 b and c) suggests that after 50,000 years, the proportion of unique marsupial 30 bp sequences in an extract, as well as the probability of a 30 bp fragment surviving 50,000 years, is expected to be between  $10^{-42}$  and  $10^{-7}$  (Table 3.2). A ‘best case scenario’ would predict that 1 in 10 million 30 bp fragments would still be available today, which would be equivalent to about 10 random 30 bp DNA fragments surviving per cell, assuming a genome size of about 2.9 Gbp (the size of the Tammar wallaby, *Macropus eugenii*, genome; Renfree et al. 2011). Even if only a few hundred cells survive in a bone sample, it cannot be excluded that a small amount of authentic DNA could be obtained from much older fossils. We have also shown that the DNA fragmentation at RCEC is not entirely dependent on age, meaning that the DNA loss in megafaunal bones from the older layers may have been slower than the younger layers. Improvements in DNA technology such as target enrichment through hybridisation capture and higher sequencing coverage may improve the chances of retrieving low-copy number DNA (Llamas et al. 2014), particularly mtDNA. Post-excavation conditions will also influence the retrieval of aDNA. For instance, it has been observed that recently excavated, dry sieved, and untouched fossil bones contain more endogenous DNA than bones that have been stored and excavated under standard procedures (Pruvost et al. 2007). Other factors including exposure to UV radiation, thermal stress, human contact, and wet sieving all reduce the amount of DNA that can be obtained (Bollongino et al. 2008). Because the bones sampled here were handled and washed (albeit out of necessity), it is highly likely that DNA retrieval was adversely affected, which could explain the stochasticity of amplification (and contamination) observed in the older layers. This may also have influenced estimates of the decay rate. As such, we would expect DNA retrieval to be even better from fossils that have been

**TABLE 3.2** | EMPIRICAL DECAY CONSTANT ( $k_{AV}$  PER SITE PER YEAR) FOR LAYERS 1, 3, AND 6, modeled using  $\lambda$  as determined from the exponential part of the fragment length distributions (Figure 3.4) and the average, maximum, and minimum estimates of bone age (Figure 3.2) (where  $k = \lambda/\text{age}$ ; Allentoft et al. 2012). The decay constant was not calculated for Layer 6 because the sample size (i.e., number of fragments in the distribution) was too small to accurately estimate  $k$ . The GC content (% of bases either G or C), and T content bias (% of bases T) at the 5' and 3' ends of the unique marsupial sequences was also calculated for Layer 1, 3, and 6 based on the shotgun data.

Parameter	Layer 1	Layer 3	Layer 6
Av. empirical decay rate, $k_{av}$ (per site per year) (Min - Max)	$6.33 \times 10^{-5}$ ( $5.88 \times 10^{-5}$ - $6.84 \times 10^{-5}$ )	$9.95 \times 10^{-6}$ ( $6.31 \times 10^{-6}$ - $1.43 \times 10^{-5}$ )	NA
Av. fragment length of extract (bp)	16	19	NA
Av. fragment length of extract after 50,000 yr (bp)	0.31 (0.29-0.34)	2 (1.40-3.17)	NA
Av. number of years until the av. fragment length is 1 bp (Min - Max)	15,806 (14,619-17,007)	100,452 (69,930-158,479)	NA
Av. $k_{30}$ (per site per year), 30 bp (Min - Max)	$1.9 \times 10^{-3}$ ( $1.8 \times 10^{-3}$ - $2.0 \times 10^{-3}$ )	$3.0 \times 10^{-4}$ ( $1.9 \times 10^{-4}$ - $4.3 \times 10^{-4}$ )	NA
Av. half life (year), 30 bp (Min - Max)	366 (393-338)	2321 (3622-1616)	NA
Av. proportion of 30 bp fragments left after 50,000 years (Min - Max)	$6.7 \times 10^{-42}$ ( $3.1 \times 10^{-45}$ - $5.4 \times 10^{-39}$ )	$3.3 \times 10^{-7}$ ( $4.98 \times 10^{-10}$ - $7.8 \times 10^{-5}$ )	NA

<b>Av. probability of a 30 bp fragment surviving 50,000 years (Min - Max)</b>	$6.1 \times 10^{-42}$ ( $2.8 \times 10^{-45}$ - $5.0 \times 10^{-39}$ )	$3.3 \times 10^{-7}$ ( $4.8 \times 10^{-10}$ - $7.6 \times 10^{-5}$ )	NA
<b>GC content (%)</b>	42.5	43.1	43.8
<b>T content first base (%)</b>	40.2	43.6	37.5
<b>T content last base (%)</b>	63.0	63.0	75.0

freshly excavated under sterile conditions. While there remains the potential for aDNA to be retrieved from even older layers within RCEC, it will always be difficult to predict the cumulative effects of the many factors that could influence DNA preservation and the probability of capturing megafaunal DNA from Naracoorte; ultimately, it needs to be empirically tested.

Nevertheless, the aDNA preservation in bone at RCEC is largely comparable to that of other cave systems around Australia that have yielded aDNA. For instance, the sediments found in the Kelly Hill Caves complex on Kangaroo Island, South Australia are made up of primarily silty sand and are largely similar to the sediments of the NCWHA caves, including RCEC (McDowell et al. 2013). Mammal aDNA between 115–175 bp was amplified from bones preserved at this site that were between ca. 6800 to over 20,000 cal a BP (Haouchar et al. 2014). However, several layers between 6800–10,000 cal a BP and over 20,000 cal a BP did not yield mammal DNA (Haouchar et al. 2014) and the mammalian diversity identified through aDNA was much lower than this study (see section 3.5.2 below)—about four families across 15 layers. This may indicate that fragment length sizes were smaller than could be amplified with the primer sets used, which suggests that they may be shorter than 100 bp—similar to fragment lengths observed in this study. In another study, aDNA was isolated from archaeological bulk bone samples up to 46,000 years old from southwest Australia (Devil’s Lair and Tunnel Cave; Murray et al. 2013). Using two mammalian specific primer sets to amplify 150 bp, eight mammalian families were detected spanning ca. 4,000–22,000 cal a BP and 10,000–46,000 cal a BP at Tunnel Cave and Devil’s lair, respectively (Murray et al. 2013). At these sites, aDNA preservation is clearly exceptional; in comparison, much shorter fragments were retrieved from much younger samples at RCEC. More recently, 37–121 unique mtDNA sequences averaging 37–40 bp were recovered from 40,000–50,000 cal a BP megafaunal bone from high altitude caves in Mt Cripps, Tasmania (Llamas et al. 2014); however, no DNA was able to be amplified through PCR, and libraries were enriched for endogenous DNA. While there were too few endogenous sequences to estimate  $k$  for Layer 6, we still managed to retrieve mtDNA and nuDNA from bones over 18,000 years old without enrichment. This further supports the notion that enrichment may be necessary to capture megafaunal DNA from Naracoorte, but also that the likelihood of success could be even smaller given the warmer climate of

Naracoorte compared with Mt Cripps. To date, no study has conducted a thorough investigation into the factors that have the greatest impact on aDNA preservation in Australian environments. Thus, it would be highly informative to compare aDNA preservation between a variety of environmental conditions in order to identify the factors that are most conducive to aDNA preservation; this will assist in predicting those sites with the likelihood of best aDNA preservation.

### 3.5.2 BIODIVERSITY

DNA attributable to a variety of taxa was amplified and sequenced with high coverage from all six layers of RCEC using two primer sets covering mitochondrial loci of differing lengths (Table 3.3). Approximately 300,000 and 100,000 sequence reads were generated in total for the 12SAO and 12SMarsMini primer sets (Table S3.8.5), yielding 103 bp and 41 bp of informative sequence, respectively. We were able to identify a range of marsupial and placental mammal, bird, and amphibian families from each layer spanning the ages of ca. 950–18,500 cal a BP, including all of the small mammal families that were identified morphologically by Macken and Reed (2013). In addition, we identified several large-bodied marsupial families that have not been previously described from RCEC fossil record (Vombatidae, Macropodidae), as well as families that can be difficult to distinguish morphologically (e.g., amphibian taxa, such as *Limnodynastes*, *Litoria*). This adds weight to the validity of the both the morphological and molecular taxonomic identifications, and is yet another example of how aDNA analysis of bulk bone is sensitive enough to rapidly capture a large portion of the biodiversity (Murray et al. 2013; Haouchar et al. 2014), including some hitherto undescribed biodiversity. Of the taxa detected, highly credible native animal families (90–95% sequence similarity between query and reference, a representative of all other members present in GenBank, and is known in the region) include Vombatidae (wombats), Petauridae (striped and gliding possums), Dasyuridae (carnivorous/insectivorous marsupials), Burramyidae (pygmy possums), Potoroidae (potoroos), Peramelidae (bandicoots), Macropodidae (kangaroos and wallabies), Phalangeridae (possums), Pseudocheiridae (ring-tailed possums), Miniopteridae (long-winged bats), Muridae (rodents), Rhipiduridae (fantails), Columbidae (pigeons), Cracticidae (magpies), Hylidae (Australasian tree frogs) and Myobatrachidae (Australian ground frogs) (Table 3).

**TABLE 3.3** | SUMMARY OF THE TAXA IDENTIFIED FROM RCEC FOR LAYER 1-6 by comparison of sequence information for the metabarcoding primer sets 12SAO (regular type-face) and 12SMarsMini (bold type-face) to GenBank’s nr nucleotide database using BLASTn. Sequence similarity between query and reference cut-offs (with 100% query coverage) for each taxonomic level include: 90%-95% for family, 96-97% for genus, and 98%-100% for species (note that for the 12SMarsMini primer set, species are only considered when the similarity to the reference is 100%). Taxonomic assignments are classified as: ‡ highly credible, being within the cut-off percentage, all other members present in GenBank, and is known to region; † credible, being within the cut-off percentage, but where other members of the taxon are equally possible (equal similarity to multiple members or other members not present in GenBank); Δ uncertain, being within the cut-off percentage, but not all genus members present in Genbank and isn’t known from region; and § inferred, not within the cut-off but no other genus members described, and is known from the region. \* Indicates the possible presence of an extinct genus known from the region, but not represented in GenBank; # indicates the taxon was identified morphologically from RCEC in Macken and Reed (2013); ^ indicates the taxon was identified morphologically in single-source bones from the same accumulation (unpublished). Shaded squares indicate that the taxon was detected in at least two or more replicates (either subsample replicates or extraction replicates).

Family 90-95% similarity	Genus 96-97% similarity	Species 98-100% similarity	Common name	Layer						
				1	2	3	4	5	6	
Acrobatidae#	<i>Acrobates#</i>					‡		‡	‡	
						‡		‡	‡	
		<i>pygmaeus#</i>				‡		‡	‡	
Vombatidae^	<i>Vombatus^</i>			‡/‡				‡		
			Common wombat	‡/‡				‡		
		<i>ursinus^</i>		‡/‡				§		
Petauridae#	<i>Petaurus#</i>					‡				
		<i>breviceps#</i>	Sugar glider			†				
Dasyuridae#^	<i>Dasyurus#</i>			‡	‡	‡	‡/‡	‡		
			Quoll		‡					
Burramyidae#^	<i>Cercartetus#</i>					‡			‡	
						‡			‡	

Marsupial#^			Eastern pygmy possum				‡							
		<i>nanus#</i>												
	Potoroidae#						‡	‡	‡/‡	‡/‡	‡			
		<i>Bettongia#</i>		Bettong			‡	‡	‡/‡	‡/‡				
	Peramelidae#						‡/‡	‡	‡/‡	‡/‡	‡	‡/‡		
		<i>Isoodon#</i>		Short nosed bandicoots			†							
	Macropodidae^													
		<i>Lagorchestes*</i>					‡/‡	‡/‡	‡/‡	‡/‡	‡/‡	‡		
		<i>Onychogalea*</i>							‡	‡	‡	‡		
				Northern nail tail wallaby					‡/‡	‡				
		<i>Macropus unguifera*</i>							Δ / Δ	Δ		Δ		
		<i>Macropus agilis</i>		Agile wallaby			‡	‡/‡			‡	‡		
				Western grey kangaroo			†				†			
		<i>fuliginosus</i>		Red-necked wallaby										
		<i>rufogriseus</i>												
Phalangeridae#^														
	<i>Trichosurus#</i>					‡/‡	‡/‡	‡/‡	‡/‡	‡/‡	‡			
			Common brushtail possum			‡/‡	‡/‡	‡/‡	‡/‡	‡/‡	‡			
Pseudocheiridae#^														
	<i>Pseudocheirus#</i>					†/†	†/†	†/†	†/†	†/†	†			
			Common ringtail possum											
	<i>peregrinus#</i>					‡	‡	†/†	‡					
Placental														
	Miniopteridae^					‡	‡	‡			‡			
	<i>Miniopterus fuliginosus</i>		Eastern bent-winged bat			‡	‡	‡			‡			
						Δ	Δ	Δ			Δ			
Muridae^						‡	‡	‡	‡	‡	‡			
	<i>Rattus</i>													
Rhipiduridae														
	<i>Rhipidura fuliginosa</i>		New Zealand fantail								Δ			
Avian^														
	Columbidae					‡	‡			‡				

Amphibian	<i>Phaps</i>		‡	‡	‡			
		<i>chalcoptera</i>	†	†	†			
	Cracticidae/ Artamidae					‡		
		<i>Cracticus</i>				‡		
		<i>tibicen</i>				†		
	Hylidae			‡	‡	‡		‡
	<i>Litoria</i>		‡	‡	‡		‡	
Myobatrachidae				‡	‡	‡	‡	
	<i>Limnodynastes</i>						‡	

Highly credible species assignments could not be made for most of the native taxa due to the incompleteness of reference databases, the possibility that sequences may derive from extinct (or extirpated) species, and that DNA damage, amplification error and sequencing error can confound evolutionary change leading to false identifications (Greal et al. 2015). This is particularly true for the very small 12SMarsMini amplicons because, depending on the locus, a few base changes could result in the misidentification of species; however, we estimate that a mutation rate of 3% (well above the error rate of *Illumina* sequencing platforms) would result in a genus misidentification only 0.1% of the time, with family misidentifications being even rarer (S3.8.20; Table S3.8.11, Table S3.8.12). Misidentifications at the genus and family-level are further reduced by stringent quality and abundance filtering, high coverage, independent confirmation with another primer set, and replication with multiple extractions. Where species identifications were able to be confidently made, the taxa were either not detected using multiple primer sets (e.g., *Cercartetus nanus*; Table 3.3) or were not detected in more than one extract within a layer (e.g., *Pseudocheirus peregrinus*; Table 3.3), which casts some doubt on their authenticity due to lack of repeatability. Nevertheless, the presence of regionally extinct genera such as *Lagorchestes* support the authenticity of the aDNA obtained, as it is extremely unlikely that these sequences could have originated from modern contamination. Non-native taxa are likely to be contaminants (e.g., *Homo sapiens*, *Ovis*; S3.8.19, Table S3.8.10) as they are very well described in GenBank and there is no evidence of their presence at Naracoorte between 950 and 18,500 cal a BP. Improvements in bulk bone approaches, such as enrichment through the use of phylogenetically broad baits in hybridisation capture, may potentially improve the resolution of the molecular taxonomic identifications, remove contaminating sequences, and recover a larger proportion of the biodiversity, allowing us to better characterise biodiversity change through time.

The results presented here suggest that aDNA analysis of fossils at Naracoorte can be applied to investigate a range of palaeontological questions regarding adaptation of animal taxa to environmental change, as well as population shifts in space and time. For example, it is hypothesised from fossil evidence that *Petaurus norfolcensis* underwent a range expansion into the Naracoorte region during the late Pleistocene; however, modern genetic analysis of SA populations remains inconclusive (Macken

and Reed 2013). Targeted aDNA approaches (for instance, use of species-specific primers, whole genome sequencing, or hybridisation capture) using single source bone from fossil specimens of *P. norfolcensis* have the potential to clarify this. In addition, the identification of extinct taxa can help investigate extinction processes: the genus *Lagorchestes* was identified with high confidence, but the only member of the genus known from SA is the extinct species *Lagorchestes leporides*, which does not have any representative sequences in GenBank. Similarly, the genus *Onychogalaea* identified may represent the extinct species *O. lunata* or the locally extirpated species *O. fraenata*. Information about such species can be used investigate the timing, mechanisms, and causes of past biodiversity change, which is an important step towards better conservation and management of endangered extant relatives (Bohman et al. 2014, Pansu et al. 2015). Thus, when integrated with paleontological and geological data, aDNA from Naracoorte has the potential to offer new insights into historical faunal turnover in South Australia. Such approaches may also be able to resolve ambiguous species IDs, such as *Pseudomys novaehollandiae*, which is difficult to distinguish morphologically from *P. apodemoides* when dentaries are absent (Macken and Reed 2013).

### 3.6 CONCLUSION

---

The World Heritage listed Naracoorte Caves, SA are of significant educational and environmental importance due to their extensive paleontological fossil deposits. These deposits play an integral role in generating an appreciation for scientific research and cultivating awareness of environmental change and human impact in Australia. While many in-depth morphological studies of these fossils have been carried out, aDNA research at these caves has been limited despite its potential to help answer a variety of palaeoecological questions. Here, the successful purification of aDNA from Holocene and late Pleistocene bulk bone samples from RCEC within the NCWHA, coupled with NGS technology, enabled an assessment of aDNA preservation at this site that was found to be comparable to other Australian cave sites. For the first time, it was possible to identify a range of marsupial and placental mammal, bird, and amphibian families and genera with high confidence using aDNA as old as ca. 18,500 cal a BP, some of which are extinct, or may be difficult to distinguish morphologically. While we were able to recover aDNA from layers

dating back as far as the LGM, there remains a small possibility that aDNA may be retrieved from even older specimens within RCEC; however, future improvements in technology may improve this likelihood. Thus, our results suggest that further aDNA research at Naracoorte is warranted, especially since aDNA preservation in other organic materials from the caves, such as plants (including leaves, nuts, seeds, and cones), pollens, and sediments has yet to be assessed. This study shows that aDNA can be a valuable and complementary tool for understanding the past, and will add to the continuing impact of the NCWHA to palaeontological research in Australia.

### 3.7 REFERENCES

---

- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B-Biological Sciences* **1745**.
- Allentoft ME. 2013. Recovering samples for ancient DNA research—guidelines for the field archaeologist. *Antiquity* **87**: 338.
- Allentoft ME, Heller R, Oskam CL, Lorenzen ED, Hale ML, et al. 2014. Extinct New Zealand megafauna were not in decline before human colonization. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4922-4927.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Ayliffe LK, Marianelli PC, Moriarty KC, Wells RT, McCulloch MT, et al. 1998. 500 ka precipitation record from southeastern Australia: evidence for interglacial relative aridity. *Geology* **26**: 147-150.

- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Ecology and Evolution* **29**: 358-367.
- Blankenberg D, Von Kuster G, Coraro N, Ananda G, Lazarus R, et al. 2010. Galaxy: a web-based genome analysis tool for experimentalists. In: *Current protocols in molecular biology* (ed. al., FMAe) **Unit 19.10**: 11-21.
- Bollongino R, Tresset A, Vigne JD. 2008. Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* **7**: 91-98.
- Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, et al. 2007. Patterns of damage in genomic DNA sequences from a Neanderthal. *Proceedings of the National Academy of Sciences* **104**: 14616-14621.
- Bronk Ramsey C. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**: 337-360.
- Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, et al. 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**: 172-175.
- Bunce M, Worthy TH, Phillips MJ, Holdaway RN, Willerslev E, et al. 2009. The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 20646-20651.
- Campos PF, Craig OE, Turner-Walker G, Peacock E, Willerslev E, et al. 2012. DNA in ancient bone—Where is it located and how should we extract it? *Annals of Anatomy-Anatomischer Anzeiger* **194**: 7-16.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335-336.

- Cooper A. 1994. Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthistidae). *Experientia* **50**: 558-563.
- Corinaldesi C, Beolchini F, Dell'Anno A. 2008. Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Molecular Ecology* **17**: 3939-3951.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15758-15763.
- Darrénougué ND, De Deckker P, Fitzsimmons KE, Norman MD, Reed L, et al. 2009. A Late Pleistocene record of aeolian sedimentation in Blanche Cave, Naracoorte, South Australia. *Quaternary Science Reviews* **28**: 2600-2615.
- Eglinton G, Logan GA. 1991. Molecular Preservation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **333**: 315-328.
- Elsner J, Schibler J, Hofrieter M, Schlumbaum A. 2014. Burial condition is the most important factor for mtDNA PCR amplification success in Palaeolithic equid remains from the Alpine foreland. *Archaeological and Anthropological Sciences* **6**: 4.
- Fallon S, Fifield L, Chappell J. 2010. The next chapter in radiocarbon dating at the Australian National University: Status report on the single stage AMS. *Nuclear Instruments and Methods in Physics research* **268**: 298-901.
- Forbes MS, Bestland EA. 2007. Origin of the sedimentary deposits of the Naracoorte Caves, South Australia. *Geomorphology* **86**: 369-392.

- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Gilbert MTP, Willerslev E, Hansen AJ, Barnes I, Rudbeck L, et al. 2003. Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics* **72**: 779-779.
- Goecks J, Nekrutenko A, Taylor J, The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biology* **11**.
- Grealy A, McDowell M, Scofield P, Murray D, Fusco D, et al. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional palaeontological methods. *Quaternary Science Reviews* **128**: 37-47.
- Gutiérrez-García TA, Vázquez-Domínguez E, Arroyo-Cabrales J, Kuch M, Enk J. 2014. Ancient DNA and the tropics: a rodent's tale. *Biology Letters* **10**.
- Hadly EA, Ramakrishnan U, Chan YL, van Tuinen M, O'Keefe K, et al. 2004. Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLOS Biology* **2**: 1600-1609.
- Haile J. 2012. Ancient DNA extraction from soils and sediments. *Methods in molecular biology* **840**: 57-63.
- Hansen AJ, Mitchell DL, Wiuf C, Panikert L, Brand TB, et al. 2006. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* **173**: 1175-1179.
- Haouchar D, Haile J, McDowell MC, Murray DC, White NE, et al. 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews* **84**: 56-64.

- Haynes S, Searle JB, Bretman A, Dobney KM. 2002. Bone preservation and ancient DNA: the application of screening methods for predicting DNA survival. *Journal of Archaeological Science* **29**: 585-592.
- Heintzman PD, Elias SA, Scott A, Moore K, Paszkiewicz K, et al. 2014. Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). *Molecular Ecology Resources* **14**.
- Hofreiter M, Betancourt JL, Sbriller AP, Markgraf V, McDonald HG. 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quaternary Research* **59**: 364-378.
- Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Pääbo S. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* **29**: 4793-4799.
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24**: 1304-1307.
- Hunter AA, Macgregor AB, Szabo TO, Wellington CA, Bellgard MI. 2012. Yabi: An online research environment for grid, high performance and cloud computing. *Source code for biology and medicine* **7**: 1-1.
- Jaenicke-Després V, Buckler ES, Smith BD, Gilbert MTP, Cooper A, et al. 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206-1208.
- Jørgensen T, Kjaer KH, Haile J, Rasmussen M, Boessenkool S, et al. 2012. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical

records and sedimentary ancient DNA meta-barcoding. *Molecular Ecology* **21**: 1980-1988.

Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage— Building and working in an ancient DNA laboratory. *Annals of Anatomy- Anatomischer Anzeiger* **194**: 3-6.

Kuhn TS, McFarlane KA, Groves P, Mooers AØ, Shapiro B. 2010. Modern and ancient DNA reveal recent partial replacement of caribou in the southwest Yukon. *Molecular Ecology* **19**: 1312-1323.

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, et al. 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**: 1613-1616.

Macken AC and Reed EH. 2013. Late Quaternary small mammal faunas of the Naracoorte Caves World Heritage area. *Transactions of the Royal Society of South Australia* **137**: 53-67.

Macken AC and Reed EH. 2014. Postglacial reorganisation of a small-mammal paleocommunity in southern Australia reveals thresholds of change. *Ecological Monographs* **84**: 563-577.

Macken AC. 2009. Palaeoecological investigation of the Grant hall fossil deposit, Victoria Fossil Cave, naracoorte, South Australia. In: *Flinders University Flinders University South Australia, Australia*.

Macken AC, Prideaux GP, Reed EH. 2012. Variation and pattern in responses of mammal faunas to late Pleistocene climatic change in south eastern South Australia. *Journal of Quaternary Science* **27**: 415–424.

Magyari EK, Major A, Bálint M, Nédli J, Braun M, et al. 2011. Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evolutionary Biology* **11**.

- McDowell MC, Bestland EA, Bertuch F, Ayliffe LK, Hellstrom JC, et al. 2013. Chronology, strigraphy and palaeoenvironmental interpretation of a Late Pleistocene to mid-Holocene cave accumulation on Kangaroo Island, South Australia. *BOREAS* **42**: 974-994.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS One* **10**.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. 2013. Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific Reports* **3**: 3371.
- Murray DC, Pearson SG, Fullagar R, Chase BM, Houston J, et al. 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews* **58**: 135-145.
- O'Keefe K, Ramakrishnan U, Van Tuinen M, Hadly EA. 2009. Source-sink dynamics structure a common montane mammal. *Molecular Ecology* **18**: 4775-4789.
- Olade I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**: 225-228.
- Orlando L, Ginolhac A, Raghavan M, Vilstrup J, Rasmussen M, et al. 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research* **21**: 1705-1719.

- Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, et al. 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **277**: 1991-2000.
- Otoni C, Koon HE, Collins MJ, Penkman KE, Craig OE. 2009. Preservation of ancient DNA in thermally damaged archaeological bone. *Naturwissenschaften* **96**: 257-278.
- Pansu J, Giguet-Covex C, Ficetola GF, Gielly L, Boyer F, et al. 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Molecular Ecology* **24**: 1485-1498.
- Prideaux GJ, Roberts RG, Megirian D, Westaway KE, Hellstrom JC, et al. 2007. Mammalian responses to Pleistocene climate change in southeastern Australia. *Geology* **35**: 33-36.
- Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, et al. 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 739-744.
- Ramsey BC. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**: 337-360.
- Rawlence NJ, Wood JR, Armstrong KN, Cooper A. 2009. DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B-Biological Sciences*, **276**: 3395-3402.
- Rawlence NJ, Lowe DJ, Wood JR, Young JM, Churchman GJ, et al. 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science* **29**: 610-626.

- Reed EH. 2003. Vertebrate taphonomy of large mammal bone deposits, Naracoorte Caves World Heritage Area. Unpublished PhD Thesis, Flinders University of South Australia.
- Reed EH. 2009. Decomposition and disarticulation of kangaroo carcasses in caves at Naracoorte, South Australia. *Journal of Taphonomy* **7**: 265-283.
- Reed EH. 2012. Of mice and megafauna: new insights into Naracoorte's fossil deposits. *Journal of the Australasian Cave and Karst Management Association* **86**: 7-14.
- Renfree MB, Papenfuss AT, Deakin JE, Lindsay J, Heider T, et al. 2011. Genome sequence of an Australian Kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biology* **12**.
- Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS One* **7**.
- Shapiro B and Hofreiter M. 2012. *Ancient DNA: methods and protocols*. Humana Press (Springer Science), New York.
- Shepherd LD and Lambert DM. 2008. Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* **17**: 2174-2184.
- Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**: 203-217.
- Wells RT, Moriarty K, Williams DLG. 1984. The fossil vertebrate deposits of Victoria Fossil Cave, Naracoorte: an introduction to the geology and fauna. *The Australian Zoologist* **21**: 305-333.

- Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791-795.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Wood JR, Wilmshurst JM, Rawlence NJ, Bonner KI, Worthy TH, et al. 2013a. A megafauna's microfauna: gastrointestinal parasites of New Zealand's extinct moa (aves: Dinornithiformes). *PLOS One* **8**.
- Wood JR, Wilmshurst JM, Richardson SJ, Rawlence NJ, Wagstaff SJ, et al. 2013b. Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes). *Proceedings of the National Academy of Sciences of the United States of America* **110**: 16910-16915.
- Wood JR, Wilmshurst JM, Worthy TH, Cooper A. 2012. First coprolite evidence for the diet of *Anomalopteryx didiformis*, an extinct forest ratite from New Zealand. *New Zealand Journal of Ecology* **36**: 164-170.

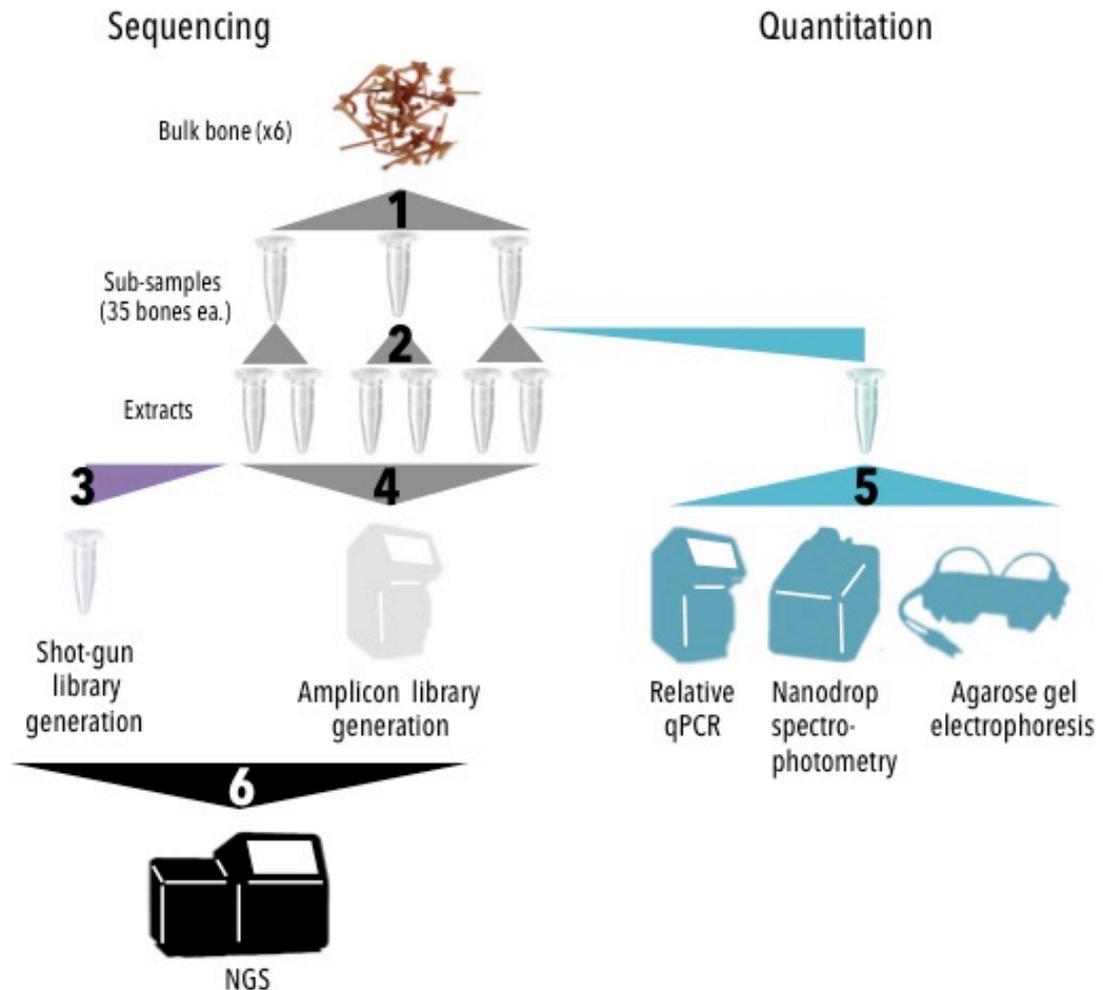
### 3.8 SUPPLEMENTARY INFORMATION

#### S3.8.1 RADIOCARBON DATING

**TABLE S3.8.1** | RADIOCARBON DATES FOR LAYERS 1 THROUGH 6 (SQUARE 1) FROM ROBERTSON CAVE ENTRANCE CHAMBER. Radiocarbon dates were calibrated using the SHCal13 calibration curve (Hogg 2013) in OxCal (Ramsey 2009).

Sample	Excavation layer	S-ANU#	d13C	Percent Modern Carbon (pMC)	14C age (cal a BP)	95.4% CI (2 sigma; cal a BP)	Sampled from	Depth below datum	Material
RCC5U17-1CDL1	Layer 1	27618	-24.7432 ±0.608	87.373 ±0.214	1085 ±30	906-1054	L1: 0 to 5 cm	26 cm	Charcoal
RCC5U17-1CDL2	Layer 2	27619	-23.6277 ±0.383	73.332 ±0.177	2490 ±30	2358-2705	L2: 0 to 5 cm	35 cm	Charcoal
RCC5U17-1CDL2	Layer 2	27632	-22.2624 ±0.349	52.783 ±0.192	5135 ±40	5731-5926	L2: 0 to 5 cm	Picked from bag	Bone
RCC5U17-1CDL3	Layer 3	27633	-16.7731 ±0.238	64.046 ±0.198	3580 ±35	3697-3958	L3: 0 to 5 cm	Picked from bag	Bone
RCC5U17-1CDL3 DUP	Layer 3	27635	-15.7852 ±0.0.129	63.888 ±0.195	3600 ±35	3718-3971	L3: 0 to 5 cm	Picked from bag	Bone
RCC5U17-1CDL3	Layer 3	27620	-24.048 ±0.567	39.067 ±0.129	7550 ±35	8202-8398	L3: 0 to 5 cm	45 cm	Charcoal
RCC5U17-1CDL4	Layer 4	27637	-24.2773 ±0.275	32.187 ±0.111	9105 ±40	10165-10366	L4: 5 to base	60 cm	Charcoal
RCC5U17-1CDL5	Layer 5	27621	-24.3256 ±0.612	15.770 ±0.087	14840 ±55	17820-18200	L5: 0 to 5cm	74 cm	Charcoal
RCC5U17-1CDL6	Layer 6	27623	-31.2762 ±0.447	15.004 ±0.081	15235 ±55	18294-18629	L6: 0 to 5 cm	83 cm	Charcoal

### S3.8.2 EXPERIMENTAL DESIGN



**FIGURE S3.8.1** | A DIAGRAM DEPICTING THE WORKFLOW OF THE METHODS (SEQUENCING AND QUANTITATION). One bulk bone sample from each of the six layers was divided into three subsamples of 35 bones each; two extractions were performed on bone powder from each subsample and the DNA was used to build both shotgun and amplicon sequencing libraries. One additional extraction was performed per layer for quantitation via qPCR, Nanodrop spectrophotometry, and agarose gel electrophoresis.

### **S3.8.3 IMPLEMENTATION of STANDARD ADNA PROTOCOLS**

Bones were kept in airtight containers at 4°C. Samples were prepared in an ancient DNA clean room preparation area. A full body suit, double gloves, mask, eyewear, boots, and a hood were worn as personal protective equipment, and to avoid contaminating the bone samples with human DNA. Before beginning, all surfaces (including tray, table, chair, electronic balance, and gloves) were cleaned with a solution of 10% bleach, followed by a solution of 70% ethanol (Willerslev and Cooper 2005; Knapp et al. 2012). A clean pot was used when beginning a new pool to avoid cross-contamination between pools. Pots were cleaned by rinsing with 10% bleach, followed by a scrubbing with DNAaway, followed by another scrubbing with 10% bleach, then rinsed in ultrapure water and dried with 100% ethanol. Pots were left for one hr in a UV hood for further decontamination. Pots were rinsed with 500 µL of Ultrapure water, which was subsequently extracted as an additional control. All pot controls were negative in all PCR reactions.

### **S3.8.4 MODIFICATIONS to DABNEY ET AL.'S (2013) ADNA EXTRACTION METHOD**

DNA was eluted in 30 µL EB buffer (*QIAGEN*, cat. No. 19086) warmed to 37°C. The eluent was then passed back through the column again after incubating for five minutes at room temperature. Eluent was stored in a 1.5 ml Lo-Bind *Eppendorf* tube containing 1.5 µL of 1% TE-Tween-20.

### **S3.8.5 MEASURING TOTAL GENOMIC DNA via SPECTROPHOTOMETRY**

The total DNA concentration (Table S3.8.7) of one extract from each layer was quantified by measuring 1 µL on a Nanodrop spectrophotometer (*Thermo Scientific*), following the manufacturer's instructions and using 1 µL of EB buffer (*QIAGEN*) to blank. The purity was measured via the 260/280 ratio reported by the Nanodrop spectrophotometer.

### **S3.8.6 VISUALISING GENOMIC DNA with AGAROSE GEL ELECTROPHORESIS**

5  $\mu\text{L}$  of genomic DNA from each layer was combined with 0.5  $\mu\text{L}$  of 6X loading dye and run on a 3% agarose gel electrophoresis (3.3 g agarose, 110 mL 1X TAE buffer, 8  $\mu\text{L}$  GelRed), alongside 3  $\mu\text{L}$  of 50 bp DNA ladder (GeneRuler, *Fermentas*). The gel was run for 1.5 hr at 70 V. The gel was visualised and photographed using a *BioRad* transilluminator. Any changes to the contrast and brightness of the image were applied to the whole image.

### S3.8.7 QPCR QUANTIFICATION of DNA and FUSION-TAGGING

**DETERMINING THE EFFICIENCY of AMPLIFICATION.** qPCR provides a way to assess the relative quantity and quality of the amplifiable DNA from each layer by measuring the accumulation of DNA through fluorescence as it amplifies exponentially in the PCR reaction. The amount of starting template molecules in an extract relative to another can be estimated by comparison of the  $C_T$ -values, the cycle at which DNA is amplified beyond a certain threshold of detection. The  $C_T$ -value is inversely related to the number of starting template molecules in an extract—the lower the  $C_T$ -value, the greater the number of starting template molecules. The quality of the DNA is also indicated by the amplification efficiency, which should ideally be between 95-105%. An amplification efficiency outside these limits can indicate the presence of contaminants that can inhibit PCR amplification. Efficiency is assessed through the generation of a standard curve, where the  $C_T$ -values of serial dilutions of a sample are plotted against the log concentration; the slope and  $R^2$  value of this relationship gives a measure of the amplification efficiency and how well the relationship fits the observed data. A reaction with perfect amplification efficiency would have a slope of -3.32 and an  $R^2$  value of  $> 0.98$ . The reproducibility of the reaction will also be reflected in the  $R^2$  value; for instance, replicate reactions with largely different  $C_T$ -values will lower the  $R^2$  value.

For each layer, the extract that was not used for fusion tagging was diluted 1/2, 1/5, 1/10, and 1/15 in ultrapure water, and each dilution was amplified in duplicate using both the 12SAO and 12SMarsMini primer sets, as below. No-template water controls were included and were negative for amplification. The cycle-threshold ( $C_T$ ) values for each layer at each dilution are recorded in Table S3.8.8. The percent efficiency for the reactions was calculated by plotting the log concentration against the  $C_T$ -

values for each layer (Figure S3.8.3), and using the formula: % efficiency =  $(10^{1/\text{slope}} - 1) * 100$ . The  $R^2$  value (Table S3.8.9) gives an estimate of the fit of the linear equation to the data. Amplification curves for each layer for one replicate of the neat, 1/5 and 1/10 dilutions are shown in Figure S3.8.4.

**SCREENING GENOMIC DNA for INHIBITION, DETERMINING RELATIVE TEMPLATE COPY NUMBER via QPCR, and FUSION TAGGING.** The neat and a 1/10 dilution of each extract were amplified using both 12SAO and 12SMarsMini primer sets. The PCR reaction contained reagents in final concentrations of: 0.4 mg/ml *Fisher* BSA, 1X *ABI* GeneAmp PCR buffer, 2 mM *ABI* MgCl<sub>2</sub>, 0.4 μM *IDT* forward primer (Table S3.8.2), 0.4 μM *IDT* reverse primer, 0.25 mM *Bioline* dNTPs, 2 U *ABI* AmpliTaq Gold DNA polymerase, 0.6 μL of a 1:2000 dilution of 10 000X SYBR green dye, and 2 μL of DNA in a total reaction of volume of 25 μL (including 14.2 μL *GIBCO* HPLC-grade water). Thermocycling conditions were: 95°C for five min, 50 cycles of 95°C for 30 sec, 57°C (12SAO) or 55°C (12SMarsMini) for 30 sec, and 72°C for 45 sec, followed by a melt curve of 95°C for 15 sec, 60°C for one min, 95°C for 15 sec, 72°C for 10 min. C<sub>T</sub>-values were recorded for each sample, and amplification and melt curves were analysed to give a rough indication of PCR efficiency. The dilution exhibiting the least inhibition and greatest PCR efficiency were selected for subsequent fusion-tagging (Table S3.8.2). Fusion-tag PCR reaction and cycling conditions were as above.

### **S3.8.8 MODIFICATIONS to THE AMPURE XP PCR PURIFICATION PROTOCOL**

1.2 μL of AMPure XP SPRI beads (*Beckman-Coulter*) was added to 1 μL of PCR product (i.e., 1.2X the volume of PCR product). Cleaned PCR products were eluted in 40 μL EB buffer (*QIAGEN*).

### **S3.8.9 VISUALISING PCR PRODUCTS and LIBRARIES with AGAROSE GEL ELECTROPHORESIS**

5 μL of PCR product was combined with 0.5 μL of 6X loading dye and run on a 2% agarose gel electrophoresis (2.2 g agarose, 110 mL 1X TAE buffer, 8 μL GelRed),

**TABLE S3.8.2 | PRIMERS USED FOR AMPLICON SEQUENCING.** The architecture of the fusion-tag indexed primers is: 5' - MiSeq P5 – 5' Sequencing Adapter – Index – Forward Gene Specific Primer– 3' (forward), and 5' – MiSeq P7 – Index – Reverse Gene Specific Primer -3' (reverse).

Name	Sequence (5'-3')	Target Taxa or Function	Gene	Reference	Amplicon size (bp)	Insert size (bp)	Annealing Temp (°C)
12SA (Forward)	CTGGGATTAGATACCCCACTAT	Bird and mammal (with 12SO)	12S	(Cooper 1994)	150	103	57
12SO (Reverse)	GTCGATTATAGGACAGGTCCTCTA	Bird and mammal (with 12SA)	12S	(Cooper 1994)			57
12SMarsMini (Forward)	TAGTTAGACCTACACATGCAAGTT	Marsupial (universal)	12S	(Haouchar et al. <i>unpublished</i> )	85	41	55
12SMarsMini (Reverse)	CCTGATACCCGCTCCTRTTR	Marsupial (universal)	12S	(Haouchar et al. <i>unpublished</i> )			55

alongside 3  $\mu\text{L}$  of 50 bp DNA ladder (GeneRuler, *Fermentas*). The gel was run for 40 minutes at 96 V. The gel was visualised and photographed using a *BioRad* transilluminator.

### **S3.8.10 PCR REACTION and THERMOCYCLING CONDITIONS for QUANTIFYING THE SEQUENCING LIBRARY**

The sequencing library was diluted 1/10, 1/100, 1/1000, 1/5000, 1/25000, 1/125000, 1/625000 in EB buffer (*QIAGEN*). Each dilution was qPCR amplified in duplicate along side a standard of known concentration ( $10^8$ ,  $10^6$ ,  $10^5$ ,  $10^4$ ,  $10^3$ ,  $10^2$  molecules) using primers complementary to the sequencing adapters, in order to quantify the number of template molecules in the final sequencing library. PCR ‘no-template controls’ were included. The PCR reaction contained reagents in final concentrations of: 1X *ABI* Power SYBR Master Mix, 0.4  $\mu\text{M}$  *IDT* forward primer P5, 0.4  $\mu\text{M}$  *IDT* reverse primer P7 (Table S3.8.3), and 2  $\mu\text{L}$  of library in a total reaction of volume of 25  $\mu\text{L}$  (including 8.5  $\mu\text{L}$  *GIBCO* HPLC-grade water). Thermocycling conditions were: 95°C for five min, 40 cycles of 95°C for 30 sec, 60°C for 45 sec, followed by a melt curve of 95°C for 15 sec, 60°C for one min, and 95°C for 15 sec.  $C_T$ -values were recorded and compared to the standard in order to calculate the number of copies in each dilution, and determine the volume of library to input into the sequencing reaction.

### **S3.8.11 MODIFICATIONS to THE MiSEQ 300 v.2 KIT**

14 billion input library template molecules were targeted for the sequencing reaction using a MiSeq 300 v.2 kit. 4.6  $\mu\text{L}$  of the 1/10 dilution of the library was combined with 13.4  $\mu\text{L}$  of EB buffer (*QIAGEN*), and 2  $\mu\text{L}$  of 1 M molecular biology-grade NaOH and incubated for five minutes at 25°C, then placed on ice. 10  $\mu\text{L}$  of this mixture was then added to 990  $\mu\text{L}$  of HT1 buffer, and placed on ice. 550  $\mu\text{L}$  of this mixture was then combined with 50  $\mu\text{L}$  of 20 pM denatured PhiX, and placed on ice. 600  $\mu\text{L}$  of this mixture was added to the reagent cartridge in slot 17. Slot 12 was pierced with a pipette tip, and the contents were removed using a Pasteur pipette and placed in a 1.5 mL *Eppendorf* tube. 3  $\mu\text{L}$  of the custom sequencing primer was added

**TABLE S3.8.3** | PRIMERS USED FOR QUANTIFY THE FINAL SEQUENCE LIBRARY BY QPCR.

Name	Sequence (5'-3')	Target Taxa	Annealing Temp (°C)
P5	AATGATACGGCGACCACCGAGATCTACAC	Forward qPCR quant primer	60
P7	CAAGCAGAAGACGGCATACGAGAT	Reverse qPCR quant primer	60

to this, and the mixture was placed back into slot 2 using a Pasteur pipette. The options selected when creating the sample sheet were: Other/ FastQ only/ Sample preparation kit = TruSeq LT/ No index reads/ Single end/ 325 cycles/ No custom primer/ No trimming.

### **S3.8.12 MODIFICATIONS *to* THE SINGLE-STRANDED LIBRARY BUILDING METHOD *by* GANSAUGE AND MEYER (2013)**

Modifications to the adapters used are listed in Table S3.8.4. An extraction control, no-template (water) control, and CL104 positive control were also included in the library building process. At step 1, reactions were performed in 0.2 ml 8-well PCR strip tubes. 12  $\mu$ L of DNA extract was used, and *Afu* UDG was replaced by Ultrapure water. At step 7, ligation products were stored overnight at -20°C. At step 13, tubes were incubated in a rotating hybridisation oven for 2 min at 65°C as opposed to a thermal shaker. At step 13, tubes were transferred to a thermal shaker pre-cooled to 15°C as opposed to a thermocycler. Steps 14, 15, 18, 19, 23, and 25 were performed in a thermal shaker. Step 22 was performed in a rotating hybridisation oven. At step 25, the supernatant was stored in a 1.5 mL Lo-Bind *Eppendorf* tube at -20°C. The qPCR standard used in step 27 is listed in Table S3.8.4. After step 28, the PCR products were run on a 2% agarose gel electrophoresis (S3.8.9) in order to confirm the library building process worked.

### **S3.8.13 PCR REACTION *and* THERMOCYCLING CONDITIONS *for* THE SHOTGUN LIBRARY FUSION-TAG/INDEXING PCR**

At step 30 (Gansauge and Meyer 2013), the libraries were amplified in quintuplicate with unique fusion-tag indexing primers (Table S3.8.4). The PCR reaction contained reagents in final concentrations of: 1X *ABI* Power SYBR Master Mix, 0.4  $\mu$ M forward indexing primer, 0.4  $\mu$ M reverse indexing primer, 1  $\mu$ L of neat library, made up to a total of 25  $\mu$ L final volume with HPLC-grade water. Thermocycling conditions were: 95°C for two min, 26 cycles of 95°C for 15 sec, 60°C for 30 sec, 72°C for one min. Replicate reactions were combined and purified using the AMPure XP PCR purification kit (*Beckman-Coulter*; S3.8.8). Step 33 was not

**TABLE S3.8.4** | LIGATION ADAPTERS AND SEQUENCING PRIMERS USED FOR SHOTGUN SEQUENCING (MODIFICATIONS FROM GANSAUGE AND MEYER, 2013). The architecture of the fusion-tag indexed primers is: **5'** - MiSeq P5 – RD1–Index—**5'** SS adapter **-3'** (forward), and **5'** - MiSeq P7—RD2—Index—**3'** SS adapter - **3'** (reverse). Unique indexes are represented by NNNNNNNN (any base) below.

Name	Sequence (5'-3')	Function	Annealing Temp (°C)
CL53	ACACGACGCTCTC-ddC	Double-stranded adapter, strand 1	RT
CL78	[Phosphate]AGATCGGAAG[C9Spacer]3[TEG-biotin]	Single-stranded adapter	60
CL105_106_Std	ACACTCTTCCCTACACGACGCTCTCCTCGTCGTTGGTATGGCTTCTAT CGUATCGATCGATCGACGATCAAGGCGAGTTACATGAAGATCGGAAGA GCACACGTCTGAACTCCAGTCAC	Synthetic qPCR standard	-
P5-RD1-Index-Fwd adapter	AATGATACGGCGACCACCGAGATCTACAC- TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-NNNNNNN- ACACTCTTCCCTACACGACGCTCTT	Forward indexing primer	60
P7-RD2-Index-Rev adapter	CAAGCAGAAGACGGCATAACGAGAT- GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-NNNNNNN- GTGACTGGAGTTCAGACGTGT	Reverse indexing primer	60
RD1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	Standard sequencing primer	-

performed. Libraries were quantified via qPCR as per S3.8.10. Libraries were pooled in equal amounts and the final shotgun sequencing library was quantified once again via qPCR as per S3.8.10.

#### **S3.8.14 MODIFICATIONS to THE MISEQ 150 v.3 KIT**

At step 34 (Gansauge and Meyer 2013), a standard sequencing primer was used (S3.8.4). 20 billion input library template molecules were targeted for the sequencing reaction using a MiSeq 150 v.3 kit. 4.0  $\mu\text{L}$  of the 1/100 dilution of the library was combined with 12  $\mu\text{L}$  of EB buffer (*QIAGEN*), and 2  $\mu\text{L}$  of 1 M molecular biology-grade NaOH and incubated for 5 min at 25°C, then placed on ice. 10  $\mu\text{L}$  of this mixture was then added to 990  $\mu\text{L}$  of HT1 buffer, and placed on ice. 550  $\mu\text{L}$  of this mixture was then combined with 50  $\mu\text{L}$  of 20 pM denatured PhiX, and placed on ice. 600  $\mu\text{L}$  of this mixture was added to the reagent cartridge in slot 17. The options selected when creating the sample sheet were: Other/ FastQ only/ Sample preparation kit = TruSeq LT/ No index reads/ Single end/ 175 cycles/ No custom primer/ No trimming.

### S3.8.15 AMPLICON SEQUENCING COVERAGE

**TABLE S3.8.5** | AMPLICON SEQUENCING READ ABUNDANCE OBTAINED FOR EACH EXTRACT (12SAO AND 12SMARSMini) AFTER QUALITY FILTERING, CHIMERA FILTERING, AND ABUNDANCE FILTERING. The number of unique reads are given in the brackets. Asterisked extracts were used for shotgun sequencing.

Extract ID	Layer	Subsample	Extract	Number of reads	
				12SAO	12SMarsMini
MB2339	1	1	a	11628 (27)	783 (31)
MB2340*		2		12579 (104)	3573 (31)
MB2341		3		19661 (14)	8480 (16)
MB2410		1	b	444 (137)	1422 (17)
MB2411		2		519 (168)	12724 (23)
MB2412		3		775 (205)	3868 (11)
MB2342	2	1	a	9906 (30)	9697 (36)
MB2343		2		15464 (31)	3173 (28)
MB2344		3		11511 (18)	2034 (9)
MB2413		1	b	364 (110)	16383 (28)
MB2414		2		455 (159)	14199 (20)
MB2415		3		583 (185)	7138 (12)
MB2345	3	1	a	10220 (22)	8207 (39)
MB2346*		2		15926 (32)	NA
MB2347		3		11417 (27)	NA
MB2416		1	b	369 (109)	9895 (42)
MB2417		2		472 (134)	2820 (15)
MB2418		3		19815 (68)	9057 (33)

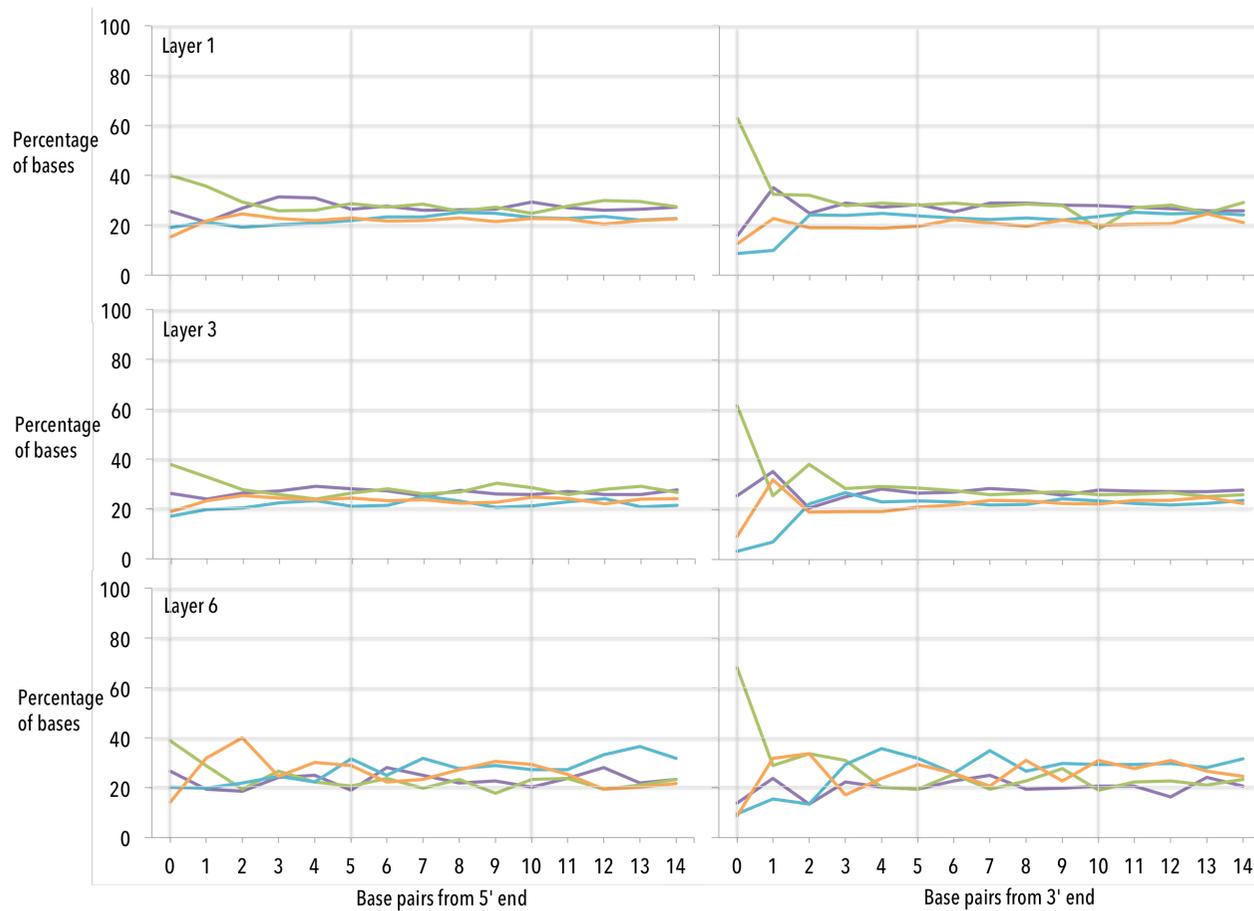
MB2348	4	1	a	5527 (qw19)	NA
MB2349		2		2 (2)	NA
MB2350		3		5190 (17)	2838 (21)
MB2419		1	b	9190 (28)	NA
MB2420		2		16158 (26)	293 (16)
MB2421		3		6034 (27)	993 (47)
MB2351	5	1	a	NA	NA
MB2352		2		2317 (4)	NA
MB2353		3		3731 (7)	NA
MB2422		1	b	2196 (19)	NA
MB2423		2		20553 (21)	145 (15)
MB2424		3		11386 (34)	108 (16)
MB2354	6	1	a	7634 (7)	NA
MB2355*		2		52167 (17)	NA
MB2356		3		NA	NA
MB2425		1	b	4054 (11)	60 (6)
MB2426		2		NA	69 (18)
MB2427		3		5488 (10)	23 (5)

### S3.8.16 BASE COMPOSITION ANALYSIS *from* SHOTGUN SEQUENCING DATA

**TABLE S3.8.6** | BASE COMPOSITION OF 15 BASES FROM THE 5' END AND THE 3' END OF UNIQUE MARSUPIAL SHOTGUN READS (PERCENTAGE OF BASES) FOR LAYERS 1, 3, AND 6.

Layer	End	Bases	Bases from end														
			0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	5'	A	27.4	24.5	26.4	30.4	32.8	26.2	29.4	26.4	26.7	28.4	28.5	29.8	28.4	26.4	31.4
		T	40.2	32.4	30.6	27.4	25.3	29.7	27.7	28.4	26.2	26	24.5	28.8	30.9	28.5	26.2
		C	17.4	19.9	16.9	19.1	20.1	20.3	20.4	20.9	25	25	22.8	21.2	21.8	22.6	23.3
		G	15	23.1	26.2	23.1	21.8	23.8	22.5	24.3	22	20.6	24.2	20.3	18.9	22.5	19.1
	3'	A	15.4	35.5	25.7	30.9	28	28.5	27.2	32.9	31.4	29.9	30.2	28.5	28	25.2	26.9
		T	63	31.1	33.4	26.5	27.7	31.3	29.9	24.5	27.5	27.7	28.4	24	27.9	25.2	29.1
		C	8.6	0.1	25.2	26.7	25.2	21.6	22.8	21.8	20.9	21.6	23.6	27	25	25.3	24
		G	13	23.5	15.7	15.9	19.1	18.6	20.1	20.8	20.1	20.8	17.7	20.4	19.1	24.3	20.1
3	5'	A	23.6	22.9	26.1	31	27.5	24.9	27.3	25.3	29.6	25.7	26.4	30.3	27.5	26.3	29
		T	43.6	33.9	29.4	27.5	25.7	29.8	29.4	24.7	30	29.6	27.5	24.9	26.9	30.6	27.3
		C	12.5	16.8	18.1	20.5	23.3	19.3	20.1	24.3	19.6	18.8	21.2	21.7	24.9	20.6	19.2
		G	20.4	26.4	26.4	21	23.5	26	23.3	25.7	20.8	25.9	24.9	23.1	20.6	22.5	24.5
	3'	A	24.5	34.7	21.2	29.6	24.8	28.6	28.7	30.4	32.3	24.8	26.5	27.9	28.7	32.8	28.2
		T	63	31.6	38.5	28.6	32.7	28.4	27.3	26.4	26	23.9	27.6	27.5	25.5	21.4	25.9
		C	4.2	5.9	23.2	25.9	22.9	21.8	20.6	20.4	21	26.7	24.1	23.6	25.1	20.6	21.2
		G	8.3	27.7	17.2	16	19.6	21.2	23.3	22.8	20.6	24.7	21.7	21	20.8	25.1	24.8

6	5'	A	37.5	18.8	18.8	25	50	12.5	18.8	31.3	31.3	31.3	31.3	18.8	25	31.3	43.8
		T	37.5	50	43.8	31.3	18.8	31.3	18.8	25	43.8	31.3	37.5	37.5	12.5	12.5	18.8
		C	6.3	6.3	0	12.5	25	18.8	31.3	25	18.8	12.5	18.8	18.8	37.5	43.8	6.3
		G	18.8	25	37.5	31.3	6.3	37.5	31.3	18.8	6.3	25	12.5	25	25	12.5	31.3
	3'	A	12.5	43.8	12.5	25	18.8	37.5	25	25	18.8	18.8	18.8	18.8	18.8	25	31.3
		T	75	37.5	31.3	37.5	37.5	6.3	31.3	43.8	37.5	31.3	6.3	37.5	37.5	18.8	12.5
		C	0	12.5	25	12.5	25	31.3	18.8	6.3	25	37.5	50	25	18.8	31.3	37.5
		G	12.5	6.3	31.3	25	18.8	25	25	25	18.8	12.5	25	18.8	25	25	18.8



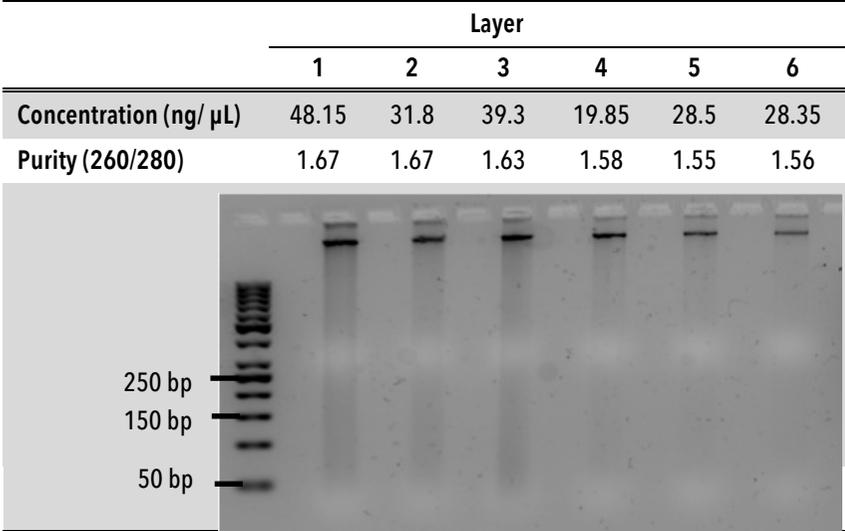
**FIGURE S3.8.2** | BASE COMPOSITION OF 15 BASES FROM THE 5' END AND THE 3' END OF UNIQUE MARSUPIAL SHOTGUN READS (PERCENTAGE OF BASES) FOR LAYERS 1, 3, AND 6.

### **S3.8.17 QUALITY *and* QUANTITY *of* GENOMIC DNA *as* DETERMINED *by* SPECTROPHOTOMETRY *and* GEL ELECTROPHORESIS**

There is a general pattern of decreasing total genomic through time; Layer 1 contains more total genomic DNA (measured by the NanoDrop spectrophotometer) than Layer 3, which contains more than Layer 6 (Table S3.8.7). However, Layer 4 contains the least amount of total genomic DNA as measured by the NanoDrop spectrophotometer. Overall, the total amount of genomic DNA recovered from Layers 1-3 is similarly higher than the total amount of genomic DNA recovered from Layers 4-6 (Table S3.8.7). The purity of DNA from Layers 1-3, as measured by the 260/280 ratio (DNA:protein, where 2 indicates a pure sample), is similar (approximately 1.65), while DNA from Layers 4-6 is less pure, with a 260/280 ratio of approximately 1.56: the presence of certain proteins and organic inhibitors may degrade the DNA, as well as contaminate the little DNA that is left.

The agarose gel electrophoresis (Table S3.8.7) shows darker smears of similar intensity for Layers 1-3 and fainter smears of similar intensity for Layers 4-6. The agarose gel also indicates that the DNA from all layers is highly degraded, as would be expected for aDNA. The large, intense band at the top of the gel (long DNA fragments), likely indicating modern contamination, while the short, degraded fragments present in the 'smear' are most likely to be endogenous aDNA. These results do not provide any information about the quality and quantity of endogenous DNA. For this, we assessed the amplification of endogenous DNA through qPCR and subsequent sequencing (see S3.8.18 below).

**TABLE S3.8.7** | TOTAL DNA CONCENTRATION AND PURITY (260/280) AS DETERMINED BY A NANODROP SPECTROPHOTOMETER (*Thermo Scientific*). The gel image embedded shows the fragment length distributions and relative total DNA concentration of 5  $\mu$ L of genomic extracts from Layers 1, 3 and 6 run on a 3% agarose gel electrophoresis at 70V for 1 hr (lane 1 includes a 50 bp DNA ladder, *GeneRuler, Fermentas*).



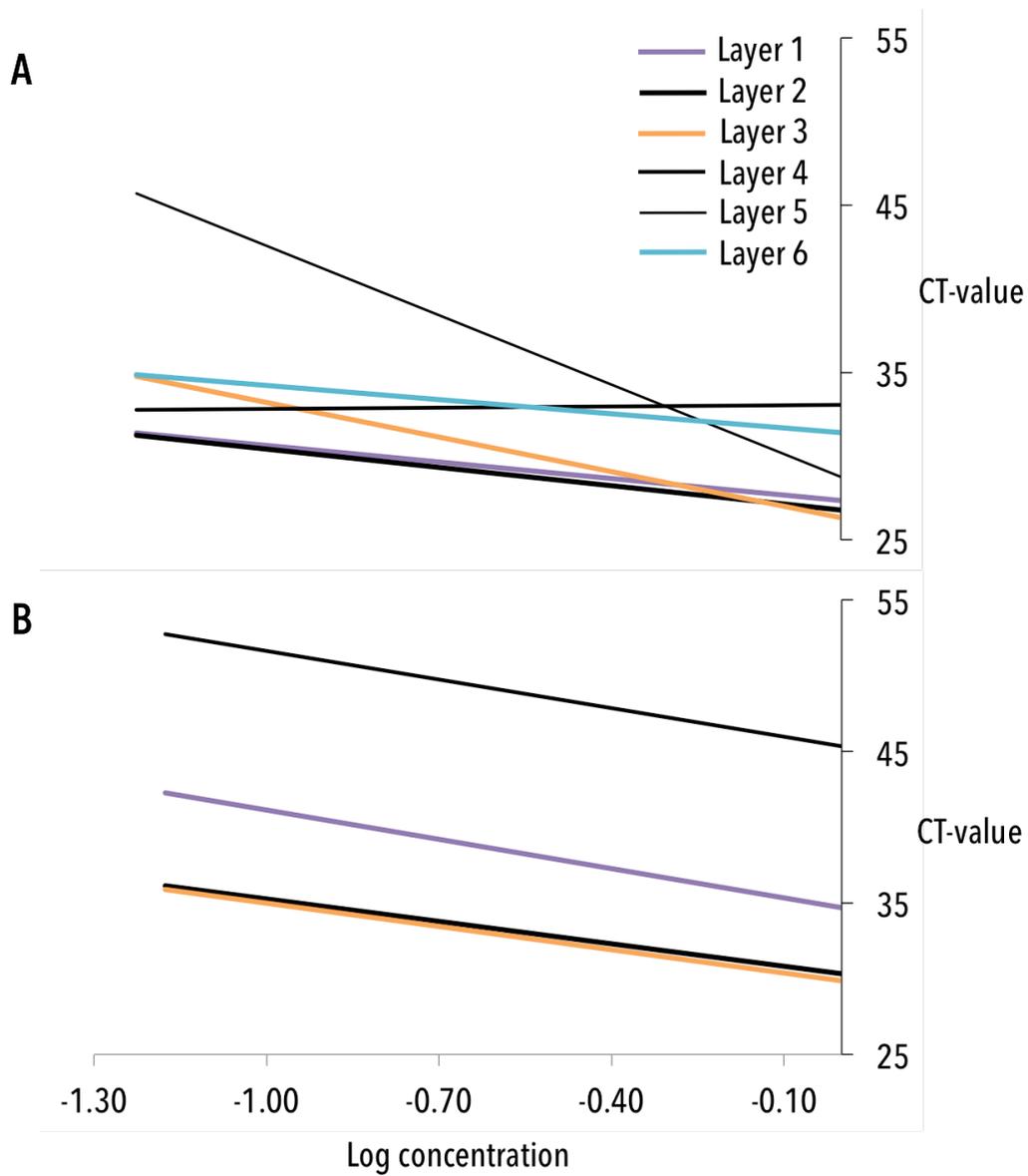
### S3.8.18 qPCR AMPLIFICATION RESULTS

**TABLE S3.8.8** | qPCR CYCLE-THRESHOLD ( $C_T$ ) VALUES FOR EACH REPLICATE OF EACH DILUTION PER LAYER, FOR BOTH 12SAO AND 12SMARSMini PRIMER SETS. Grey squares indicate no amplification.

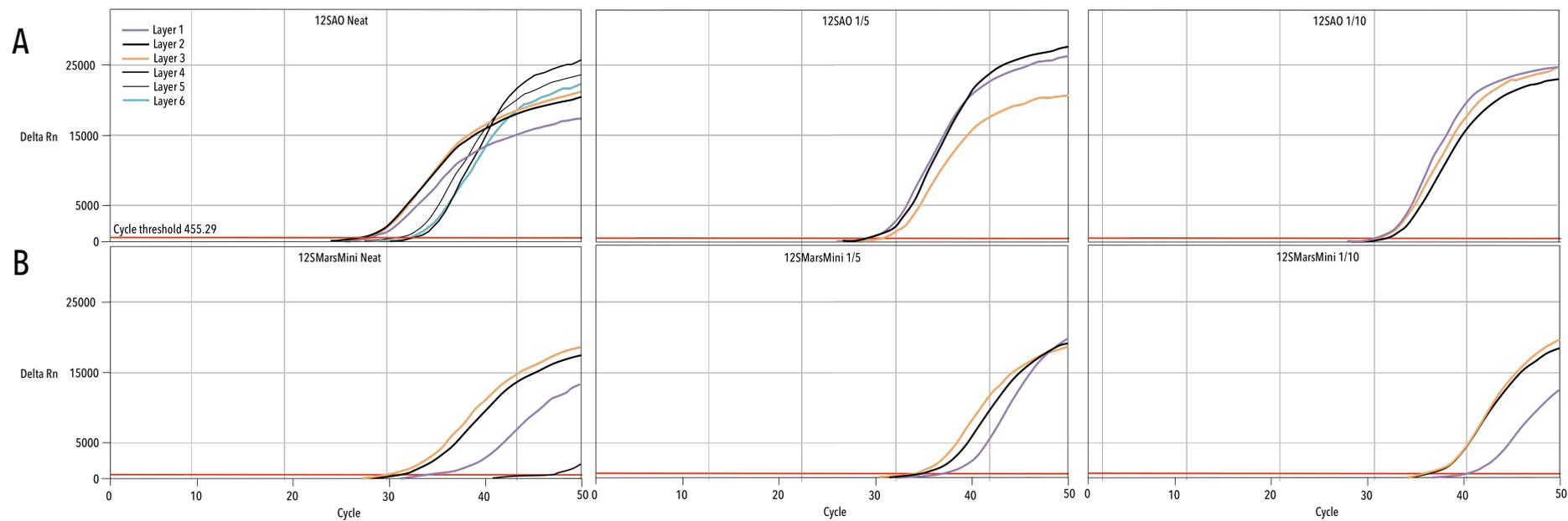
		12SAO						12SMarsMini					
Dilution	Log concentration	Layer	Layer 2	Layer 3	Layer	Layer	Layer	Layer 1	Layer 2	Layer	Layer 4	Layer 5	Layer 6
		1			4	5	3						
Neat	0.00	27.48	26.55	26.77	32.73	29.86	30.87	38.21	29.69	30.15	45.35		
		27.56	26.01	27.58	30.33	30.83	33.44	33.72	31.56	29.77			44.11
1in2	-0.30	28.45	28.25	27.85		32.37	30.71	37.91	31.56	31.27			
		28.24	28.02	27.77	38.13			34.23	31.35	31.44			
1in5	-0.70	29.65	29.56	30.81	31.18		32.17	38.54	33.63	33.14	49.75		
		29.29	30.81	31.50		32.35		36.46	33.46	33.43			
1in10	-1.00	30.26	31.75	32.48	32.45		35.74	40.7	35.64	35.67			
		30.79	29.81	30.78				39.3	35.75	34.42			
1in15	-1.18	32.67	30.59			46.41				36.05	35.25		
		30.90	30.56	38.52	32.78	49.06		46.51	35.96	36.66			

**TABLE S3.8.9** | THE PERCENT AMPLIFICATION EFFICIENCY OF THE QPCR REACTIONS FOR DNA EXTRACTS FROM LAYERS 1-6 AMPLIFIED WITH 12SAO AND 12SMARS MINI PRIMER SETS.  $R^2$  refer to the fit of the regressions fitted in Figure S3.8.3 below.

Primer set		Layer					
		1	2	3	4	5	6
12SAO	Slope	-3.44	-3.80	-7.18	0.25	-14.42	-2.95
	% Amplification efficiency	95.24	83.20	37.78	-99.99	17.32	118.27
	$R^2$	0.90	0.83	0.76	0.00	0.84	0.39
12SMarsMini	Slope	-6.45	-4.94	-5.13	-6.30	NA	NA
	% Amplification efficiency	42.92	59.37	56.71	44.16	NA	NA
	$R^2$	0.56	0.94	0.96	1.00	NA	NA



**FIGURE S3.8.3** | CYCLE THRESHOLD ( $C_T$ ) VALUES PLOTTED AGAINST LOG CONCENTRATION FOR EACH LAYER, for **A** 12SAO, and **B** 12SMarsMini primer sets.



**FIGURE S3.8.4** | QPCR AMPLIFICATION PLOTS of the neat (undiluted), 1/5, and 1/10 dilutions of DNA from Layers 1-6 (fluorescence per cycle) amplified with **A** 12SAO, and **B** 12SMarsMini primer sets (table I).

### **S3.8.19 FULL DATA TABLE**

The table below is available as an Excel spreadsheet with the online version of the Supplementary Information.

**TABLE S3.8.10** | FULL DATA TABLE SHOWING THE TAXA DETECTED FOR EACH SUBSAMPLE AND EACH EXTRACT REPLICATE WITHIN SUBSAMPLE (AS WELL AS EXTRACTION CONTROLS) FOR BOTH 12SAO AND 12SMARSMINI. Refer to the key on Table 3.3 for a definition of the symbols used. ! indicates a likely contaminant (non-native, no evidence of its presence at the time). Dark grey columns indicate that the sample did not amplify for either primer set. Medium grey columns indicate that the sample did not amplify for 12SMarsMini. Light grey columns indicate that the sample did not amplify for 12SAO. Numbers inside brackets indicate the number of reads assigned to the taxon.

### **S3.8.20 ERROR *and* TAXONOMIC RESOLUTION *of* THE 12SMARSMINI PRIMER SET**

The identity matrices presented in Table S3.8.11 exemplify that although the 12SMarsMini primer set amplifies only 41-43 bp, changes in 1-2 bp (that could be caused by DNA damage, amplification error, or sequencing error) would not change genus or family-level taxonomic identifications; at least 3 bp needs to differ between two samples to potentially change the genus-level identification, and more than 5 bp differences are needed to change the family-level identification. Thus, it is unlikely that error would result in genus or family misidentifications.

To further model the effect of damage and error on taxonomic assignments at the genus and family level for several genera, we applied a 1%, 2% and 3% mutation rate to a single sequence and generated 1000 sequences with randomly dispersed mutations, that were then BLAST to the GenBank database (Table S3.8.12). This was done using an open source python script (available: <https://www.biostars.org/p/12417/>). These results show that even with a 3% mutation rate (on average every sequence having at least one error; well above the combined error expected damage, amplification and sequencing), most sequences were not

assigned to a different genus or family; rather, erroneous sequences were pushed back to a higher taxonomic classification (e.g., species come genus, family becomes order) or were discarded because hits no longer aligned over 100% of the reference sequence. The one *Macropus* sequence generated from a sequence that was originally *Lagorchestes* (grey cell), occurred only at a mutation rate of 3%, and only resulted in 1/1000 sequences being able to be assigned to *Macropus*; this would normally be removed by abundance filtering.

**TABLE S3.8.11** | IDENTITY MATRICES OF SEQUENCES AMPLIFIED WITH 12SMARSMINI IN THREE EXAMPLE SAMPLES; each cell shows the percent similarity between two sequences and the number of base pairs that are different between them. Sequences are grouped by family and genus (different colours).

Example 1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	MB2343_12SMARSMINI_Phalangeridae-Trichosurus																					
2	MB2343_12SMARSMINI_Phalangeridae-Trichosurus	1/98%																				
3	MB2343_12SMARSMINI_Macropodidae	6/86%	6/86%																			
4	MB2343_12SMARSMINI_Macropodidae	7/83%	7/83%	3/93%																		
5	MB2343_12SMARSMINI_Macropodidae	6/86%	6/86%	3/93%	2/95%																	
6	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	7/83%	7/83%	3/93%	2/95%	2/95%																
7	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	8/81%	8/81%	4/90%	3/93%	3/93%	3/93%															
8	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	7/83%	7/83%	3/93%	2/95%	2/95%	2/95%	3/93%														
9	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	8/81%	8/81%	4/90%	3/93%	3/93%	3/93%	4/90%	3/93%													
10	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	8/81%	8/81%	4/90%	3/93%	3/93%	3/93%	4/90%	4/90%	4/90%												
11	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	8/81%	8/81%	4/90%	3/93%	3/93%	3/93%	4/90%	1/98%	2/95%	4/90%											
12	MB2343_12SMARSMINI_Macropodidae-Lagorchestes	7/83%	7/83%	3/93%	2/95%	2/95%	2/95%	3/93%	2/95%	1/98%	3/93%	3/93%										
13	MB2343_12SMARSMINI_Macropodidae-Macropus	7/83%	7/83%	1/98%	2/95%	4/90%	4/90%	5/88%	4/90%	5/88%	5/88%	5/88%	4/90%									
14	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	2/95%	3/93%	5/88%	5/88%	6/86%	5/88%	4/90%	6/86%	6/86%	3/93%	1/98%								
15	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	2/95%	3/93%	5/88%	5/88%	6/86%	5/88%	6/86%	6/86%	6/86%	5/88%	1/98%	2/95%							
16	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	2/95%	1/98%	3/93%	3/93%	4/90%	3/93%	4/90%	4/90%	4/90%	3/93%	1/98%	2/95%	2/95%						
17	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	2/95%	3/93%	5/88%	5/88%	6/86%	5/88%	6/86%	6/86%	6/86%	5/88%	1/98%	2/95%	2/95%	2/95%					
18	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	5/88%	4/90%	6/86%	4/90%	7/83%	6/86%	7/83%	7/83%	7/83%	6/86%	4/90%	5/88%	5/88%	5/88%	5/88%				
19	MB2343_12SMARSMINI_Macropodidae-Macropus	8/81%	8/81%	2/95%	3/93%	5/88%	5/88%	6/86%	5/88%	6/86%	6/86%	6/86%	5/88%	1/98%	2/95%	2/95%	2/95%	2/95%	5/88%			
20	MB2343_12SMARSMINI_Macropodidae-Macropus	6/86%	6/86%	2/95%	1/98%	3/93%	3/93%	4/90%	3/93%	4/90%	4/90%	4/90%	3/93%	1/98%	2/95%	2/95%	2/95%	2/95%	3/93%	2/95%		

Example 2		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	MB2417_12SMARSMINI_Burramyidae-Cercartetus															
2	MB2417_12SMARSMINI_Burramyidae-Cercartetus	2/95%														
3	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%													
4	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%												
5	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%	2/95%											
6	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%	2/95%	2/95%										
7	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%	2/95%	2/95%	2/95%									
8	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%	2/95%	2/95%	2/95%	2/95%								
9	MB2417_12SMARSMINI_Burramyidae-Cercartetus	1/98%	3/93%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%							
10	MB2417_12SMARSMINI_Macropodidae-Onychogalea	7/83%	7/83%	8/80%	8/80%	8/80%	8/80%	8/80%	8/80%	8/80%						
11	MB2417_12SMARSMINI_Macropodidae-Onychogalea	5/88%	7/83%	6/85%	6/85%	6/85%	6/85%	6/85%	6/85%	2/95%						
12	MB2417_12SMARSMINI_Pseudocheiridae-Pseudocheirus	9/78%	11/73%	10/76%	10/76%	10/76%	10/76%	10/76%	10/76%	9/78%	7/83%					
13	MB2417_12SMARSMINI_Pseudocheiridae-Pseudocheirus	9/78%	11/73%	10/76%	10/76%	10/76%	10/76%	10/76%	10/76%	9/78%	7/83%	1/98%				
14	MB2417_12SMARSMINI_Phalangeridae-Trichosurus	7/83%	9/78%	8/80%	8/80%	8/80%	8/80%	8/80%	8/80%	8/80%	6/85%	8/80%	8/80%			
15	MB2417_12SMARSMINI_Perameles	12/71%	14/67%	13/69%	13/69%	13/69%	13/69%	11/73%	13/69%	13/69%	14/67%	12/71%	13/69%	13/69%	11/73%	

Example 3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
1 MB2340_12SMARSMINI_Phalangeridae-Trichosurus																														
2 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%																													
3 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%																												
4 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%																											
5 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%																										
6 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%																									
7 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	2/95%	3/93%	3/93%	3/93%	3/93%	1/98%																								
8 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	3/93%	4/90%	4/90%	4/90%	4/90%	4/90%	5/88%																							
9 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%																						
10 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%																					
11 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%																				
12 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%																			
13 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%																		
14 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%																	
15 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	1/98%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%																
16 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%															
17 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%														
18 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%													
19 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%												
20 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%											
21 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%										
22 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%									
23 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%								
24 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%							
25 MB2340_12SMARSMINI_Phalangeridae-Trichosurus	1/98%	2/95%	2/95%	2/95%	2/95%	2/95%	3/93%	4/90%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%	2/95%						
26 MB2340_12SMARSMINI_Macropodidae-Macropus	9/79%	10/77%	9/79%	10/77%	10/77%	10/77%	9/79%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	8/81%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	10/77%	
27 MB2340_12SMARSMINI_Vombatidae-Vombatus	7/84%	8/81%	7/84%	8/81%	8/81%	8/81%	9/79%	10/77%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	9/79%	
28 MB2340_12SMARSMINI_Vombatidae-Vombatus	6/86%	7/84%	6/86%	7/84%	7/84%	7/84%	8/81%	9/79%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	7/84%	9/79%	1/98%	
29 MB2340_12SMARSMINI_Vombatidae-Vombatus	7/84%	8/81%	7/84%	8/81%	8/81%	8/81%	9/79%	10/77%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	8/81%	6/86%	8/81%	8/81%	8/81%	8/81%	8/81%	10/77%	2/95%	1/98%	

**TABLE S3.8.12** | TAXONOMIC ASSIGNMENTS OF 1000 SEQUENCES GENERATED FROM MUTATING ONE SEQUENCE OF EACH GENUS AT A RATE OF 1%, 2% AND 3%. Grey cells indicate taxonomic misidentifications at the genus level; however, these sequences would be discarded during filtering due to their low abundance.

Genus	Error rate		
	1%	2%	3%
<b>Lagorches</b>	Unassigned = 2	Unassigned = 11	Unassigned = 26
	Macropodidae = 38	Macropodidae = 72	Macropodidae = 98
	Lagorches = 959	Lagorches = 917	Macropus = 1
			Lagorches = 874
<b>Macropus</b>	Macropodidae = 14	Unassigned = 17	Unassigned = 46
	Macropus = 986	Macropodidae = 17	Macropodidae = 24
		Macropus = 966	Macropus = 930
<b>Cercartetus</b>	Cercartetus = 1000	Unassigned = 5	Unassigned = 13
		Cercartetus = 995	Cercartetus = 987
<b>Pseudocheirus</b>	Unassigned = 6	Unassigned = 7	Unassigned = 22
	Pseudocheirus = 994	Pseudocheiridae = 1	Pseudocheirus = 978
		Pseudocheirus = 991	
<b>Onychogalea</b>	Unassigned = 11	Unassigned = 34	Unassigned = 118
	Macropodidae = 46	Macropodidae = 966	Macropodidae = 884
	Onychogalea = 943		

<b>Trichosurus</b>	Unassigned =4	Unassigned =14	Unassigned = 36
	Phalangeridae =5	Phalangeridae = 10	Phalangeridae =22
	Trichosurus =991	Trichosurus = 976	Trichosurus =942

---

### S3.8.21 SUPPLEMENTARY REFERENCES

- Cooper A. 1994. Ancient DNA-sequences reveal unsuspected phylogenetic relationships with New-Zealand wrens (Acanthisttidae). *Experientia* **50**: 558-563.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15758-15763.
- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage—Building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Ramsey BC. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**: 337-360.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.

### 3.9 EPILOGUE

---

In this chapter, we have demonstrated how bulk bone can be used to evaluate general aDNA preservation at historical sites located within climates in which DNA is typically poorly preserved. The retrieval of aDNA from one such site in South Australia, Naracoorte, had not been achieved before now. The recent advancements made in next-generation sequencing technology as applied to aDNA extracted from bulk fossil bone (Chapter 2), has allowed us to acquire new information regarding past biodiversity in South Australia, as we describe several vertebrate families not previously recorded in the fossil record using traditional morphological approaches. A solid understanding of biodiversity loss in the area is key to developing informed conservation strategies that will better preserve the remaining biodiversity; however, it is vital that reference sequences continue to be added to genetic databases in order to obtain better taxonomic assignments (although this was beyond the scope of the study).

For the first time, a shotgun next-generation sequencing approach was applied to bulk bone aDNA extracts in order to characterise levels of exogenous contamination and DNA degradation that was then used to predict the limits of aDNA survival at Naracoorte. This is particularly important in warm climates because this method allows us to prioritise the best-preserved fossil deposits for aDNA extraction, without risking the destruction of precious specimens that, in all probability, will not yield informative aDNA. For example, in an additional experiment, we attempted to extract aDNA from bulk bone material that came from the same layer as a fossil specimen of the giant extinct snake *Wonambi* (Kelly Hill Cave, Kangaroo Island); this bulk bone did not yield amplifiable DNA, and consequently, a decision was made to not extract aDNA from rare fossil specimen, which may have damaged it fruitlessly.

Nevertheless, with the right molecular tool kit, aDNA can be recovered from extremely old fossil bone from environments where DNA survival was never thought possible. As the caves are of significant economic, educational, and environmental importance, this study advocates for further aDNA research to be carried out at

Naracoorte, and adds yet another reason to continue to protect these unique and treasured caves.

Chapters 2 and 3 have explored the use of bulk bone as a complementary tool to the morphological study of fossil bones, particularly in Australia. However, many sites, especially archaeological middens, do not yield any morphologically identifiable bone due to poor preservation conditions and food processing such as cooking. Can bulk bone be used to identify fossil assemblages where no morphological record exists at all? Chapter 4 investigates this possibility.

— CHAPTER 4 —

---

**TROPICAL ANCIENT DNA FROM BULK ARCHAEOLOGICAL FISH  
BONE REVEALS THE SUBSISTENCE PRACTICES OF A HISTORIC  
COASTAL COMMUNITY *in* SOUTHWEST MADAGASCAR**

---

*Invisibility—there are things we can't see now, that are there, that are embedded, that it really takes time order to be able to see. There are many ghosts that are lurking around and lingering through us that takes the technology of another generation or so in order to uncover...*

- Lynn Hershman Leeson

## 4.1 PROLOGUE

---

Chapter 3 showed that DNA preservation in fossil bone is often highly degraded in warm climates, and as such, aDNA can be difficult to retrieve using traditional substrates such as single-source fossil bone. The last chapter demonstrated that bulk bone can be used as an alternative to retrieve highly degraded aDNA from fossils deposited in warm climates.

In this chapter, we demonstrate how ancient DNA analysis of archaeological fish bone fragments from a tropical site can be a valuable tool for resolving faunal assemblages when morphological identification is hindered by taphonomic processes, as is often the case in tropical, coastal midden sites. The identification of archaeological fish assemblages is vital for understanding how ancient coastal peoples derived livelihoods from exploitation of the marine environment, and how that environment has subsequently been impacted.

Using the bulk bone metabarcoding method, we provide the first description of an archaeological fish assemblage from southwest Madagascar, and one of the few ancient DNA studies of Sub-Saharan archaeological material, entirely using molecular techniques. We compare the taxa identified to modern fisheries data in order to determine how and why subsistence practices of local people have changed over the past several hundred years. This is particularly important to establish as accelerated rates of environmental degradation, resource over-exploitation and loss of faunal diversity in recent times have generated important concerns about the future of Madagascar's natural communities and the ability of human communities to derive sustainable livelihoods from the marine environment. This study highlights how historic biodiversity information obtained from archaeological ancient DNA may inform future conservation and management decisions. We can only be aware of what has been lost if we know that it existed in the first place.

The study presented in this chapter resulted in a manuscript published in a 2016 issue of *Journal of Archaeological Science* (Grealy et al. 2016b, Vol. 75, Pg. 82-88), a facsimile of which can be found in Appendix I. This chapter is a reproduction of the

aforementioned manuscript (formatting, including in-text referencing and headings, excepted).

#### **4.1.1 ACKNOWLEDGEMENTS**

The archaeological investigations carried out in the Velondriake Marine Protected Area, were made possible with funding from the National Science Foundation Graduate Research Fellowship Program, the P.E.O. Scholar Award, the Yale Institute of Biospheric Studies, the Yale MacMillan Center for International and Area Studies and the Yale Council on Archaeological Studies. Research permissions were granted by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Autorisation Numéro 128/13-MESupReS/SG/DGRP and by the Centre de Documentation et de Recherche sur l'Art et les Traditions Orales Malgaches (CEDRATOM), under the auspices of the Memorandum of Understanding between the University of Toliara, under the direction of Dr. Barthélémy Manjakahery, Director of the CEDRATOM, and Yale University, under the direction of Dr. Roderick McIntosh, Professor of Anthropology. Local permission to carry out archaeological research was granted by the Office du Maire, Commune de Befandefa and by the Chefs de Fokontany of Andavadoaka, Nosy Ve, Antsaragnangy, Lamboara, Ampasilava and Salary. Permits for the export of archaeological materials for the purposes of laboratory analysis were granted by the Secretariat Général of the Ministère de l'Artisanat de la Culture et des Patrimoines, Direction Régionale de la Culture et du Patrimoine Atsimo Andrefana, Visas de Sorties Numéro 09/06-MCP/SG/DRCP.AA; Numéro 05/14-MACP/SG/DRCP.AA; Numéro 08/14-MACP/SG/DRCP.AA in accordance with Avis Numéro 375, 02/02/1978. We acknowledge the support of the Australian Research Council Discovery Project grant DP160104473 (MB).

We give special thanks to the Morombe Archaeological Project (MAP) team, to the people of Andavadoake, Madagascar, and to Blue Ventures conservation for sharing data pertaining to modern Velondriake Fisheries. We would also like to thank James Taylor for his assistance with sample preparation and sequencing, and Joey DiBattista for offering insights into fish behavior. High-throughput BLAST searches were performed through the Pawsey Supercomputing Centre (Perth).

#### 4.1.2 AUTHOR CONTRIBUTIONS

KD organised and directed the archaeological excavation. AG and JH assisted with the collection of bulk bone material. AG conducted genetic analyses with assistance from MB. JH designed the primers used. CG provided modern fisheries data for comparison. CB rendered the line drawings and figure. AG and KD drafted the manuscript with contributions and edits from all co-authors.

#### 4.1.3 AUTHOR DECLARATIONS

All necessary permits were obtained for the described study, which complied with all relevant regulations.

The authors declare no competing interests.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work (e.g., journal articles) in which the copyright is held by another party (e.g. publisher, co-author). Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged. Reprinted from Publication *Journal of Archaeological Science*, Vol. 75, Grealy AC et al., Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar, Pg. 82-88, Copyright (2016), with permission from Elsevier.

All supplementary data related to this article can be found at in the file “Supplementary Information” published with the online version of the article at <http://dx.doi.org/10.1016/j.jas.2016.10.001>, as well as section 4.8.

Sequencing data is deposited on the online data repository Data Dryad, and is available at: [doi:10.5061/dryad.8d1p9](https://doi.org/10.5061/dryad.8d1p9).

**TROPICAL ANCIENT DNA FROM BULK ARCHAEOLOGICAL FISH BONE REVEALS THE  
SUBSISTENCE PRACTICES OF A HISTORIC COASTAL COMMUNITY IN SOUTHWEST  
MADAGASCAR**

—*in*—

***JOURNAL OF ARCHAEOLOGICAL SCIENCE* (2016) | VOL. 75 | PG. 82-88**

Alicia Grealy<sup>1\*</sup>, Kristina Douglass<sup>2</sup>, James Haile<sup>3</sup>, Chriselle Bruwer<sup>4</sup>, Charlotte Gough<sup>5</sup>, Michael Bunce<sup>1</sup>

<sup>1</sup>Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia.

<sup>2</sup>Department of Anthropology, National Museum of Natural History, Smithsonian Institution, USA

<sup>3</sup>PalaeoBARN, Research Laboratory for Archaeology and the History of Art, Oxford University, UK

<sup>4</sup>Department of Anthropology and Archaeology, University of South Africa, Pretoria, South Africa, and Arc Maps and Graphics, Pretoria, South Africa

<sup>5</sup>Blue Ventures Conservation, London N7 9DP, UK

\* Corresponding author. Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia. *E-mail address*: alicia.grealy@uqconnect.edu.au

**Key words:** ancient DNA, archaeology, biodiversity, bulk-bone, fish, Madagascar, metabarcoding, subsistence

## 4.2 ABSTRACT

---

Taxonomic identification of archaeological fish bones provides important insights into the subsistence practices of ancient coastal peoples. However, it can be difficult to execute robust morphological identification of fish bones from species-rich fossil assemblages, especially from post-cranial material with few distinguishing features. Fragmentation, weathering and burning further impede taxonomic identification, resulting in large numbers of unidentifiable bones from archaeological sites. This limitation can be somewhat mitigated by taking an ancient DNA (aDNA) bulk-bone metabarcoding (BBM) approach to faunal identification, where DNA from non-diagnostic bone fragments is extracted and sequenced in parallel. However, a large proportion of fishing communities (both past and present) live in tropical regions that have sub-optimal conditions for long-term aDNA preservation. To date, the BBM method has never been applied to fish bones before, or to fossils excavated from an exposed context within a tropical climate. Here, we demonstrate that morphologically indistinct bulk fish bone from the tropics can be identified by sequencing aDNA extracted from 100-300 ya archaeological midden material in southwest Madagascar. Despite the biases of the approach, we rapidly obtained family, genus, and species-level assemblage information, and used this to describe a subset of the ichthyofauna exploited by an 18<sup>th</sup> century fishing community. We identified 23 families of fish, including benthic, pelagic, and coral-dwelling fishes, suggesting a reliance on a variety of marine and brackish habitats. When possible, BBM should be used alongside osteological approaches to address the limitations of both; however, this study highlights how genetic methods can nevertheless be a valuable tool for helping resolve faunal assemblages when morphological identification is hindered by taphonomic processes, lack of adequate comparative collections, and time constraints, and can provide a temporal perspective on fish biodiversity in the context of accelerated exploitation of the marine environment.

### 4.3 INTRODUCTION

---

THE identification of archaeological fish bone offers important insights into the subsistence practices of ancient fishing communities. Fish are a staple food for coastal peoples throughout the world; modern estimates suggest that approximately 60% of the global population lives within 100 km of the coast (Erlandson and Rick 2008). As such, there is tremendous potential for archaeological data and interpretations to provide a long-term perspective that can inform present-day marine resource management and conservation policies (Braje 2010, Braje et al. 2015, Lambrides and Weisler 2016, Speller et al. 2012). Fine-grained archaeological investigations of resource exploitation patterns are especially important because human arrival in many regions of the world has been correlated with an increase in faunal extinctions, implying that over-exploitation of local fauna has contributed to significant loss of biodiversity (Braje and Erlandson 2013). The coincidence of human colonisation and declines in floral and faunal diversity is particularly acute in island contexts (Rick et al. 2013). One example is Madagascar, which—although still considered a biodiversity ‘hotspot’ (Myers et al. 2000)—has undergone a significant loss of biodiversity over the last two thousand years that has long been thought to coincide with human arrival on the island (de la Bâthie 1921, Humbert 1927). However, many questions remain as to the timing of environmental change in Madagascar and the role early communities played in shaping the island’s land and seascapes, particularly given the challenges of investigating early forager sites (Douglass and Zinke 2015). Furthermore, disentangling anthropogenic and climatic drivers of environmental change remains a central research concern in Madagascar, since the island’s climate and environment were in constant flux well before human colonisation (Dewar and Richard 2007, Douglass and Zinke 2015). Moreover, despite the fact that Madagascar is an island, the historical ecology of Madagascar’s marine and coastal environment has received little research attention. Instead, as is the case in other parts of the world (Erlandson and Rick 2008), far more archaeological and paleontological work has been directed at understanding anthropogenic impacts on terrestrial ecologies.

Determining how humans impacted the marine environment of Madagascar during the Holocene relies on a thorough understanding of the marine taxa that were

targeted by ancient communities. Burnt or modified fish bone, or fish bones found in cultural deposits, are good gauges of direct human interaction with marine biota. As such, the identification of archaeological fish bone is essential to uncovering marine prey targets; to date, few studies comprehensively achieve this (Labridés and Weisler 2016), largely because of the limitations to morphological identification. The identification of fish bone predominantly relies on the examination of size range and diagnostic osteological features (Labridés and Weisler 2015). Cranial elements, such as teeth, are particularly important in refining taxonomic identification. However, the cranium and teeth generally make up a small proportion of the overall number of bones recovered from archaeological deposits (Yang et al. 2004); for each cranium, there may be over three times as many post-cranial elements (Jones 2009), including vertebrae, ribs, spines, and rays. Articulated specimens are even rarer because fish remains are fragile and susceptible to damage during food preparation, cooking, and consumption, as well as to post-depositional weathering (Collins 2010). Vertebrae and ribs of many fish species are often difficult, if not impossible, to distinguish from one another as they display little variation between species (Teletchea 2009). To complicate matters further, many fish exhibit different morphology throughout their development from juvenile to adult, and can also display high intraspecific morphological variability, sexual dimorphism (Teletchea 2009), and phenotypic plasticity (Labridés and Weisler 2015). Depositional bias, taphonomy, and lack of diagnostic features hinder morphological taxonomic identifications in many archaeological assemblages of fish bones, and often result in large numbers of unidentified remains.

Ancient DNA (aDNA) is a complementary method to the study of faunal remains, as it does not rely on the preservation of diagnostic morphological features. However, in Sub-Saharan African contexts, studies of archaeological aDNA are rare, despite the potential for aDNA analyses to complement traditional approaches to questions of human-environment interaction (Campana et al. 2013, Gifford-Gonzalez 2013). For fish, DNA reference collections represent a large portion of fish diversity, and DNA analysis has been used to discriminate cryptic species and morphotypes; for example, the genus *Schindleria* consists of 21 genetically distinct but morphologically cryptic species (Kon et al. 2007), while the morphologically different *Eumicrotremus spinosus* and *E. eggvinii* constitute a single species

(Byrkjedal et al. 2007). Ancient DNA has also proved to be a useful tool in studies of archaeological fish assemblages (Campana et al. 2013), albeit in a relatively small number of studies (Teletchea 2009): in a literature search, only approximately 2.5% of articles published on archaeological aDNA relate to fish. However, the studies that have been published demonstrate the value of such an approach in garnering important information about species diversity and distribution in the past (Cannon and Yang 2006, Grier et al. 2013, Speller et al. 2005, 2013, Yang et al. 2004), and the economic importance of different fish taxa to ancient communities (Nikulina and Schmölcke 2015).

The infrequent use of aDNA techniques in the analysis of archaeological fish assemblages may be due to the fact that fish bones are often too small and numerous to warrant the high cost of individual DNA extraction and sequencing—especially in the tropics that have one of the highest biodiversities of fish in the world (Lambrides and Weisler 2015). Furthermore, the majority of fishing communities occupy ‘exposed’ sites in coastal tropical and sub-tropical zones, with a consistently hot climate that fluctuates annually between dry and humid. These landscapes often lack natural and permanent shelter formations (such as caves), and are not typically conducive to aDNA preservation. Nevertheless, aDNA has been retrieved from tropical zones before (e.g., Gutiérrez-García et al. 2014, Murray et al. 2012, Nicholls et al. 2003, Schroeder et al. 2015), including Madagascar (Kistler et al. 2014, Mitchell et al. 2014, Orlando et al. 2008), and the innovation of new methods promises to increase the successful application of aDNA analysis on materials collected in tropical localities.

The recently developed ‘bulk bone metabarcoding’ (BBM) approach is one such method that allows the DNA from many bones to be extracted, amplified, and sequenced in parallel to rapidly and accurately identify many of the taxa within a sub-fossil assemblage (Grealy et al. 2015, 2016, Haouchar et al. 2014, Murray et al. 2013), which can increase the probability of characterising tropical archeofish remains. Here, we demonstrate how the BBM method can retrieve molecular taxonomic information from Malagasy 100-300 ya archaeological fish bone fragments that can then be used to examine past interactions of humans with their marine environment.

## 4.4 MATERIALS *and* METHODS

---

### 4.4.1 SITE DESCRIPTION *and* DATING

The coastal ‘Andamoty-be’ archaeological site is located just north of the village of Andavadoaka (22° 04’S, 43° 14’E) in Toliara province, Southwest Madagascar (Figure 4.1a,b), and was excavated in June 2014 by K. Douglass. The site is bordered on the east by spiny forest and by the Mozambique Channel to the west. It is located within the Velondriake Marine Protected Area, a locally managed marine area (LMMA) that encompasses the longest continuous reef system in Madagascar and is protected under the International Union for Conservation of Nature (IUCN). A decade of reef and fisheries monitoring by Blue Ventures Conservation has generated an excellent modern record of marine biodiversity within Velondriake’s shallow reef flats, sand flats, macro-algae, sea-grass and mangrove habitats (Cripps 2009, Cripps et al. 2015, Hantanirina and Benbow 2013, Harris et al. 2010, Jones et al. 2014, Nadon et al. 2007, Roy et al. 2009). Human occupation at the site is estimated to date between 100 and 300 ya based on the presence of imported 19<sup>th</sup> century British stoneware ceramics and Venetian glass trade beads found in the accumulation. The site has been described in detail elsewhere (Douglass 2016).

Two replicate 2 m x 2 m units were placed on areas with the highest density of surface scatter approximately 20 m apart and 500 m from the shoreline. Layers were excavated following the natural stratigraphy, resulting in four layers per unit with multiple sub-contexts within layers (Figure 4.1d). Excavated material was sieved on-site using 2 mm x 2 mm mesh screens. Bones were subsequently sorted from cultural material and stored at room temperature. Gloves and facemasks were worn during excavation, sieving, and sorting to minimise contamination with modern DNA. Gloves did not come into contact with modern fish at any time during the excavation to ensure that no contamination by modern fish DNA was introduced to the samples during collection.

#### 4.4.2 SAMPLE PREPARATION

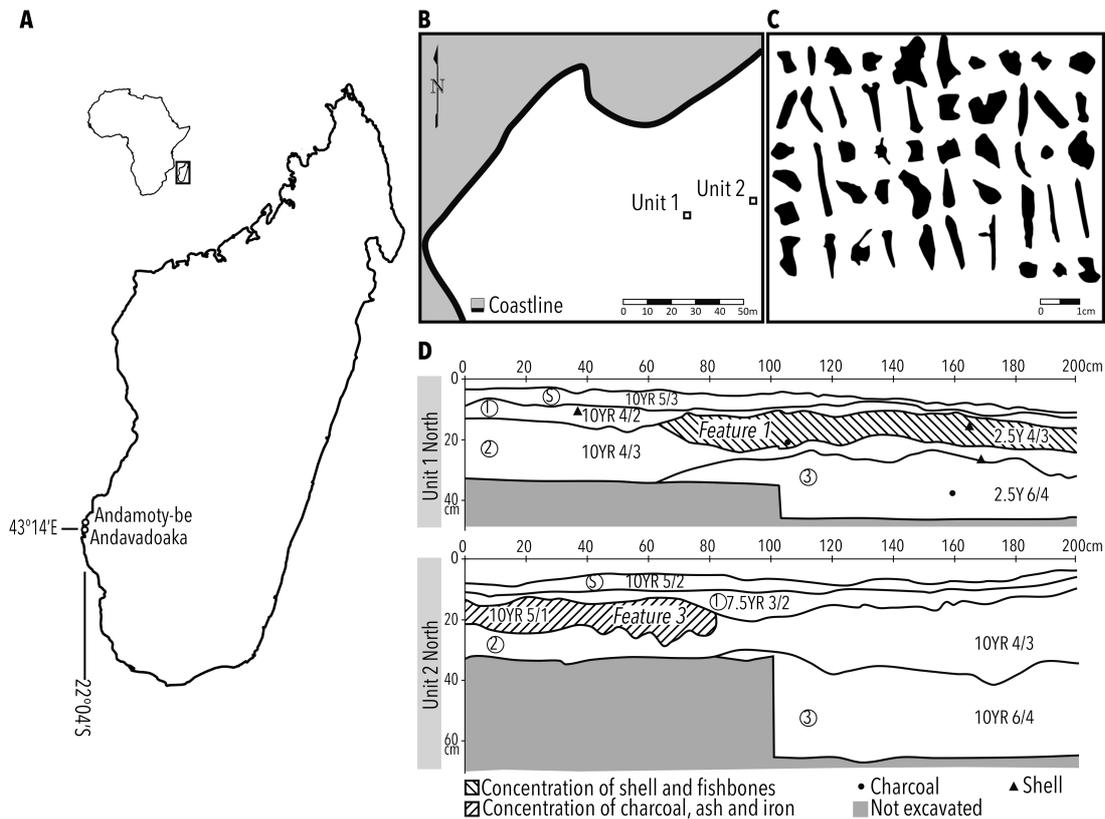
Where possible, one pool of 50 bones (Figure 4.1c) was randomly generated for each context (any layer or feature within a layer) for a total of 13 pools (note that two contexts contained fewer than 50 bones; for these, all bones were pooled). Three additional pools of 50 bones were generated for the first layer in each unit (six total) that targeted fish vertebrae fragments. A total of 887 bones were sampled, with each bone having an average mass of 123.5 mg. Approximately 20 mg of bone was subsampled from each bone within a pool and these were ground into a fine powder using the *Retsch* PM200 Planetary Ball Mill at 500 rpm for five min. Powder was stored at -20°C. All sample preparation was conducted in an isolated ultra-clean environment within Curtin University's TRACE facility (WA, Australia) following standard aDNA protocols for contamination avoidance (Willerslev and Cooper 2005; Knapp et al. 2012).

#### 4.4.3 ADNA EXTRACTION

aDNA was extracted from 100 mg of bone powder for each pool, following the methods described by Grealy et al. (2016). DNA-free controls were included throughout the extraction processes and were carried through to sequencing.

#### 4.4.4 METABARCODING *and* NEXT-GENERATION SEQUENCING

Primers targeting typical barcoding genes *COI* and *Cytb* tend to amplify regions that are too long to capture degraded DNA fragments of ancient samples (Jordan et al. 2010). Therefore, aDNA extracts were amplified via qPCR using a primer set designed to target conserved regions of the fish *12S rRNA* mitochondrial gene. At 53°C, these primers (*12S* 5'-CGCCTATATACCRCCGTC-3' and 5'-CRCTACACCTCGACCTG-3', flanked by unique indexes and *Illumina* sequencing adapters) amplify a 56 bp variable barcoding region from local members within more than 60 fish families. *In silico* analysis of the primer-binding sites shows that they are conserved across modern taxa found in the area, and are not likely to be inherently more biased towards the detection of any one taxon over another (S4.8.1, Figure S4.8.1). In most cases, the metabarcoding region differs by five or more base pairs between taxa of interest (Table S4.8.1), and it is unlikely that the combined effect



**FIGURE 4.1** | MAP OF MADAGASCAR showing the location of **A** the archaeological site examined, and **B** placement of the excavation units; **C** Silhouettes of a representative pool of 50 bones as an example of the typical size and shape of bones from the archaeological accumulation; **D** North-wall profiles of the stratigraphy for each unit depicting the layers excavated (rendered by C. Bruwer; vertical axis represents depth). Note that these are examples of the stratigraphy and do not depict all contexts (for further detail refer to Douglass 2016).

DNA damage, amplification error, and sequencing error would result in taxonomic mis-identifications (S4.8.1, Table S4.8.2). Amplification, subsequent sequencing on the MiSeq platform, trimming, and quality control were performed as per Grealy et al. (2016). DNA sequences are available on the online data repository Data Dryad and can be accessed via the doi:10.5061/dryad.8d1p9.

#### **4.4.5 TAXONOMIC ASSIGNMENT**

Taxonomy was assigned to sequences by comparison with NCBI's GenBank (Benson et al. 2006) nucleotide reference database using BLASTn (default parameters; Altschul et al. 1990) implemented through the Pawsey Supercomputing Centre (WA, Australia), and examination in MetaGenome Analyser (MEGAN v. 4.70.4; Huson et al. 2007) as per Grealy et al. (2016). Identifications were based on the similarity of query and reference sequence across 100% of the query, with similarity cut-offs for species-level IDs at > 98% similarity, genus-level IDs at 96-97% similarity, and family-level IDs at 90-95% similarity. Assignments were assigned a credibility rating (highly credible, credible, or unlikely; Table 4.1) based on whether the taxa are found in the area according to species' distribution records defined by FishBase (Froese and Pauly 2015), and whether genetic reference sequences exist in GenBank for all subtaxa within family or genus.

#### **4.5 RESULTS *and* DISCUSSION**

---

Next-generation sequencing of 56 bp *12S rRNA* sequences amplified from bulk fish bone aDNA generated a total of 77,298 reads (an average of 4024 reads per pool) and 1338 unique reads (an average of 70 unique reads per pool). After molecular taxonomic identification by comparison with a reference database, 23 families were identified with high credibility; within them, 14 genera were able to be identified with high credibility, four credibly, and within those, five species were able to be identified with high credibility and six credibly (Table 4.1). No fish DNA was amplified in any of the controls, indicating that contamination from the laboratory environment was below detectable levels.

Despite the large diversity of taxa identified, habitat associations derived from modern fisheries data (Blue Ventures) suggest a primary reliance on near-shore reef-dwelling fish, with 53% of identified families associating with coral reefs, and the remainder associating with seagrass (8%), mixed habitats (11%), or unknown (28%). These data suggest that there was a strong dependence on coral communities in terms of the exploitation of marine taxa. It appears that a range of fish sizes were targeted, and although not enough sampling has been done thus far to draw comparisons between layers, species within the family Lethrinidae (followed by Scaridae and Serranidae) were detected in more samples than any other taxa. This suggests that these typically large-bodied, high trophic-level fish may have constituted the primary staple marine food of people at Andamoty-be; in contrast, these families represent a small percentage of the catch in near-by Morombe today (Laroche et al. 1997). At Morombe, “high fishing pressure [has] led to a concentration of effort on lower trophic level species to maintain catch levels” (Laroche et al. 1997), an example of “fishing down marine food webs” (Pauly et al. 1998). This may indicate that fishing pressure in this region has increased over several hundred years, and that line fishing (the predominant method employed for catching high trophic-level fishes) may have been more commonly practiced by ancient communities than it is at present; today, only about 6% of catch in the Morombe region is by line (Laroche et al. 1997).

Other fish families detected include reef dwelling fish of the Chaetodontidae and Pomacentridae families (a mix of corallivores, planktivores, omnivores and herbivores, some of which are small and may have been used as bait fish), carnivores such as the wrasses of the Labridae family, members of the Carangidae family, and members of the Sparidae family. *Megalops cyprinoides* (Indo-Pacific tarpon) of the Megalopidae family are typically migratory fish that move between open water and inland rivers (Merrick and Schmida 1984). In the modern fisheries data (Blue Ventures), Megalopidae were recorded from catch in the coral habitat, suggesting that although adult fish could have been caught in the open sea beyond the barrier reef, they may have been netted, as adolescent *M. cyprinoides* migrate offshore from estuarine waters and mangroves (Coates 1987). Pelagic fish, like members of the Chanidae family (*Chanos chanos*, or milkfish), also possibly indicate open sea

**TABLE 4.1** | TAXA IDENTIFIED THROUGH BULK BONE METABARCODING OF ADNA. Molecular taxonomic identifications retrieved from analysis of bulk archaeological fish bone from six layers across two excavation units at Andamoty-be, Southwest Madagascar through next-generation sequencing of a short metabarcoding region of the mitochondrial *12SrRNA* gene. Asterisks designate taxa not recorded in modern fisheries data. Grey cells indicate that the taxon was found in more than one sample (relevant to Layer 1 only). ^ Indicates that the taxon may be derived from modern contamination. Taxa are classified as being: ‡ highly credible (within cut-off %ID across > 100% of the query, found in area according to FishBase, with genetic reference sequences for all subtaxa within family or genus present in GenBank); † credible (with in cut-off %ID across > 100% of the query, but not all species of the genus or genus within the family that also occur in the area according to FishBase have reference sequences in GenBank); and Δ unlikely (within cut-off ID% across 100% of the query, but not found in the area according to FishBase).

Family (90-95% ID)	Genus (96-97% ID)	Species (98-100% ID)	Common name	Surface		Layer 1		Layer 2		Layer 2		Layer 2	Layer 3	
				Unit	1	2	1	2	1	2	Feature 1	Feature 2	Feature 3	1
Dasyatidae	<i>Himantura</i>	<i>gerrardi</i>	Whitespotted whipray				‡							
							‡							
Latidae*	<i>Psammoperca</i>	<i>waigiensis</i>	Waigeo barramundi				‡							
							‡							
Ginglymo- stomatidae*			Nurse sharks											‡
														
Carcharhinidae			Requiem sharks				‡		‡					
							‡		‡					



	Subfamily									
	Epinephelinae				‡			‡	‡	
Balistidae					‡		‡		‡	
Siganidae							‡			‡
	<i>Siganus</i>		Rabbit fishes		‡	‡	‡			‡
Cichlidae*			Cichlid		‡					
Sparidae			Sea bream							‡
Megalopidae							‡			
	<i>Megalops</i>						‡			
	<i>cyprinoides</i>		Indo-Pacific tarpon				‡			
Kyphosidae								‡	‡	
	<i>Kyphosus</i>							‡	‡	
	<i>bigibbus</i>		Brown chub					‡	‡	
Pomacentridae								‡		
	<i>Abudefduf</i>		Sergeant-majors					‡		
Chanidae										‡
	<i>Chanos</i>									‡
	<i>chanos</i>		Milk fish							‡
Clupeidae*^					†					
	<i>Sardina</i> ^				Δ					
	<i>pilchardus</i> ^		European pilchard		Δ					
Haemulidae									‡	‡
	<i>Haemulon</i>								Δ	Δ
	<i>aurolineatum</i>		Tomato grunt						Δ	Δ

fishing; these fish live in large schools in surface waters over the continental shelf and generally require sophisticated fishing methods, including nets, larger outrigger canoes than are needed for fishing around coral reefs, and potentially the cooperation of several boats (Wheeler and Jones 1989). However, like members of the Megalopidae family, milkfish do migrate into brackish waters (including mangroves, estuaries, and lakes) as juveniles and return to the sea to sexually mature (Froese and Pauly 2015). Requiem sharks (*Carcharhinus*) are also known to occur in brackish and freshwater habitats. The presence of demersal fish such as whiprays (*Himantura*) may indicate the practice of bottom trawling or line fishing. In addition to serving as a food source, stingrays are commonly sought out in Velondriake today as a valuable source of abrasive material, and their tails are used as a tool to shape and sand wood (Douglass 2012).

Several taxa identified have not been recorded in modern fisheries data (Blue Ventures). These include carnivores such as *Psammoperca waigiensis* (Waigeo barramundi) of the Latidae family and Ginglymostomatidae (nurse sharks), freshwater fishes of the Cichlidae family, and forage-fishes of the Clupeidae family. The detection of nocturnal predators such as nurse shark and Waigeo barramundi may be an indication of night fishing, dive fishing, or leaving nets out overnight. In particular, nurse sharks are bottom feeders that live in shallow inshore waters with coral communities. Overfishing may be responsible for the rarity of these sharks today (Cooke 1997), indicating that there has been significant anthropogenic impact on the environment by past people. Furthermore, the “season of abundance” for sharks is predominantly April to July (Langley 2006), which may indicate that this site was inhabited during the cooler, dry season.

Detection of the Cichlidae family is interesting as this is the only non-marine family identified in the archaeological assemblage thus far, and no cichlid catch is recorded in the modern fisheries data (Blue Ventures). Cichlids are a diverse family, with 28 endemic and nine introduced species known from Madagascar (Froese and Pauly 2015). This identification is strongly suggestive of fresh and/or brackish water fishing by local people. The closest occurrence of cichlid species to Andamoty-be is the Onilahy river basin’s *Ptychochromoides betsileanus* and *Ptychochromis onilahy*: these species are classified by the IUCN as critically endangered and extinct

(respectively) as a result of habitat loss, fishing, and competition or predation by introduced species. The Onilahy River is located approximately 180 km south of Andamoty-be, so the presence of a cichlid at Andamoty-be could be an indication that the ranges of one or both of these species extended as far as Andamoty-be in the past, but underwent a range contraction as a result of human pressures. With more research, a historic range could potentially be established for these species, which may inform conservation efforts (Hofman et al. 2015, Speller et al. 2012). To confirm the identification of cichlids, additional samples and metabarcoding genes should be sequenced.

Finally, the detection of Clupeidae DNA (100% sequence similarity to *Sardina pilchardus*) in only the surface scatter layer is likely to have been derived from contamination by imported sardines: they are not native species but are a common component of human diet in Madagascar today. Although other sardine genera have been recorded in the modern fisheries data, these are generally well represented in genetic databases and their sequences differ from *Sardina pilchardus* by more than 9%, making it unlikely that this DNA originated from native sardine species.

Like osteological approaches, not all taxa can be identified to the species-level, and some taxa are likely to have not been detected at all (cf., Grealy et al. 2015 for an in-depth discussion of the biases and limitations of the method where a direct comparison with a morphological approach was possible): Table 4.1 shows that some taxa are not consistently detected between replicates, highlighting that the ability to detect a given taxon in a complex mixture can be variable. Similar to other metabarcoding approaches such as bacterial metagenomics or environmental DNA monitoring, this ability to identify taxa genetically is influenced by: (1) the intensity of sampling, (2) unique taphonomic biases that affect DNA preservation (including pH, temperature fluctuations, and exposure to humidity), (3) collection and storage (such as handling that can introduce contamination), (4) choice of barcoding region (high intraspecific variation at a locus can affect taxonomic resolution), (5) incomplete reference genetic databases (while comprehensive, some taxa may not be represented), and (6) DNA damage, PCR bias, and sequencing error (although some of these can be mitigated by adequate sequencing depth, PCR replication, diluting inhibitors, and stringent quality control). Nevertheless, the detection of one taxon is

not undermined by an inability to detect another, although we cannot confidently estimate what we did *not* find. In addition, although the amplification of longer fragments may have resulted in more refined taxonomic identifications, the degradation of aDNA in tropical environments typically results in the majority of fragments being very short; as such, there is a trade-off between the breadth of taxa identified and the specificity of identification (Grealy et al. 2015). Analysis of more samples, amplification of additional barcoding genes, and revisiting the existing data as genetic databases become more complete, will also deliver more fine-scaled molecular identifications and identify additional diversity. While not a complete audit of the past fish diversity, this is the first published description of an archaeological fish assemblage from southwest Madagascar, and demonstrates that a genetic approach provides useful zooarchaeological information in the absence of an alternative. The analysis of additional DNA extracts in the future will allow us to potentially compare archaeofish biodiversity between Andamoty-be and other archaeological sites in Madagascar.

#### 4.6 CONCLUSION

---

This study has established the first published marine zooarchaeological record for Velondriake, offering insights into how past coastal communities derived a livelihood from local marine resources. This is particularly important to establish as accelerated rates of environmental degradation, resource over-exploitation and loss of faunal diversity in recent times have generated important concerns about the future of Madagascar's natural communities and the ability of human communities to derive sustainable livelihoods, especially in a region where more than a third of the population currently engage in sea fishing (Laroche et al. 1997). The data presented here provide a baseline upon which future data collection and analysis may build, and knowledge of historic biodiversity and human exploitation of the marine environment may assist in conservation and management decisions. *Post-hoc* comparisons with morphological analysis of fossil assemblages in Velondriake may support the accuracy (or otherwise) of the molecular identifications. Nevertheless, this study suggests that other archaeological sites around Madagascar, and in other tropical regions, may benefit from aDNA analysis of bulk bone to expand the

taxonomic identifications obtained through traditional methods, and hopefully will encourage more fruitful collaborations between geneticists and archaeologists.

#### 4.7 REFERENCES

---

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2006. GenBank. *Nucleic Acids Research* **34**.

Braje TJ. 2010. Modern oceans, ancient sites: archaeology and marine conservation on San Miguel Island, California. Salt Lake City, University of Utah Press.

Braje TJ and Erlandson. 2013. Human acceleration of animal and plant extinctions: a Late Pleistocene, Holocene and Anthropocene continuum. *Anthropocene* **4**: 14-23.

Braje TJ, Rick TC, Erlandson JM, Rogers-Bennett L, Catton CA. 2015. Historical ecology can inform restoration site selection: the case of black abalone (*Haliotis cracherodii*) along California's Channel Islands. *Aquatic Conservation: Marine and Freshwater Ecosystems* **26**:470-481

Byrkjedal I, Rees DJ, Willassen E. 2007. Lumping lumpsuckers: molecular and morphological insights into the taxonomic status of *Eumicrotremus spinosus* (Fabricius, 1776) and *Eumicrotremus eggvinii* (Koefoed, 1956) (Teleostei: Cyclopteridae). *Journal of Fish Biology* **71**: 111-131.

Campana MG, Bower MA, Crabtree PJ. 2013. Ancient DNA for the archaeologist: the future of African research. *African Archaeological Review* **30**: 21-37.

- Cannon A and Yang DY. 2006. Early storage and sedentism on the Pacific Northwest Coast: ancient DNA analysis of salmon remains from Namu, British Columbia. *American Antiquity* **71**: 123-140.
- Coates D. 1987. Observations on the biology of tarpon, *Megalops cyprinoides* (Broussonet) (Pisces: Megalopidae), in the Sepik River, Northern Papua New Guinea. *Australian Journal of Marine and Freshwater Research* **38**: 529-535.
- Collins BR. 2010. Element survivability of *Salmo salar*. *Journal of Taphonomy* **8**: 291-300.
- Cooke AJ. 1997. Survey of Elasmobranch fisheries and trade in Madagascar in The trade shark and shark products in the western Indian and southeast Atlantic oceans, Marshall NT and Barnett R (eds.). Graphicor: South Africa.
- Cripps G. 2009. Understanding migration amongst the traditional fishers of West Madagascar. Blue Ventures Conservation Report for ReCoMaP.
- Cripps G, Harris A, Humber F, Harding S, Thomas T. 2015. A preliminary value chain analysis of shark fisheries in Madagascar. Indian Ocean Commission SF/2-15/34.
- de la Bâthie HP. 1921. La végétation malgache. *Annales du Musée Colonial de Marseille* **9**: 1-266.
- Dewar RE and Richard AF. 2007. Evolution in the hypervariable environment of Madagascar. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 13723-13727.
- Douglass S. 2012. Fisher's of a song: community music education in Andavadoaka, Madagascar. A thesis presented to the Faculty of the Steinhardt School of Culture, Education and Human Development at New York University in Candidacy for the Degree of Master of Arts in Music Education.

- Douglass K and Zinke J. 2015. Forging ahead by land and by sea: archaeology and paleoclimate reconstruction in Madagascar. *African Archaeological Review* **32**: 267-299.
- Douglass K. 2016. An archaeological investigation of settlement and resource exploitation patterns in the Velondriake Marine Protected Area, Southwest Madagascar, ca. 900 BC to AD 1900. A dissertation presented to the Faculty of the Graduate School of Yale University in Candidacy for the Degree of Doctor of Philosophy.
- Erlandson JM and Rick TC. 2008. Archaeology, marine ecology, and human impacts on marine environments. University of California Press; Berkeley.
- Froese R and Pauly D. 2015. FishBase. Available: [www.fishbase.org](http://www.fishbase.org). Accessed: Jul 2015.
- Gifford-Gonzalez D. 2013. Animal genetics and African archaeology: why it matters. *African Archaeological Review* **30**: 1-20.
- Grealy A, McDowell M, Scofield P, Murray D, Fusco D, et al. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional palaeontological methods. *Quaternary Science Reviews* **128**: 37-47.
- Grealy A, Macken A, Allentoft ME, Rawlence NJ, Reed E, et al. 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *Journal of Quaternary Science* **31**: 33-45.
- Grier C, Flanigan K, Winters M, Jordan LG, Lukowski S, et al. 2013. Using ancient DNA identification and osteometric measures of archaeological Pacific salmon vertebrae for reconstructing salmon fisheries and site seasonality at Dionisio Point, British Columbia. *Journal of Archaeological Science* **40**: 544-555.

- Gutiérrez-García TA, Vázquez-Domínguez E, Arroyo-Cabrales J, Kuch M, Enk J, et al. 2014. Ancient DNA and the tropics: a rodent's tale. *Biology Letters* **10**.
- Hantanirina JMO and Benbow S. 2013. Diversity and coverage of seagrass ecosystems in southwest Madagascar. *African Journal of Marine Science* **35**: 291-297.
- Haouchar D, Haile J, McDowell MC, Murray DC, White NE, et al. 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews* **84**: 56-64.
- Harris A, Manahira G, Sheppard A, Gough C, Sheppard C. 2010. Demise of Madagascar's once great barrier reef: change in coral reef condition over 40 years. *Atoll Research Bulletin* **574**:1-16.
- Hofman CA, Rick TC, Fleischer RC, Maldonado JE. 2015. Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology and Evolution* **30**: 540-549.
- Humbert H. 1927. La Destruction D'une Flore Insulaire Par Le Feu. Principaux Aspects de la Végétation À Madagascar. Documents Photographiques Et Notices. *Mémoires de l'Académie Malgache* **5**: 1-80.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377e386.
- Jones S. 2009. Food and gender in Fiji: ethnoarchaeological explorations. Lexington Books; USA.
- Jones T, Ratsimba H, Ravaorinorotsihoarana L, Cripps G, Bey A. 2014. Ecological variability and carbon stock estimates of mangrove ecosystems in Northwestern Madagascar. *Forests* **5**: 177-205.

- Jordan LG, Steele CA, Thorgaard GH. 2010. Universal mtDNA primers for species identification of degraded bony fish samples. *Molecular Ecology Resources* **10**: 225-228.
- Kistler L, Ratan A, Godfrey LR, Crowley BE, Hughes CE, et al. 2014. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *Journal of Human Evolution* **79**: 45-54.
- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage—building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Kon T, Yoshinto T, Mukai T. 2007. DNA sequences identify numerous cryptic species of vertebrate: a lesson from the gobioid fish *Schindleria*. *Molecular Phylogenetics and Evolution* **44**: 53-62.
- Lambrides ABJ and Weisler MI. 2015. Applications of vertebral morphometrics in Pacific Island archaeological fishing studies. *Archaeology in Oceania* **50**: 43-70.
- Lambrides ABJ and Weisler MI. 2016. Pacific Islands ichthyoarchaeology: implications for the development of prehistoric fishing studies and global sustainability. *Journal of Archaeological research*: **25**: 275.
- Langley J. 2006. Vezo knowledge: traditional ecological knowledge in Andavadoaka, southwest Madagascar. London: Blue Ventures.
- Laroche J, Razanoelisoa J, Fauroux E, Rabenevanana MW. 1997. The reef fisheries surrounding the southwest coastal cities of Madagascar. *Fisheries management and ecology* **4**: 285-299.
- Merrick JR and Schmida GE. 1984. Australian freshwater fishes—biology and management. Griffin Press: South Australia.

- Mitchell K, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Murray DC, Pearson SG, Fullagar R, Chase BM, Houston J, et al. 2012.. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews* **58**: 135-145.
- Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. 2013. Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific Reports* **3**: 3371.
- Myers N, Mittermeyer RA, Mittermeyer CG, da FONSECA GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858.
- Nadon MO, Griffiths D, Doherty E, Harris A. 2007. The status of coral reefs in the remote region of Andavadoaka, Southwest Madagascar. *Western Indian Ocean Journal of Marine Science* **6**: 207-218.
- Nicholls A, Matisoo-Smith E, Allen MS. 2003. A novel application of molecular techniques to Pacific archaeofish remains. *Archaeometry* **45**: 133-147.
- Nikulina EA and Schmölcke U. 2015. Archaeogenetic evidence for medieval occurrence of Atlantic sturgeon *Acipenser oxyrinchus* in North Sea. *Environmental Archaeology* **21**: 137-143.
- Orlando L, Calvignac S, Schnebelen C, Douady CJ, Godfrey LR, et al. 2008. DNA from extinct giant lemurs links archaeolemurids to extant indriids. *BMC Evolutionary Biology* **8**: 121.
- Pauly D, Christensen V, Dalsgaard J, Froese R, Torres FJ. 1998. Fishing down the marine food webs. *Science* **279**.

- Rick TC, Kirch PV, Erlandson JM, Fitzpatrick SM. 2013. Archaeology, deep history, and the human transformation of island ecosystems. *Anthropocene* **4**: 33-45.
- Roy R, Dunn S, Purkis S. 2009. Mapping Velondriake: the application of bathymetric and marine habitat mapping to support conservation planning, southwest Madagascar. Blue Ventures Conservation Report.
- Schroeder H, Ávila-Arcos M, Malaspinas A, Poznik GD, Sandoval-Velasco M, et al. 2015. Genome-wide ancestry of 17<sup>th</sup>-century enslaved Africans from the Caribbean. *Proceedings of the National Academy Sciences of the United States of America* **112**: 3669-3673.
- Shaw JLA, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, et al. 2016. Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation* **197**: 131-138.
- Speller CF, Yang DY, Hayden B. 2005. Ancient DNA investigation of prehistoric almon resource utilization at Keatley Creek, British Columbia, Canada. *Journal of Archaeological Science* **32**: 1378-1389.
- Speller CF, Hauser L, Lepofsky D, Moore J, Rodrigues AT, et al. 2012. High potential for using DNA from ancient herring bones to inform modern fisheries management and conservation. *PLOS One* **7**.
- Teletchea F. 2009. Molecular identification methods of fish species: reassessment and possible applications. *Reviews in Fish Biology and Fisheries* **19**:265-293.
- Wheeler A and Jones AKG. 1989. *Fishes*. Cambridge University Press: Cambridge.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.

Yang D, Cannon A, Saunders S. 2004. DNA species identification of archaeological salmon bone from the Pacific Northwest Coast of North America. *Journal of Archaeological Science* **31**: 619-631.

## 4.8 SUPPLEMENTARY INFORMATION

---

### S4.8.1 BIAS, ERROR *and* TAXONOMIC RESOLUTION *of* THE SHORT *12S* FISH PRIMER SET

The short *12S* fish primer set used here (see main text) were designed to detect and distinguish between over 100 modern fish taxa that are known from the region across more than 50 families. An alignment of 80 of these taxa performed in Geneious (Figure S4.8.1) shows that the primer-binding sites are conserved across these taxa, and are not likely to be inherently more biased towards the detection of any one taxon over another. Taxa with one or two base pairs that differ from the primer are expected to still be amplified by the primers because mismatches generally do not occur toward the 3'-end of the primer (the end where mismatches are more likely to affect primer binding efficiency), and the primers were annealed at a low temperature (53°C) that allows relaxed primer-template specificity. Primers were checked *in silico* for the formation of secondary structures (hairpins, self-dimers, and heterodimers) using *IDT's* OligoAnalyzer web tool (<https://sg.idtdna.com/calc/analyzer>). Primers were also compared with NCBI's Genbank nucleotide database using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) to ensure that they did not pick up non-specific targets (i.e., non-fish). Primers were experimentally optimised using qPCR for use with ancient DNA templates, alongside a single-source modern fish positive-control. PCR products were run on a 2% agarose gel electrophoresis to ensure that one-sized product only was amplified, ensuring that there was no amplification of non-specific products in practice. Samples were also screened for inhibition prior to fusion-tag PCR in order to reduce amplification bias caused by the presence of inhibitors. The use of multiple independent replicate PCR reactions per sample is also likely to reduce the effects of stochastic PCR bias.

Degenerate primers that amplify longer regions (i.e., 150 bp) of both fish *12S rRNA* and *16S rRNA* were also tested. While longer primer sets can provide greater taxonomic resolution, fewer taxa are recovered and fewer samples contain any amplifiable DNA of this length (data available upon request). The taxonomic and sample drop-out caused by attempting to amplify longer DNA templates is typical of highly degraded samples, and could be argued to bias taxonomic composition more so

FWD 5'-3' CGCCTATATACCRCC-GTC-INSERT-  
 REV 3'-5' -INSERT-TCAGGT-CGAGGTGTAGYG

Belonidae, <i>Ablennes hians</i>	AB373007	-----G-----INSERT-----C-
Sparidae, <i>Acanthopagrus australis</i>	KJ774772	-----A-----INSERT-----C-
Acanthuridae, <i>Acanthurus mata</i>	AY057237	-----A-----INSERT-----C-
Acanthuridae, <i>Acanthurus triostegus</i>	AY057246	-----A-----INSERT-----C-
Serranidae, <i>Aethaloperca rogae</i>	NC_022141	-----G-----INSERT-----C-
Myliobatidae, <i>Aetobatus narinari</i>	JX978310	-----G-----INSERT-----C-
Carangidae, <i>Alectis ciliaris</i>	NC_025566	-----A-----INSERT-----T-
Lutjanidae, <i>Apsilus dentatus</i>	FJ694227	-----A-----INSERT-----C-
Ariidae, <i>Arius madagascariensis</i>	FJ626110	-----G-----INSERT-----C-
Tetradontidae, <i>Arothron hispidus</i>	NC_015336	-----A-----INSERT-----C-
Atherenidae, <i>Atherinomorus stipes</i>	AF150001	-----G-----INSERT-----C-
Balistidae, <i>Balistoides viridescens</i>	AY700250	-----A-----INSERT-----C-
Scaridae, <i>Calotomus carolinus</i>	AY081074	-----G-----INSERT-----C-
Carangidae, <i>Carangoides armatus</i>	AP004444	-----A-----INSERT-----T-
Carangidae, <i>Caranx sexfaciatus</i>	KM082979	-----G-----INSERT-----C-----T-
Carangidae, <i>Caranx tille</i>	NC_029421	-----A-----INSERT-----T-
Serranidae, <i>Cephalopholis argus</i>	NC_022142	-----G-----INSERT-----C-
Serranidae, <i>Cephalopholis sonnerati</i>	NC_022143	-----G-----INSERT-----C-
Chanidae, <i>Chanos chanos</i>	NC_004693	A-----G-----INSERT-----C-
Labridae, <i>Cheilinus trilobatus</i>	AJ810128	-----G-----INSERT-----C-
Labridae, <i>Cheilinus undulates</i>	NC_013842	-----G-----INSERT-----C-
Labridae, <i>Cheilio inermis</i>	AY279583	-----G-----INSERT-----C-
Chirocentridae, <i>Chirocentrus dorab</i>	AP006229	-----G-----INSERT-----C-----C-
Scaridae, <i>Chlorurus sordius</i>	AP006567	-----G-----INSERT-----C-
Congridae, <i>Conger japonicas</i>	KR131863	AA-----G-----INSERT-----TT
Dasyatidae, <i>Dasyatis akajei</i>	NC_021132	-----G-----INSERT-----C-
Haemulidae, <i>Diagramma picta</i>	NC_009856	-----A-----INSERT-----C-
Diodontidae, <i>Diodon holocanthus</i>	NC_009866	-----G-----C-----INSERT-----C-
Sparidae, <i>Diplodus sargus</i>	KC987056	-----A-----INSERT-----C-
Serranidae, <i>Epinephelus lanceolatus</i>	NC_011715	-----G-----INSERT-----C-
Serranidae, <i>Epinephelus polyphekadion</i>	AY279560	-----G-----INSERT-----C-
Serranidae, <i>Epinephelus quoyanus</i>	NC_021450	-----G-----INSERT-----C-
Clupeidae, <i>Gilchristella aestuaria</i>	AP011606	-----A-----INSERT-----C-

Scombridae, <i>Gymnosarda unicolor</i>	NC_022496	-----G-----INSERT-----C-
Muraenidae, <i>Gymnothorax kidako</i>	AP002976	-----G-----INSERT-----C-
Dasyatidae, <i>Himantura uarnak</i>	AF447997	-----G-----INSERT-----C-
Hemiramphidae, <i>Hyporhamphus regularis</i>	AF092206	-----G--T----INSERT-----A-----CA
Scaridae, <i>Leptoscarus vaigiensis</i>	AY081076	-----G-----INSERT-----C-
Lethrinidae, <i>Lethrinus harak</i>	KU680995	-----A-----INSERT-----C-
Lethrinidae, <i>Lethrinus lentjan</i>	AY484981	-----A-----INSERT-----C-
Lethrinidae, <i>Lethrinus mahsena</i>	KU680997	-----A-----INSERT-----C-
Mugilidae, <i>Liza affinis</i>	JF911709	-----G-----INSERT-----C-
Lutjanidae, <i>Lutjanus argentimaculatus</i>	NC_016661	-----A-----INSERT-----C-
Lutjanidae, <i>Lutjanus rivulatus</i>	NC_009869	-----A-----INSERT-----C-
Lutjanidae, <i>Lutjanus sanguineus</i>	AY484976	-----A-----INSERT-----C-
Myliobatidae, <i>Manta birostris</i>	KF413894	-----G-----INSERT-----CA
Monodactylidae, <i>Monodactylus argenteus</i>	NC_009858	-----A-----INSERT-----C-
Lethrinidae, <i>Monotaxis grandoculis</i>	NC_010957	-----A-----INSERT-----CT
Mugilidae, <i>Mugil cephalus</i>	NC_003182	-----G-----INSERT-----C-
Mugilidae, <i>Mulloidichthys vanicolensis</i>	AP012310	T-----G-----INSERT-----CA
Acanthuridae, <i>Naso brachycentron</i>	AY057256	-----A-----INSERT-----C-
Acanthuridae, <i>Naso lituratus</i>	AF055603	-----A-----INSERT-----C-
Labridae, <i>Novaculichthys taeniourus</i>	AY279627	-----G-----INSERT-----C-
Labridae, <i>Oxycheilinus digramma</i>	EU601221	-----G-----INSERT-----C-
Dasyatidae, <i>Pastinachus sephen</i>	JX978312	-----G-----INSERT-----C-
Haemulidae, <i>Plectorhinchus cinctus</i>	NC_024587	-----A-----INSERT-----C-
Haemulidae, <i>Plectorhinchus orientalis</i>	NC_027097	-----A-----INSERT-----C-
Haemulidae, <i>Pomadasyys perotaei</i>	AY368293	-----G-----INSERT-----C-
Pomacanthidae, <i>Pygoplites diacanthus</i>	NC_026545	-----A-----INSERT-----C-
Sparidae, <i>Rhabdosargus sarba</i>	NC_025301	-----A-----INSERT-----C-
Scaridae, <i>Scarus rubroviolaceus</i>	NC_011343	-----G-----INSERT-----C-
Nemipteridae, <i>Scolopsis vosmeri</i>	KT692978	A-----G-----INSERT-----C-
Scombridae, <i>Scomberomorus commerson</i>	HM003557	-----G-----INSERT-----CA
Carangidae, <i>Selar crumenophthalmus</i>	NC_023954	-----A-----INSERT-----T-
Carangidae, <i>Seriola lalandi</i>	NC_016869	-----A-----INSERT-----T-
Sillaginidae, <i>Sillago sihama</i>	NC_016672	-----G-----INSERT-----C-
Sphyraenidae, <i>Sphyraena barracuda</i>	NC_022484	-----G-----INSERT-----T-
Sphyraenidae, <i>Sphyrna lewini</i>	JX827259	--T-----G-----INSERT-----CA
Engraulidae, <i>Stolephorus chinensis</i>	AP011566	-----A-----INSERT-----CA
Belonidae, <i>Strongylura incise</i>	AF231550	-----G-----INSERT-----C-

Balistidae, <i>Sufflamen chrysopterum</i>	AY700251	-----G-----INSERT-----C-
Synodontidae, <i>Synodus variegatus</i>	AY524977	-----G-----INSERT-----C-
Dasyatidae, <i>Taeniura lymma</i>	NC_026210	-----G-----INSERT-----C-
Dasyatidae, <i>Taeniura meyeri</i>	NC_019641	-----G-----INSERT-----C-
Terapontidae, <i>Terapon jarbua</i>	NC_027281	-----G-----INSERT-----C-----C-
Carangidae, <i>Trachurus trachurus</i>	AB108498	-----G-----INSERT-----C-----T-
Trichiduridae, <i>Trichiurus lepturus</i>	NC_018791	A---C-----G-----INSERT-----G-CT
Mugilidae, <i>Valamugil speigleri</i>	KF374995	-----G-----INSERT-----C-
Serranidae, <i>Variola louti</i>	NC_022138	-----G-----INSERT-----C-
Acanthuridae <i>Zebrasoma flavescens</i>	NC_009874	-----A-----INSERT-----C-

---

**FIGURE S4.8.1** | AN ALIGNMENT OF 80 LOCAL FISH TAXA SHOWING THE CONSERVED PRIMER-BINDING REGIONS FLANKING THE *12S* METABARCODING REGION ('INSERT'). Forward (FWD) and reverse (REV) primers are shown at the top. Conserved bases are indicated by a '-'. The sites with the most variability (e.g., base 13 in the forward primer) were accounted for through the use of degenerate bases in the primer sets.

Consensus		GT-CAGCTTACCCTGT-GAAGG--MCYAATA-G-TAAGCA-MAATTGGCACA-RCC--CAAAACGT
Belonidae, <i>Ablennes hians</i>	AB373007	..-....CC.....-.....-A.TCG...-.....-A...CA.T...-A.T--.....C
Sparidae, <i>Acanthopagrus australis</i>	KJ774772	..T.....-.....-GTA..G.-.....-A.GC.....TC-A...-..GT....
Acanthuridae, <i>Acanthurus mata</i>	AY057237	..-.....-.....-ATT.....-.....-A..C.....A.-G.....
Acanthuridae, <i>Acanthurus triostegus</i>	AY057246	.....A.....-.....-GTT.....-.....-A..C.....A.-G.....
Serranidae, <i>Aethaloperca rogaa</i>	NC_022141	.....A.....-.....-C.CCT.....-.....-T.....-G.....
Myliobatidae, <i>Aetobatus narinari</i>	JX978310	.C.....C.....A.....-.....-TTT..C.....-.....-A..A.C.T--..CT...C....
Carangidae, <i>Alectis ciliaris</i>	NC_025566	.C.....-.....-C.T.....-.....-C.....-G.....
Lutjanidae, <i>Apsilus dentatus</i>	FJ694227	.....-.....-C.TT.....-.....-A.....-G.....
Ariidae, <i>Arius madagascariensis</i>	FJ626110	.....-.....-C...C.....-.....-G...G..TT..CC.---A.....
Tetradontidae, <i>Arothron hispidus</i>	NC_015336	.....C.....-.....-A.T.....-.....-G...-A.G.....-G.....
Atherenidae, <i>Atherinomorus stipes</i>	AF150001	.....-.....-ATT.....-.....-A.GCCA.....-G.T--.....C
Balistidae, <i>Balistoides viridescens</i>	AY700250	.....T.....-.....-A.TTT..-A.....-A.....T.A.-A...-T.....
Scaridae, <i>Calotomus carolinus</i>	AY081074	T.-.T.....A.....-.....-A.AC.....-G.....-T.....TTTT-A...-T.....
Carangidae, <i>Carangoides armatus</i>	AP004444	.C.....-.....-C.T.....-.....-C.....T.-G.....
Carangidae, <i>Caranx sexfaciatus</i>	KM082979	.....-.....-ATT.....-.....-C.....T...-A.....
Carangidae, <i>Caranx tille</i>	NC_029421	.C.....-.....-A.T.....-.....-C...C.....-G.....
Serranidae, <i>Cephalopholis argus</i>	NC_022142	.....A.....-.....-C.CTTC.....-.....-T.....T...-A.....
Serranidae, <i>Cephalopholis sonnerati</i>	NC_022143	.....A.....-.....-C.CCT.....-.....-C.....T...-A.....
Chanidae, <i>Chanos chanos</i>	NC_004693	.....-.....-ATGCC.....-.....-A...GA...-GA.T--.....
Labridae, <i>Cheilinus trilobatus</i>	AJ810128	.....C-A.....-A.CT.C.....-A.....TA.CC..C.....C
Labridae, <i>Cheilinus undulates</i>	NC_013842	.....AC-A.....-ATCT.A.....-A.....TA.CC..C.....C
Labridae, <i>Cheilio inermis</i>	AY279583	.....C.....A.....-.....-C.CC.A.....-G...-A.....-G.....
Chirocentridae, <i>Chirocentrus dorab</i>	AP006229	.C.....-.....-A.....-ATT.A.A.....-C..CGA.A.TTAC.T--.....
Scaridae, <i>Chlorurus sordius</i>	AP006567	-.C...TC.....-A.....-A.CC.....-G..G-T.....TTT.-G...-C.G...A
Congridae, <i>Conger japonicas</i>	KR131863	C.....G..T...GA.....-A.GG.....-C.....-C..A...GTT.TA...-G..A..
Dasyatidae, <i>Dasyatis akajei</i>	NC_021132	.....C.....-.....A--TAC..C.....-T...GAC.TTTTA..CT...C....
Haemulidae, <i>Diagramma picta</i>	NC_009856	.....-.....-C.C..C.....-G.....CC.....
Diodontidae, <i>Diodon holocanthus</i>	NC_009866	.....CA...G...-A.CC.C.....-T...A...T.-G.....
Sparidae, <i>Diplodus sargus</i>	KC987056	.....-.....-AC..A.A.....-A..C.....C-A...-G.....
Serranidae, <i>Epinephelus lanceolatus</i>	NC_011715	.....A.....-.....-T.C.T.....-.....-C.....T...-A.....
Serranidae, <i>Epinephelus polyphemus</i>	AY279560	.....A.....-.....-C.CTT.....-.....-C.....T...-A.....
Serranidae, <i>Epinephelus quoyanus</i>	NC_021450	.C.....A.....-.....-C.CCC.....-G.....TG.....T.A.-A.....
Clupeidae, <i>Gilchristella aestuaria</i>	AP011606	.C.....-.....-CACT.C.....-G...GA.T..TTCT--.....
Scombridae, <i>Gymnosarda unicolor</i>	NC_022496	.....-.....-C.CT.....-.....-A.....T..C-A...-G.....
Muraenidae, <i>Gymnothorax kidako</i>	AP002976	.C...T...T.T.GA.....-AAC..C.....-A..CG..T.TT-A...-C....
Dasyatidae, <i>Himantura uarnak</i>	AF447997	.....C.....A.....-.....-TTC.....-.....CA...-A.C.TCT..CT...C....

Hemiramphidae, <i>Hyporhamphus regul.</i>	AF092206	.G-.....-....CT-C.TG.C.T.C.....TGTG.CA.T.A.-G.T--.....C
Scaridae, <i>Leptoscarus vaigiensis</i>	AY081076	T.-....C.....-A.....-T.TC.C.-..GG...-C.....-G.--.....
Lethrinidae, <i>Lethrinus harak</i>	KU680995	..-.....C.-.....-TTC.....-C.....-G.--...G....
Lethrinidae, <i>Lethrinus lentjan</i>	AY484981	..-.....C.-..G.-..-T.C.....-C.....-G.--.....
Lethrinidae, <i>Lethrinus mahsena</i>	KU680997	.....C.....-TTT.....-C.....-G.....
Mugilidae, <i>Liza affinis</i>	JF911709	.....C.....-T.CCG.....-T-AGG.....G-A.--.....C
Lutjanidae, <i>Lutjanus argentimacula.</i>	NC_016661	.....A.....-A.TT.....-AG.....T.-G.--.....
Lutjanidae, <i>Lutjanus rivulatus</i>	NC_009869	.....-.....-ATTC.....-AG.....T.-G.--.....
Lutjanidae, <i>Lutjanus sanguineus</i>	AY484976	.....-.....-G.CT.....-AG.....-G.--.....
Myliobatidae, <i>Manta birostris</i>	KF413894	.....C.....A.....-CAT..C.-.....-A..A.CT.CCT.TT...T....
Monodactylidae, <i>Monodactylus argen.</i>	NC_009858	.....A.....-C.CT.....-A.....T.-G.--.....
Lethrinidae, <i>Monotaxis grandoculis</i>	NC_010957	.....CC.....-A.GC.....-GC...-A.....T.T-G.--.....C
Mugilidae, <i>Mugil cephalus</i>	NC_003182	.....CC.....-G.-..-T.C.....-G....-GG..C..T.G.-A.--.....C
Mugilidae, <i>Mulloidichthys vanicole.</i>	AP012310	.....C.....-G..C.TAG.GG..T-A..C...T..T-A.--.TT..T..
Acanthuridae, <i>Naso brachycentron</i>	AY057256	.....-.....-CCT..TAG.....-A.....A.-G.--.....
Acanthuridae, <i>Naso lituratus</i>	AF055603	.....-.....-C.CT.....-A.....A.-G.--.....
Labridae, <i>Novaculichthys taeniourus</i>	AY279627	.....-.....-C.A.....-A..C.....T-G.--.....
Labridae, <i>Oxycheilinus digramma</i>	EU601221	.....C-A.....-ATCT.A.-.....-A.....T..CC..C-.....C
Dasyatidae, <i>Pastinachus sephen</i>	JX978312	.....C....CA.-..G.-..-CTC..C.-.....-T..CGAC.TT.AT..T...T....
Haemulidae, <i>Plectorhinchus cinctus</i>	NC_024587	.....-.....-C.CT.C.-.....-A.....C..CC---T.....C
Haemulidae, <i>Plectorhinchus orienta.</i>	NC_027097	.....-.....-C.C..C.-.....-G.....CC---.....
Haemulidae, <i>Pomadasy perotaei</i>	AY368293	.....-.....-AT.A.A.-.....-G.....-G.--.....C
Pomacanthidae, <i>Pygoplites diacanth.</i>	NC_026545	.....T.....-.....-C.C..C.-.....-T..C.....-G.--.....
Sparidae, <i>Rhabdosargus sarba</i>	NC_025301	.....T.-..G.-..-A.A.....-.....-A..C....TC-A.--.GC....
Scaridae, <i>Scarus rubroviolaceus</i>	NC_011343	-C...TC.....-A.....-A.CC.....-G..G-A.....T...-G.--.C.G...A
Nemipteridae, <i>Scolopsis vosmeri</i>	KT692978	.C-.....C....CT.-..G..T-AAT..C.T.-G...T-C..CC..TTA.CC..C...T.A..
Scombridae, <i>Scomberomorus commerson</i>	HM003557	..-A.....-.....-C.T.....-A.....C-G.--.G.....
Carangidae, <i>Selar crumenophthalmus</i>	NC_023954	.....-.....-ATTT.....-G....-T..C...T.A.-G.--.G.....
Carangidae, <i>Seriola lalandi</i>	NC_016869	.....-.....-A.G.....-.....-C.....-G.--.T....
Sillaginidae, <i>Sillago sihama</i>	NC_016672	.C-A......C.-..G.-..-A.A.....-.....T-C..CC....T-G.--.G.....
Sphyraenidae, <i>Sphyraena barracuda</i>	NC_022484	.....-.....-C.A.....-.....-C..C.....-G.--.GT....
Sphyraenidae, <i>Sphyrna lewini</i>	JX827259	.....C.....A.....-AT..A.A.-.....-A.....-A.T.AAA.TC.....
Engraulidae, <i>Stolephorus chinensis</i>	AP011566	.C-.....A.....-.....-A.A.-.....-A..AA.A.GTTT--.GG.....
Belonidae, <i>Strongylura incise</i>	AF231550	.....CA.....-ATTC.....-A..CCA.T...-A.T--.....C
Balistidae, <i>Sufflamen chrysopterum</i>	AY700251	.....T.....-.....-A.CCT..A--.....-A.....T.A.-G.--T.GT....
Synodontidae, <i>Synodus variegatus</i>	AY524977	.....T.-.....-TA..G.A.-.....T-C...A.TT.A.AA.--A..C....
Dasyatidae, <i>Taeniura lymma</i>	NC_026210	.....C....CA.-..G.-..-TTT..C.-.....CA.--.A.CTTCAT.CT...T....

Dasyatidae, <i>Taeniura meyeni</i>	NC_019641	..-.....C.....-.....--CAT..C.-.-.....TA.--..A.CTTTA..CT...T....
Terapontidae, <i>Terapon jarbua</i>	NC_027281	.C-.....A.-.....---TC.C.CAG.....-G.....-G....-G.....
Carangidae, <i>Trachurus trachurus</i>	AB108498	..-.....-.....-.....-A.T.....-.....G-C.....T...-A.-.....
Trichiduridae, <i>Trichiurus lepturus</i>	NC_018791	T.-...T.....A.-...T.--T.C..C.-.-...T..-C.....T.-G....-G.....
Mugilidae, <i>Valamugil speigleri</i>	KF374995	..-.....C.....A.-...G...-ATC..C.-.-.....-A....A...TG-G....-.....
Serranidae, <i>Variola louti</i>	NC_022138	..-.....CA..G....-AAC.C.CAG.....-C.....T.-.TA....-.....
Acanthuridae <i>Zebrasoma flavescens</i>	NC_009874	..-.....-.....---TT..A.CAG.....-A.....T.-G....-.....

---

**FIGURE S4.8.2** | AN ALIGNMENT OF 80 LOCAL FISH TAXA SHOWING THE VARIABLE METABARCODING REGION (LABELED ‘INSERT’ IN FIGURE S4.8.1). A consensus sequence is shown at the top. Bases that are identical to the consensus are indicated by a ‘.’. ‘-’ represent gaps in the alignment. Forward (FWD) and reverse (REV) primers are shown at the top. Conserved bases are indicated by a ‘-’. The sites with the most variability (e.g., base 13 in the forward primer) were accounted for through the use of degenerate bases in the primer sets.

because long aDNA templates are exponentially rarer than short ones (cf., Grealy et al. 2015).

The variability in the barcoding region in Figure S4.8.2, summarized by the identity matrices presented in Table S4.8.1, exemplify that although the primers amplify only 56 bp, changes in 1-2 bp (that could be caused by DNA damage, amplification error, or sequencing error) would not change genus or family-level taxonomic identifications (with only two exceptions) and at least 3 bp needs to differ between two samples to potentially change the family-level identification; however, in most cases, five or more base-pair differences are needed to change the family-level identification. Thus, it is unlikely that error would result in genus or family misidentifications. In the situations where no differences exist between the two taxa (e.g., *Diagramma picta* and *Plectorhinchus orientalis*), had this sequence been detected it would have been identified to the family-level only.

To further model the effect of damage and error on taxonomic assignments at the species, genus, and family level for every sequence/taxon identified in our data, we applied a 1%, 2% and 3% mutation rate to a single sequence and generated 1000 sequences with randomly dispersed mutations, that were then BLAST to the GenBank database (Table S4.8.2). This was done using an open source python script (available: <https://www.biostars.org/p/12417/>). This approach has been used with short metabarcoding primer sets in the past (Grealy et al. 2016). These results show that even with a 3% mutation rate (on average, almost every sequence having two errors; well above the combined error expected damage, amplification and sequencing), most sequences were not assigned to a different species, genus or family; rather, erroneous sequences were pushed back to a higher taxonomic classification (e.g., species come genus, genus becomes family, family becomes order, etc.) or were discarded because hits no longer aligned over 100% of the reference sequence. Where sequences would result in a taxonomic mis-identification, these typically occur at a rate of 0.1% and would be removed by abundance filtering (the removal of rare sequences). This, coupled with stringent filtering (base quality filtering and chimera filtering, as well as abundance filtering), high sequencing coverage, and independent PCR amplification reactions, makes it unlikely that damage/error would result in genus or family-level misidentifications. Finally, if a

sequence has arisen through error, we would expect to see both the “parent” (true, higher abundant sequence) and “descendent” (erroneous, lower abundant sequence) taxa: the fact that we did not observe any of the mis-identified taxa in Table S4.8.2 (except Carangidae, but this ID was not found in the same sample as *Psammoperca*) suggests that the quality control measures employed here were adequate to remove erroneous sequences.

Taxonomic assignments were also made conservatively: query sequences needed to be aligned to a reference across the entire length of the query, the similarity of the query to the reference needed to be within the pre-defined identity cut-offs (see main text), and *I2S* genetic reference sequences for all subtaxa within a family or genus also needed to be present in order to call an ID at the species- or genus-level. Sequences that were equally similar to multiple taxa were identified only to the level of the last common ancestor. For example, a sequence 100% identical to *Epinephelus* and *Aethaloperca* would be assigned to the family Serranidae; a sequence 100% identical to *Mulloidichthys flavolineatus* only would be assigned to the genus *Mulloidichthys* because not every species within this genus has a representative *I2S* sequence in GenBank.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
41 <i>Lethrinus mahsena</i>	16	13	5	5	8	15	5	5	14	8	10	12	16	7	8	8	12	9	15	16	15	10	20	21	26	20	9	13	11	7	9	14	16	11	20	17	24	11	2	3	
42 <i>Liza affinis</i>	12	18	16	17	12	21	16	13	20	13	15	17	19	17	15	17	13	12	17	17	19	14	26	21	30	20	16	18	16	12	13	14	22	12	24	18	26	17	15	14	
43 <i>Lutjanus argentimaculatus</i>	15	14	6	6	8	19	8	4	16	8	12	9	13	8	12	10	11	10	15	13	13	9	20	19	25	22	11	10	13	10	9	11	17	10	19	20	23	14	9	8	
44 <i>Lutjanus rivulatus</i>	14	12	4	6	9	19	8	5	16	8	10	10	14	8	10	10	14	11	12	14	14	10	20	18	25	21	11	11	12	12	12	12	17	11	19	20	23	13	8	9	
45 <i>Lutjanus sanguineus</i>	15	14	8	8	8	21	8	4	13	9	12	11	17	10	13	11	11	10	15	13	15	9	21	20	27	22	7	13	11	10	9	12	13	8	21	20	23	14	8	7	
46 <i>Manta birostris</i>	24	22	19	18	19	11	20	19	19	22	23	25	24	18	21	23	19	20	24	20	20	20	24	29	32	18	17	21	22	20	20	23	22	22	25	16	31	23	20	21	
47 <i>Monodactylus argenteus</i>	16	13	7	6	5	19	7	3	14	9	13	10	13	7	13	11	8	7	15	12	12	6	20	18	25	20	8	9	11	8	6	10	14	7	18	18	23	14	7	6	
48 <i>Monotaxis grandoculis</i>	14	20	13	15	14	23	14	12	20	11	16	17	15	16	18	16	20	16	19	16	18	12	28	19	27	23	16	18	17	17	17	18	21	15	25	22	28	14	15	14	
49 <i>Mugil cephalus</i>	12	21	16	17	15	23	16	15	16	16	16	16	17	17	17	15	17	15	13	20	18	20	15	26	21	30	21	15	17	18	11	13	12	20	13	22	23	23	16	15	12
50 <i>Mulloidichthys vanicolensis</i>	25	23	21	22	24	28	24	22	24	21	24	23	22	26	22	26	22	22	26	27	29	20	28	28	32	30	23	27	18	21	22	24	26	19	25	30	31	21	24	24	
51 <i>Naso brachycentron</i>	17	17	9	10	8	24	11	8	18	12	16	9	18	12	16	14	13	10	16	15	17	11	23	21	29	24	12	14	12	11	11	13	19	12	23	23	23	16	11	10	
52 <i>Naso lituratus</i>	15	13	5	6	6	20	6	2	13	8	12	8	15	7	12	10	9	8	14	10	12	7	20	18	25	20	7	11	10	9	7	10	13	6	19	19	21	13	7	6	
53 <i>Novaculichthys taeniourus</i>	15	11	6	7	7	16	7	5	12	9	9	13	13	9	11	11	11	9	15	14	16	8	18	20	25	19	9	13	9	9	9	13	15	7	16	16	23	13	9	8	
54 <i>Oxycheilinus digramma</i>	18	19	13	15	14	21	16	13	15	16	14	17	19	18	15	18	16	15	17	3	2	13	21	20	26	20	12	17	15	15	14	19	18	13	21	18	28	17	14	15	
55 <i>Pastinachus sephen</i>	24	21	19	18	18	18	22	21	19	23	20	27	23	21	21	23	17	20	21	22	21	20	23	26	34	12	19	14	22	20	20	22	22	22	23	17	32	25	19	19	
56 <i>Plectorhinchus cinctus</i>	17	18	13	14	12	21	12	8	9	14	15	13	20	14	17	16	12	13	18	9	12	12	22	23	29	22	5	16	14	14	12	17	12	11	23	20	24	17	13	12	
57 <i>Plectorhinchus orientalis</i>	17	15	10	11	9	19	8	7	6	11	14	14	18	10	13	12	10	10	16	11	14	10	20	22	27	19	0	13	11	10	10	14	9	10	20	19	22	14	9	8	
58 <i>Pomadasy perotaei</i>	15	16	11	12	9	21	10	8	14	12	12	14	16	12	13	13	12	11	17	12	13	9	18	20	27	22	10	13	10	11	10	15	16	10	22	22	20	13	11	10	
59 <i>Pygoplites diacanthus</i>	16	16	10	11	8	19	8	7	12	9	13	15	18	9	12	8	8	11	17	15	18	10	22	17	29	17	7	12	12	11	11	14	14	11	18	20	22	13	9	8	
60 <i>Rhabdosargus sarba</i>	19	8	9	10	14	20	13	11	19	13	15	14	15	13	15	13	16	14	18	18	19	15	23	23	25	21	14	14	9	13	14	17	21	10	15	20	30	20	12	10	
61 <i>Scarus rubroviolaceus</i>	17	24	18	20	16	28	19	16	22	13	20	20	20	20	17	19	17	16	22	19	21	13	28	3	30	28	20	20	22	17	17	20	25	16	22	26	31	14	17	18	
62 <i>Scolopsis vosmeri</i>	27	28	22	23	27	25	24	27	23	28	25	27	26	25	24	24	27	25	28	22	23	28	27	32	32	26	24	24	26	24	25	24	26	27	24	28	30	27	25	24	
63 <i>Scomberomorus commerson</i>	16	13	9	10	9	19	8	6	14	10	12	14	17	10	12	10	13	11	17	16	18	10	22	22	25	22	11	15	10	11	11	15	18	6	21	20	22	14	11	10	
64 <i>Selar crumenophthalmus</i>	18	14	6	8	13	22	10	9	18	13	13	11	15	11	10	10	14	14	18	17	17	16	21	21	24	25	15	16	13	14	13	13	18	11	21	24	25	16	12	13	
65 <i>Seriola lalandi</i>	16	13	7	9	9	19	4	6	14	8	12	12	16	6	8	6	13	10	15	15	17	11	22	19	22	21	9	13	12	9	10	15	17	11	20	20	25	13	6	6	
66 <i>Sillago sihama</i>	20	18	14	15	15	20	13	15	20	15	15	21	19	14	15	9	17	16	21	22	23	17	21	24	28	24	18	17	16	15	16	19	21	15	20	22	29	19	14	11	
67 <i>Sphyrna barracuda</i>	16	15	12	13	9	20	7	9	14	11	13	17	19	8	9	5	11	10	17	17	19	11	22	19	24	19	11	15	13	10	10	15	18	9	21	19	24	14	9	9	
68 <i>Sphyrna lewini</i>	20	23	19	18	20	22	21	19	22	21	21	22	23	20	19	22	19	19	21	23	21	19	18	30	33	20	21	21	17	18	19	22	23	20	24	23	29	24	21	21	
69 <i>Stolephorus chinensis</i>	26	23	19	18	21	22	20	20	24	23	24	23	25	20	23	20	21	21	23	25	23	21	20	31	32	30	20	24	17	21	21	21	21	21	21	28	26	29	26	21	21
70 <i>Strongylura incisa</i>	8	18	11	11	13	21	16	13	17	16	6	15	16	18	12	16	14	11	12	18	16	14	19	23	29	24	17	13	16	12	12	15	18	13	19	22	21	19	15	16	
71 <i>Sufflamen chrysopterum</i>	17	18	13	15	11	26	16	13	18	16	18	8	18	17	16	16	14	11	18	17	19	14	25	20	28	26	17	16	15	12	12	14	22	11	22	25	26	19	15	15	
72 <i>Synodus variegatus</i>	22	19	17	17	19	23	19	19	15	22	20	19	21	20	16	21	18	16	20	22	21	20	21	27	31	23	17	22	18	15	16	19	21	18	21	24	28	23	16	18	
73 <i>Taeniura lymma</i>	26	20	20	19	23	12	24	22	24	23	23	26	24	23	22	26	22	23	25	22	21	23	28	32	36	17	22	20	23	21	23	26	26	23	25	8	35	25	21	21	
74 <i>Taeniura meyeni</i>	23	18	18	19	21	13	20	18	21	19	21	23	23	19	19	22	19	21	23	20	22	21	24	28	32	12	19	23	19	21	21	24	22	19	22	10	30	23	21	22	
75 <i>Terapon jarbua</i>	21	20	15	14	12	19	13	13	18	17	18	18	20	15	17	14	15	14	21	20	20	13	22	26	31	26	15	15	15	15	15	16	19	15	23	25	24	16	15	15	
76 <i>Trachurus trachurus</i>	12	17	10	12	11	19	7	9	13	10	13	11	14	8	2	8	10	8	15	16	18	13	18	17	21	21	12	15	14	7	8	13	18	9	19	21	24	13	11	10	
77 <i>Trichiurus lepturus nanhaiensis</i>	25	21	18	19	17	25	17	17	21	19	22	23	22	16	19	17	18	18	22	21	24	18	27	23	26	24	17	16	19	16	18	22	25	17	24	27	30	16	16	14	
78 <i>Valamugil speigleri</i>	15	16	11	11	12	15	16	15	20	14	13	18	17	14	14	16	14	14	17	16	12	20	22	30	17	14	8	15	13	14	17	18	14	18	17	26	17	12	11		
79 <i>Variola louti</i>	21	22	19	20	18	27	19	20	21	20	23	18	20	19	15	20	16	15	19	19	21	19	23	25	29	22	18	18	18	16	16	19	20	17	23	27	28	19	19	17	
80 <i>Zebрасoma flavescens</i>	19	13	7	8	13	21	11	8	18	11	13	12	19	11	13	13	16	15	17	18	17	12	20	23	29	23	12	15	10	14	15	18	19	14	21	22	24	16	10	11	

**TABLE S4.8.2** | TAXONOMIC ASSIGNMENTS OF 1000 SEQUENCES GENERATED FROM MUTATING ONE SEQUENCE OF EACH TAXON AT A RATE OF 1%, 2% AND 3%. Dark grey cells indicate potential taxonomic misidentifications. ‘>Family’ refers to matches that have been pushed beyond the family-level (e.g., to order-level), and would be discarded. ‘Other’ refers to matches to other taxa that have not aligned across the total length of query (and would be discarded). \* not all genera within the family have a reference *12S* sequence in GenBank and this hit would therefore not be called at the genus-level.

Taxon	Error rate		
	1%	2%	3%
<i>Himantura gerrardi</i>	<i>Himantura gerrardi</i> = 970 <i>Himantura</i> = 28	<i>Himantura gerrardi</i> = 928 <i>Himantura</i> = 68 Dasyatidae = 4	<i>Himantura gerrardi</i> = 889 <i>Himantura</i> = 100 Dasyatidae = 10 > Family = 1
<i>Psammoperca waigiensis</i>	<i>Psammoperca waigiensis</i> = 945 > Family = 55	<i>Psammoperca waigiensis</i> = 880 > Family = 120	<i>Psammoperca waigiensis</i> = 816 Bathycluepidae = 1 Carangidae = 1 > Family = 182
<b>Ginglymostomatidae</b>	Ginglymostomatidae = 859 > Family = 141	Ginglymostomatidae = 758 > Family = 242	Ginglymostomatidae = 657 > Family = 343
<i>Carcharhinus</i>	<i>Carcharhinus</i> = 994 Carcharhinidae = 6	<i>Carcharhinus</i> = 977 Carcharhinidae = 18 >Family = 1	<i>Carcharhinus</i> = 972 Carcharhinidae = 17 Other = 11
<b>Carangidae</b>	Carangidae = 982 >Family = 18	Carangidae = 953 >Family = 47	Carangidae = 920 >Family = 80
<i>Mulloidichthys</i>	<i>Mulloidichthys</i> = 999	<i>Mulloidichthys</i> = 993	<i>Mulloidichthys</i> = 982

	>Family = 1	Mullidae = 5 > Family = 2	Mullidae = 11 > Family = 7
<b>Liza</b>	<i>Liza</i> = 975 Mugilidae =25	<i>Liza</i> = 970 Mugilidae =30	<i>Liza</i> = 957 Mugilidae =43
<b>Monotaxis grandoculis</b>	<i>Monotaxis grandoculis</i> =1000	<i>Monotaxis grandoculis</i> =1000	<i>Monotaxis grandoculis</i> =998 > Family = 2
<b>Lethrinus lentjan</b>	<i>Lethrinus lentjan</i> = 924 <i>Lethrinus</i> = 2 > Family = 74	<i>Lethrinus lentjan</i> = 886 <i>Lethrinus</i> = 3 > Family = 111	<i>Lethrinus lentjan</i> = 835 <i>Lethrinus</i> = 7 > Family = 158
<b>Acanthurus triostegus</b>	<i>Acanthurus triostegus</i> =965 <i>Acanthurus</i> = 3 Acanthuridae =16 > Family = 16	<i>Acanthurus triostegus</i> =915 <i>Acanthurus</i> = 5 Acanthuridae =36 > Family = 44	<i>Acanthurus triostegus</i> =863 <i>Acanthurus</i> = 6 Acanthuridae =53 > Family = 78
<b>Gerres oyena</b>	<i>Gerres oyena</i> = 1000	<i>Gerres oyena</i> = 1000	<i>Gerres oyena</i> = 1000
<b>Scaridae</b>	Scaridae=1000	Scaridae=999 > Family = 1	<b>Scaridae</b> =6 Scaridae =993 Other = 1
<b>Hemigymnus melapterus</b>	<i>Hemigymnus melapterus</i> =998 Labridae =2	<i>Hemigymnus melapterus</i> =991 Labridae = 7 > Family = 2	<i>Hemigymnus melapterus</i> =983 Labridae =14 > Family = 1 Other = 2
<b>Novaculichthys taeniourus</b>	<i>Novaculichthys taeniourus</i> =958 Labridae = 7 > Family = 35	<i>Novaculichthys taeniourus</i> =930 Labridae = 11 > Family = 59	<i>Novaculichthys taeniourus</i> =880 Labridae = 15 > Family = 105
<b>Epinephelus</b>	<i>Epinephelus</i> = 1000	<i>Epinephelus</i> =999 > Family = 1	<i>Epinephelus</i> = 997 Epinephelinae=1 > Family = 2

<b>Ballistidae</b>	<i>Rhinecanthus</i> = 2 Ballistidae = 998	<i>Rhinecanthus</i> = 7 Ballistidae = 992 > Family = 1	<i>Rhinecanthus</i> = 3 Ballistidae = 994 > Family = 3
<b><i>Siganus</i></b>	<i>Siganus</i> = 988 > Family = 12	<i>Siganus</i> = 965 > Family = 35	<i>Siganus</i> = 954 > Family = 46
<b>Cichlidae</b>	<i>Pelmatochromis</i> = 3* Cichlidae = 835 > Family = 59 Other = 103	<i>Pelmatochromis</i> = 1* Cichlidae = 668 > Family = 21 Other = 290	Cichlidae = 560 > Family = 292 Other = 147
<b>Sparidae</b>	Sparidae = 802 > Family = 198	Sparidae = 646 > Family = 110 Other = 244	Sparidae = 552 > Family = 129 Other = 316
<b><i>Megalops cyprinoides</i></b>	<i>Megalops cyprinoides</i> = 994 <i>Megalops</i> = 6	<i>Megalops cyprinoides</i> = 982 <i>Megalops</i> = 18	<i>Megalops cyprinoides</i> = 965 <i>Megalops</i> = 35
<b><i>Kyphosus bigibbus</i></b>	<i>Kyphosus bigibbus</i> = 984 <i>Kyphosus</i> = 4 > Family = 12	<i>Kyphosus bigibbus</i> = 956 <i>Kyphosus</i> = 12 Bythitidae = 1 > Family = 30 Other = 1	<i>Kyphosus bigibbus</i> = 947 <i>Kyphosus</i> = 8 > Family = 41 Other = 1
<b><i>Abudefduf</i></b>	<i>Abudefduf</i> = 941 Pomacentridae = 15 > Family = 44	<i>Abudefduf</i> = 894 Pomacentridae = 28 > Family = 78	<i>Abudefduf</i> = 819 Pomacentridae = 44 > Family = 137
<b><i>Chanos chanos</i></b>	<i>Chanos chanos</i> = 996 > Family = 2 Other = 1	<i>Chanos chanos</i> = 981 > Family = 18 Other = 1	<i>Chanos chanos</i> = 950 Other = 50
<b><i>Sardina pilchardus</i></b>	<i>Sardina pilchardus</i> = 1000	<i>Sardina pilchardus</i> = 999 Clupeidae = 1	<i>Sardina pilchardus</i> = 996 Clupeidae = 3

<i>Haemulon aurolineatum</i>	<i>Haemulon aurolineatum</i> =884 > Family =115	<i>Haemulon aurolineatum</i> =798 Haemulidae = 4 > Family =4 Other = 194	> Family = 1 <i>Haemulon aurolineatum</i> =713 Haemulidae =6 <b>Sinipercidae= 1</b> Other = 280
------------------------------	--	---	---

#### **S4.8.2 SUPPLEMENTARY REFERENCES**

Grealy A, McDowell M, Scofield P, Murray D, Fusco D, et al. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional palaeontological methods. *Quaternary Science Reviews* **128**: 37-47.

Grealy A, Macken A, Allentoft ME, Rawlence NJ, Reed E, et al. 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *Journal of Quaternary Science* **31**: 33-45.

## 4.9 EPILOGUE

---

Given the results of this study, several more subsamples have been taken from the most diverse layers (namely, layer 1 in units 1 and 2) in order to test whether most of the diversity has been detected. In addition, another primer set targeting a short region of the mitochondrial fish *16S rRNA* gene has been amplified from each extract to independently confirm taxa IDs (which is advocated in Chapter 2). These results will be integrated with morphological analysis of faunal assemblages from other archaeological sites within the Velondriake Marine Protected Area, and will be published by Dr. Kristina Douglass in a paper describing the faunal assemblage recovered from different archaeological sites in Velondriake using multi-analytical approaches. We aim to publish this article in a thematic volume of *Quaternary International* on African archaeozoology by May 2017.

In this chapter, we identified archaeological fish bone from Madagascar using an aDNA bulk bone metabarcoding approach without any prior information about the assemblage derived from morphological analysis. 23 fish families were identified that suggest past people relied on both marine and brackish-water habitats. Some taxa that existed in the past are absent today, which may indicate over-fishing by past people. In this context, the method has wider applications for the fields of molecular ecology, and marine and conservation biology, which could be of interest to local fisheries and governing bodies. This chapter draws together the concepts examined in Chapter 2 and 3, showing that aDNA analysis of bulk bone can provide insights into past biodiversity even without independent confirmation from morphology.

This study also demonstrates that aDNA can be retrieved from fossil bone in tropical coastal sites that have been exposed to the elements since deposition. While the DNA obtained from this site in Madagascar is relatively recent, it is extremely degraded. Consequently, the possibility of retrieving informative aDNA from older fossil bone specimens is small. Other fossil substrates, such as fossil avian eggshell, can also be reservoirs for aDNA, and can be more promising alternatives for retrieving aDNA from extinct animals in warm climates than bone. Most aDNA extracted from fossil eggshell has come from eggs of the extinct moa, which lived in the cool climates of New Zealand. The next chapter (Chapter 5) explores how high-quality aDNA can

also be retrieved from ca. 1000 ya fossil avian eggshell in the warm, tropical environment of Madagascar.

— CHAPTER 5 —

---

**EGGSHELL PALAEOGENOMICS: PALAEOGNATH EVOLUTIONARY  
HISTORY REVEALED THROUGH ANCIENT NUCLEAR *and*  
MITOCHONDRIAL DNA FROM MADAGASCAN ELEPHANT BIRD  
(*AEPYORNIS* SP.) EGG SHELL**

---

*The ratites reached their present separated homelands without the benefit of flight. How did they get there?...They walked. All the way. What we now know as separate continents were jointed together, and the great flightless birds walked...[The ratites] shared ancestor was flightless too.*

- Richard Dawkins  
The Elephant Bird's Tale

*The great tragedy of science—the slaying of a beautiful hypothesis by an ugly fact.*

- TH Huxley

## 5.1 PROLOGUE

---

As illustrated in Chapters 3 and 4, the aDNA retrieved from fossil bone in warm, tropical environments like Madagascar is typically highly degraded and contaminated by exogenous bacterial DNA, making the likelihood of obtaining aDNA from bone samples older than several hundred years slim. Madagascar in particular is a biodiversity hotspot, home to a multitude of vastly different biomes and ecosystems that have come under considerable strain over the past millennia since human colonisation. As such, there has been a significant loss of biodiversity, with many species, especially Madagascar's megafauna (including giant lemurs, tortoises, and elephant birds), becoming extinct. Consequently, little is known about the evolutionary history of many of Madagascar's most unique taxa.

Although aDNA has the potential to shed light on the evolutionary history of Madagascar's lost species, there has been little success in isolating aDNA from fossil bone because many are more than 1000 years old. For elephant birds in particular, very few skeletal specimens have been found, and it was not until recently that a whole mitochondrial genome was recovered from such specimens (Mitchell et al. 2014). Until then, most of what was known about the relationship of elephant birds to other birds was largely assumed or inferred by morphological analysis of incomplete skeletons. Nevertheless, what we now know about the evolutionary history of elephant birds is based only on mitochondrial DNA as nuclear DNA has yet to be recovered, presumably due to poor preservation.

Like bulk bone, fossil avian eggshell can be an alternative source of aDNA that may offer some success for additional palaeogenomic information to be retrieved from early-Holocene elephant birds. However, little more than a few hundred base pairs has been extracted from elephant bird eggshell before now. In this chapter, we apply new shotgun library preparation and targeted enrichment methods to fossil eggshell for the first time in order to recover both a whole mitogenome, as well as nuclear loci, from the elephant bird. The aDNA information retrieved is then used to independently verify their phylogenetic relationship with other birds, and to explore

the prospect of obtaining further palaeogenomic information from elephant bird eggshell in the future.

The work presented in this chapter was largely conducted in 2013 for publication in 2014; however, due to release of a partial mitogenome study in 2014 by Mitchell et al. (*Science*, Vol. 344, Pg. 898-900), the focus of our study necessitated a greater emphasis on retrieving nuclear DNA information, which was not a trivial task. The collection of this extra nuclear data resulted in delayed submission of this manuscript, which has subsequently been published in the peer-reviewed journal *Molecular Phylogenetics and Evolution* in January 2017 (Vol. 109, Pg. 151-163), a facsimile of which can be found in Appendix I. This chapter is a reproduction of the aforementioned manuscript (formatting, including in-text referencing and headings, excepted). In the time since the work described in this chapter was accepted, Yonezawa et al. (2016) published a study in *Current Biology* (doi: 10.1016/j.cub.2016.10.029) reporting nuclear DNA recovery from elephant bird fossil bone. This study recovers the same phylogenetic topology but differs in some of the inferred dates within the palaeognath radiation.

### 5.1.1 ACKNOWLEDGEMENTS

This work was funded by a grant from the Australian Research Council (ARC) grants awarded to JH. (DE120100107). MB was supported in this research by an ARC future fellowship (FT0991741). MTPG was supported by ERC Consolidator Grant (681396-Extinction Genomics). GF wishes to acknowledge the National Geography Society, as well as the Easterbrook Distinguished Scientist Award (from the Division of Quaternary Geology and Geomorphology Division of the Geological Society of America), for funding sample collecting trips in Madagascar. Richard Allcock at Lottery West performed high-throughput sequencing at Royal Perth Hospital, WA. Computing resources at the University of Copenhagen and Pawsey Supercomputer Centre (WA) were used to perform BLAST searches. We thank Guojie Zhang and Erich Jarvis for access to avian genome data, and Leon Huynen for constructive comments on the manuscript.

### 5.1.2 AUTHOR CONTRIBUTIONS

JH and MB conceived and supervised the study. JH and AG designed the experiments. AG performed the experiments. AG and MP analysed the data. J-MR synthesised and supplied tinamou nuclear MYbaits. GM supplied and dated eggshell samples. MTPG and DL provided databases of *D. novaehollandiae*, *R. americana* and *A. mantelli* nuclear sequences. AG drafted the manuscript. All authors contributed to editing the manuscript.

### 5.1.3 AUTHOR DECLARATIONS

The authors declare no competing interests.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work (e.g., journal articles) in which the copyright is held by another party (e.g., publisher, co-author). Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged. Reprinted from Publication *Molecular Phylogenetics and Evolution*, Vol. 109, Grealy AC et al., Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell, Pg. 151-163, Copyright (2017), with permission from Elsevier.

All supplementary data related to this article can be found at in the file “Supplementary Information” published with the online version of the article, as well as section 5.9.

Mitochondrial and nuclear genome sequences for the studied *Aepyornis* sp. specimen have been deposited within GenBank (available [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)) under the accession code KY412176. Data can be accessed through DataDryad at [doi:10.5061/dryad.6h3q7](https://doi.org/10.5061/dryad.6h3q7). Correspondence and requests for materials should be address to AG ([alicia.grealy@uqconnect.edu.au](mailto:alicia.grealy@uqconnect.edu.au)).

**EGGSHELL PALAEOGENOMICS: PALAEOGNATH EVOLUTIONARY HISTORY REVEALED THROUGH ANCIENT NUCLEAR AND MITOCHONDRIAL DNA FROM MADAGASCAN ELEPHANT BIRD (*AEPYORNIS* SP.) EGG SHELL**

—*in*—

***MOLECULAR PHYLOGENETICS AND EVOLUTION* (2017) | VOL. 109 | PG. 151-163**

Alicia Grealy<sup>1\*</sup>, Matthew Phillips<sup>2</sup>, Gifford Miller<sup>1,3</sup>, M Thomas P Gilbert<sup>1,4,5</sup>, Jean-Marie Rouillard, David Lambert<sup>6</sup>, Mike Bunce<sup>1</sup>, & James Haile<sup>7</sup>

<sup>1</sup>Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia.

<sup>2</sup>Vertebrate Evolution Group, School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia

<sup>3</sup>INSTAAR and Geological Sciences, University of Colorado, Boulder, CO 80309, USA

<sup>4</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350, Denmark.

<sup>5</sup>NTNU University Museum, N-7491 Trondheim, Norway

<sup>6</sup>Environmental Futures Research Institute, School of Environment, Griffith University, Brisbane, QLD 4111, Australia

<sup>7</sup>Research Laboratory for Archaeology and the History of Art, Oxford University, Oxford OX12JD, UK.

\* Corresponding author. Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia. *E-mail address*: alicia.grealy@uqconnect.edu.au

**Key words:** ancient DNA, biogeography, eggshell, high-throughput sequencing, Palaeognath, phylogeny

## 5.2 ABSTRACT

---

Palaeognaths, the sister group of all other living birds (neognaths), were once considered to be vicariant relics from the breakup of the Gondwanan supercontinent. However, recent molecular studies instead argue for dispersal of volant ancestors across marine barriers. Resolving this debate hinges upon accurately reconstructing their evolutionary relationships and dating their divergences, which often relies on phylogenetic information from extinct relatives and nuclear genomes. Mitogenomes from the extinct elephant birds of Madagascar have helped inform the palaeognath phylogeny; however, nuclear information has remained unavailable. Here, we use ancient DNA (aDNA) extracted from fossil eggshell, together with target enrichment and next-generation sequencing techniques, to reconstruct an additional new mitogenome from *Aepyornis* sp. with 33.5X coverage. We also recover the first elephant bird nuclear aDNA, represented by 12,500 bp of exonic information. While we confirm that elephant birds are sister taxa to the kiwi, our data suggests that, like neognaths, palaeognaths underwent an explosive radiation between 69-52 Ma—well after the break-up of Gondwana, and more rapidly than previously estimated from mitochondrial data alone. These results further support the idea that ratites primarily diversified immediately following the Cretaceous-Palaeogene mass extinction and convergently evolved flightlessness. Our study reinforces the importance of including information from the nuclear genome of extinct taxa for recovering deep evolutionary relationships. Furthermore, with approximately 3% endogenous aDNA retrieved, avian eggshell can be a valuable substrate for recovering high quality aDNA. We suggest that elephant bird whole genome recovery is ultimately achievable, and will provide future insights into the evolution these birds.

### 5.3 INTRODUCTION

---

**P**ALAEOGNATHS (including the volant tinamous and the flightless ratites) are distinguished from other birds by the retention of an ‘ancient palate’ (Cracraft 1974). As such, the Palaeognathae are ubiquitously recognised as the sister group of all other living birds (neognaths). The common ancestor of palaeognaths split from neognaths during the middle of the Cretaceous period (Clarke et al. 2005), after which neognaths underwent an explosive radiation around the Cretaceous-Palaeogene boundary ca. 65 Ma (Jarvis et al. 2014; Claramunt and Cracraft 2015; Prum et al. 2015); however, the subsequent diversification of palaeognaths remains poorly understood. Ratites have historically been seen as archetypal relics from the breakup of the Gondwanan supercontinent (Cracraft 1973), where populations of a flightless ancestor ‘rafted’ with the drifting continents and diverged from one another in isolation to give rise to rheas (*Rhea* spp.) in South America, the ostrich (*Struthio* spp.) in Africa, emus (*Dromaius* spp.) in Australia, cassowaries (*Casuaris* spp.) in Australia and New Guinea, kiwi (*Apteryx* spp.) and moa (*Dinornis* spp., *Anomalopteryx* sp., *Emeus* sp., *Euryapteryx* sp., *Pachyornis* spp., *Megalapteryx* sp.) in New Zealand, and elephant birds (*Aepyornis* spp., *Mullerornis* spp.) in Madagascar. As such, ratites were assumed to form a monophyletic group, sister to tinamous (Cracraft 1974; Cooper et al. 2001). While aspects of this phylogeny gain some support from morphological studies (Johnston 2011; Worthy and Scofield 2012), morphology has been shown to be poor at resolving evolutionary relationships from characters that are functionally correlated (e.g., Springer and O’Brien 2007). Morphological characters such as flightlessness could be interpreted equally as synapomorphies (inherited from a common ancestor) or homoplasies (arising independently via convergent evolution), confounding phylogenetic inference (Hackett et al. 2008; Harshman et al. 2008; Allentoft and Rawlence 2012; Baker et al. 2014).

In contrast, DNA has been used to resolve the evolutionary relationships of many organisms where morphology has been inconclusive (e.g., Bunce et al. 2003, Krause et al. 2010). Ratite monophyly has been challenged by the inclusion of DNA sequence information into the phylogeny (Harshman et al. 2008), particularly

with the inclusion of ancient DNA (aDNA) sequence information from the extinct moa. The discovery that the volant South American tinamous (*Tinamus* spp., *Nothocercus* spp., *Crypturellus* spp., *Rhynchotus* spp., *Nothoprocta* spp., *Nothura* spp., *Taoniscus* spp., *Eudromia* spp., *Tinamotis* spp.) are nested within ratites as the sister taxon of moa, and that ostriches are sister to all other palaeognaths, suggests that flightlessness evolved several times among ratites (Harshman et al. 2008; Phillips et al. 2010; Smith et al. 2013; Haddrath and Baker 2012; Baker et al. 2014). Similarly, mitochondrial aDNA places Madagascar's extinct giant elephant birds as sister to New Zealand's kiwi (Phillips et al. 2010; Mitchell et al. 2014), which further supports the idea of multiple losses of flight among ratites because the timing of their divergence suggests that vicariance is unlikely to explain their biogeographic distribution. Consequently, there is renewed skepticism of the extent to which vicariance played a role in palaeognath evolution because their distribution could alternatively be explained by dispersal via flying. Resolving this debate hinges on accurately reconstructing the evolutionary relationships and timing of divergence within this group. Reconstructions of palaeognath evolutionary history have further been mired by uncertainty in the timing of the Gondwanan break-up; controversy surrounding the dating and classification of avian fossil specimens (Ksepka et al. 2014); and methods used to construct phylogenies (Ksepka and Phillips 2015), including choice of characters (mitochondrial DNA vs. nuclear DNA; Mitchell et al. 2014), choice of outgroup (Johnston 2011), choice of analytical method (Johnston 2011), choice of molecular clock calibration (Phillips et al. 2010), and taxon sampling (Cooper et al. 2001; Phillips et al. 2010; Mitchell et al. 2014), specifically the underrepresentation of extinct taxa in phylogenetic reconstructions.

Most studies that do include DNA from extinct taxa in phylogenetic analyses rely on mitochondrial aDNA alone as ancient nuclear DNA is typically more challenging to recover. The likelihood of retrieving endogenous aDNA with high coverage from the multi-copy mitochondrial genome is many-fold greater than the likelihood of retrieving endogenous aDNA from a diploid nuclear genome. Nuclear DNA has also been shown to degrade faster than mitochondrial DNA (Allentoft et al. 2012), which may be protected from microbial decay due to its stable circular structure and the double membrane of the surrounding organelle. Nevertheless,

nuclear information often proves crucial to accurately recover deep evolutionary relationships (Hackett et al. 2008): mitochondrial DNA typically evolves over an order of magnitude faster than nuclear DNA, and the resulting saturation tends to erode phylogenetic signal at deeper divergences in the tree and leaves both phylogeny and branch-length inference more susceptible to biases, such as from compositional non-stationarity (Ksepka and Phillips 2015). In addition, maternally inherited mitochondrial gene histories may not represent species evolution. For example, many groupings based on both genomic sequences and retroposons of neognaths (Jarvis et al. 2014; Prum et al. 2015) differ from earlier mitochondrial DNA-based phylogenetic studies, demonstrating that nuclear DNA is often needed to reconstruct accurate phylogenies. As emphasised by Haddrath and Baker (2012), erroneous phylogenetic topologies will result in misleading interpretations of evolutionary history and biogeography.

While nuclear DNA from moa has been retrieved, nuclear DNA sequences have not been obtained from elephant birds, presumably due to the rarity of skeletal remains, particularly those with well-preserved DNA. In addition, DNA preservation in bone is typically poor in hot, tropical climates (Höss et al. 1996) and non-endogenous DNA levels in bone are often high compared with other substrates (Gilbert et al. 2007; Ramírez et al. 2009; Schroeder et al. 2015). As such, it remains to be seen whether the relationships among elephant birds and other palaeognaths can be consistently recovered with the inclusion of elephant bird nuclear aDNA as many avian relationships based on mtDNA have been overturned by nuclear DNA analyses (e.g., Hackett et al. 2008). Because Madagascar has been isolated for over 85 Ma (McLoughlin 2001; Yoder and Nowak 2006), accurate placement of elephant birds within the palaeognath phylogeny using aDNA has the potential to better inform their evolutionary history as it offers a unique opportunity to test biogeographical hypotheses regarding the roles of vicariance and dispersal in ratite evolution (Phillips et al. 2010; Mitchell et al. 2014); if the lineage leading to elephant birds arose after Madagascar was isolated, a vicariance model of evolution could not explain their presence there.

As an alternative to bone, eggshell has been shown to provide a good source of high-quality, endogenous aDNA from extinct birds, such as moa (Oskam et al.

2010; Allentoft et al. 2011; Oskam et al. 2011). This may be because eggshell is water resistant and therefore DNA in eggshell is better protected from microbial decay and hydrolytic damage than the DNA in porous bone (Oskam et al. 2010). To date, there has been some success with recovering aDNA from elephant bird eggshell (Oskam et al. 2010). In addition, elephant bird eggshell is highly robust and abundant; certain beaches are “nearly paved with broken eggshells” (Dewar 1984), with many buried within dune systems that might offer additional protection of DNA from the environment. Thus, eggshell is the most promising substrate for the retrieval of both mitochondrial and nuclear elephant bird DNA without the destructive sampling of scarce and valuable skeletal specimens. Coupled with recent advances in aDNA methods (Gansauge and Meyer 2013; Li 2013) and high-throughput sequencing, we demonstrate how aDNA can be extracted from elephant bird eggshell to independently reconstruct the whole mitochondrial genome of *Aepyornis* sp., as well as capture nuclear loci for the first time. In doing so, we aim to confirm the relationship between elephant birds and kiwi and more accurately date the divergences among the palaeognaths, in order to clarify the processes that shaped their evolution.

## **5.4 MATERIALS *and* METHODS**

---

### **5.4.1 SITE DESCRIPTION *and* COLLECTION *of* EGGSHELL SAMPLES**

Eggshell samples screened for elephant bird aDNA (13; q.v., section 5.4.3-5.4.4) were collected from various coastal sites in Southern Madagascar (data available upon request). The sample chosen for further genomic analysis, sample M06-M028 (Figure 5.1a and b), was excavated from a rapidly deflating coastal dune system on the southern coast of Madagascar at 25.46250°S, 45.70083°E (Talaky). Eggshell was buried beneath 5–10 cm of aeolian sand; however, the long-term depth of burial is estimated to have been over a meter. The tight concentration and lack of any abrasion along the edges of eggshell fragments in this deposit suggest they were continuously buried (no exposure to the surface) since their initial burial.

#### **5.4.2 DATING *and* MORPHOLOGICAL IDENTIFICATION *of* EGGSHELL**

Genus identification of eggshell in the accumulation described above (section 5.4.1) was based on the thickness; fragments ranged from 3.6 to 3.9 mm thick, and were identified as *Aepyornis* sp. as there is a clear distinction in the thickness of eggshell between elephant bird genera (Tovondrafale 2014). The age of specimen M06-M028 was estimated to be 1100 years BP, based on <sup>14</sup>C dating (S5.9.1).

#### **5.4.3 aDNA EXTRACTION**

Genomic DNA was extracted from 200 mg of powdered *Aepyornis* sp. eggshell (13 samples) using a previously optimised protocol with minor changes (Shapiro and Hofreiter 2012; S5.9.2). Sample preparation and DNA extraction were performed in a designated ultra-clean facility at Murdoch University, WA, Australia using sterile reagents and equipment in keeping with standard aDNA practice (Willerslev and Cooper 2005). All post-PCR methods were performed in a physically separated laboratory. Extraction and subsequent sequencing of sample M06-M028 was replicated in a separate ultra-clean facility at Curtin University, WA, Australia for authentication; however, a different extraction method was employed with minor changes (Dabney et al. 2013; S5.9.2). Extraction and PCR negative controls were included and carried through to sequencing.

#### **5.4.4 SCREENING SAMPLES *via* AMPLICON SEQUENCING**

Samples were screened for endogenous aDNA by qPCR-amplifying and sequencing in parallel a 53 bp avian metabarcoding region of the mitochondrial *12S rRNA* gene, with unique 5' and 3' indexes flanking the insert (S5.9.3). Sequencing was carried out using the Ion Torrent PGM (*Life Technologies*) 200 bp kit V2 on a 314 chip following the manufacturer's instructions. The sequences were trimmed, quality filtered, abundance filtered, dereplicated (S5.9.4), and were aligned to the NCBI's GenBank nucleotide database (Benson et al. 2006) using BLAST v.2.2.30+ (Altschul et al. 1990, default parameters) to confirm that the DNA was most similar to published reference sequences of *Aepyornis* and free from avian contaminants. Of all the samples that contained amplifiable elephant bird aDNA (seven), the

sample exhibiting the lowest relative qPCR cycle-threshold value (and therefore, the sample with the most endogenous DNA, sample M06-M028) was selected for further genomic analyses.

#### **5.4.5 SHOTGUN SEQUENCING**

Four single-stranded shotgun libraries (two per extract) were prepared independently in ultra-clean facilities at both Murdoch University and Curtin University by following Gansauge and Meyer (2013) with minor changes (S5.9.5). Extraction and negative control libraries were carried through to sequencing in order to assess potential contamination. Libraries were double-indexed and sequenced from one end on both the Ion Torrent PGM (*Life Technologies*; 200 v.2 kit) and MiSeq (*Illumina*; 150 v.3 kit) platforms, with the best library further sequenced on the high-throughput sequencing platform, Ion Torrent Proton (*Life Technologies*; 200 v.2 kit) at Lottery West, Royal Perth Hospital, following the manufacturer's instructions.

#### **5.4.6 TARGETED ENRICHMENT *through* HYBRIDISATION CAPTURE**

A relaxed hybridisation capture was used to enrich the two best libraries for avian mitochondrial and nuclear sequences by following Li (2013) and the *MYcroarray* MYbaits Sequence Enrichment for Targeted Sequencing kit protocol, with minor changes (S5.9.6). Whole genome baits (MYbaits, *MYcroarray*) were synthesised by globally transcribing tinamou genomic DNA (*Nothocercus bonapartei*) into RNA baits using biotinylated rUTP (S5.9.6), and 100mer mitochondrial baits with 50 bp tiling were designed using the reference mitochondrial genomes of several avian taxa (Table S5.9.4). Enriched libraries were sequenced on the MiSeq (*Illumina*) platform as above.

#### **5.4.7 QUALITY CONTROL**

Reads were sorted by index and trimmed in Geneious v.7.1.2 ([www.geneious.com](http://www.geneious.com); Kearse 2012), allowing no mismatches in the expected sequence of the index or library adapters. FASTQ files were exported and uploaded on to the web-based

bioinformatics platform Galaxy (usegalaxy.org; Giardine et al. 2005; Blankenberg et al. 2010; Goecks et al. 2010) for quality filtering following Grealy et al. (2016) with minor changes (S5.9.4): sequences were required to have an average quality score of at least Q25, with all bases above Q10, 95% of bases above Q15, and 90% of bases above Q20. Sequences were then combined and dereplicated such that only unique sequences remained. Reads below 30 bp in length were discarded. To conservatively estimate the proportion of endogenous DNA extracted from *Aepyornis* eggshell relative to bacterial and contaminating DNA, unenriched (shotgun) sequences were assigned taxonomy using BLAST and MEGAN v.4.70.4 (Huson et al. 2007; S.5.9.7).

#### **5.4.8 MITOCHONDRIAL GENOME ASSEMBLY**

Quality filtered, unique sequences were iteratively mapped against a ratite consensus mitochondrial reference genome in Geneious (i.e., a consensus derived from aligning kiwi, emu, cassowary, ostrich, rhea, moa, and tinamou mitochondrial genomes; S5.9.7). Mapped reads were then aligned with BLAST to a custom aggregate database containing both the GenBank nucleotide database and unpublished *Struthio camelus* whole genome contigs (S5.9.7). Taxonomic assignments were examined in MEGAN. Sequences that were assigned to class Aves were extracted and re-mapped onto the consensus mitochondrial genome from the previous round of mapping in order to ensure that no contaminating sequences were mapped. Mapped tRNA-Phe was used as a seed to complete assembly of the control region in MITObim v.1.8 (Hahn et al., 2013), using all unique sequences as the read pool. Bases with coverage greater than two and matching at least 50% of sequences were called as the consensus as per Dabney et al. (2013), in order to generate a conservative contiguous reconstruction of the elephant bird mitochondrial genome. aDNA damage patterns were assessed using mapDamage v.2.0 (Jonsson et al. 2013) to authenticate the antiquity of sequences (S5.9.8; Figure S5.9.2).

#### 5.4.9 IDENTIFICATION OF NUCLEAR SEQUENCES

In order to identify orthologous, phylogenetically informative nuclear loci, unique reads were mapped, in Geneious, to a custom database of ostrich (*S. camelus*) reference sequences containing 44 nuclear protein-coding exons from Baker et al. (2014), 67 nuclear protein-coding exons from Haddrath and Baker (2012), Chojnowski et al. (2007), Hackett et al. (2008), Harshman et al. (2008), and Smith et al. (2013), and 7976 nuclear protein-coding exons from Jarvis et al. (2014) and Zhang et al. (2014) (S5.9.9). These genes are known to be single-copy and have been used in prior avian phylogenetic studies. With relatively short sequence reads and limited coverage, any non-coding data that is substantially variable also tends to provide far less certainty about orthology; rather, single-copy, protein coding genes allow frame-checking, as well as a mix of conserved and variable sites that balances homology considerations with signal considerations. Hits were accepted as putative orthologues only if: (1) the read was 80 bp or longer and aligned across the entire length of the read; (2) the percent similarity for the alignment with ostrich was higher than the similarity between ostrich and chicken for the same locus; and (3) reciprocal comparison of the ostrich genome to elephant bird sequence set identified the same read as its closest match (Vallender 2009). The minimum coverage accepted per base was one as there are far fewer copies of nuclear loci than mitochondrial loci; this is not unusual for aDNA (Rohland et al. 2010). The ostrich orthologues were then mapped against tinamou (*T. major*/*T. guttatus*; Jarvis et al. 2014), kiwi (*A. apteryx mantelli*; Le Duc et al. 2015), and unpublished emu (*D. novaehollandiae*), and rhea (*R. americana*) whole genome contigs (generated as part of the B10k avian genome sequencing project; Zhang et al. 2015) to find the corresponding loci in these ratites using the same method. These sequences were also reciprocally mapped to the elephant bird data set to further confirm orthology (S5.9.9).

#### 5.4.10 ALIGNMENTS *and* PARTITIONING

Each locus was aligned in Geneious with the corresponding locus from 16 other palaeognaths (where available), and 11 neognath outgroups using the default parameters (S5.9.10). Alignments were concatenated and translated in MUSCLE

(Edgar 2004) to find ORFs and refine the alignments. The multiple sequence alignment was then imported into Se-AL v.2.0a (Rambaut 1996) for manual checking. Based on published partitioning schemes for palaeognaths (e.g., Phillips et al. 2010), concatenated mitochondrial protein-coding genes were partitioned into first, second, and third codon positions, while RNA-coding genes were partitioned into stem and loop sites, resulting in a total of five mitochondrial partitions (m1, m2, m3, stems, loops; S5.9.10). Due to differing taxonomic composition, nuclear data were partitioned by source as well as by codon position, with loci derived from Jarvis et al. (2014) and Zhang et al. (2014) concatenated and separated from loci derived from Haddrath and Baker (2012) and Baker et al. (2014), resulting in a total of six nuclear partitions (n1, n2, n3 and hb1, hb2, hb3, respectively; S5.9.10).

#### **5.4.11 PHYLOGENETIC ANALYSES**

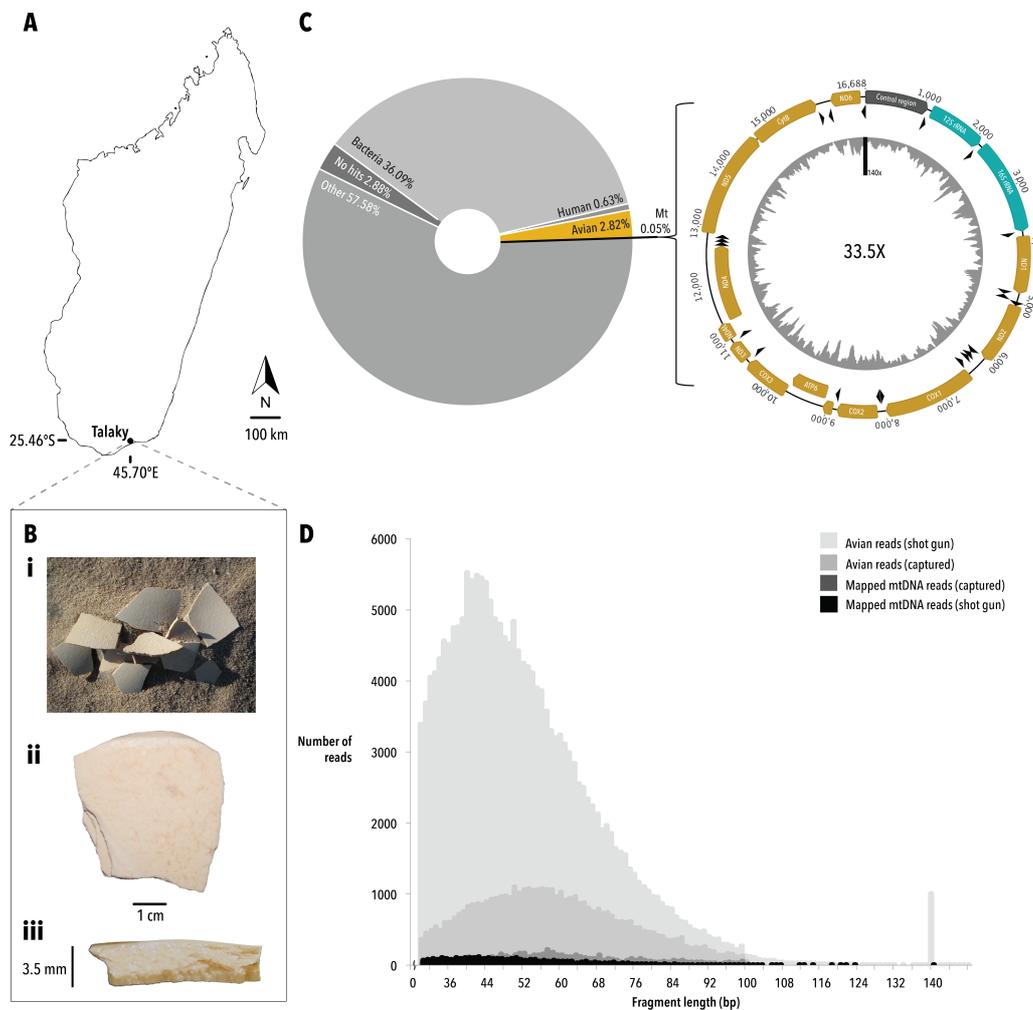
Relative composition variability (RCV) and stemminess tests (Phillips and Pratt 2008) were performed in PAUP v.4.0b (Swofford 2003) to assess base composition bias and the extent of phylogenetic signal erosion in order to determine which partitions may benefit from RY coding that will alleviate the biases (S5.9.11). Mitochondrial third codon positions had both the highest compositional heterogeneity across taxa (Table S5.9.13, Figure S5.9.4) and the lowest uncorrected stemminess (greatest phylogenetic signal erosion) and hence, were RY-coded; however, we also considered how an additional two extreme mitochondrial coding schemes impact phylogenetic support by performing analyses when (1) all mitochondrial partitions coded by standard nucleotides, and (2) first and second mitochondrial codon positions, as well as RNA stem and loop partitions, all RY-coded, with mitochondrial third codons excluded (S5.9.11).

jModelTest v.2.1.7 (Guindon 2003; Durrin 2012) was used to test the best-fitting substitution model for each partition (Table S5.9.16). Mitochondrial, nuclear and total evidence phylogenetic trees were constructed using both a maximum likelihood and Bayesian approach for each coding scheme. Phylogeny reconstruction was performed in RAxML v.1.5 (with 500 bootstrap replicates; Stamatakis 2014; Silvestro and Michalak 2012) and MrBayes v.3.2.6 (Huelsenbeck and Ronquist 2001; S5.9.11) implemented through the CIPRES online

bioinformatics pipeline (Miller 2010). Maximum likelihood analyses were run under a GTR+I+ $\Gamma$  model with 500 bootstrap replicates (S5.9.11). Bayesian analyses were run twice after a burn-in of 0.2 for 10 million generations sampling every 2500 generations and employing three chains (S5.9.11); m1, m2, stems, loops, and n1 and n2 partitions were run under a GTR+I+ $\Gamma$  model, while RY-coded mitochondrial third codon positions (m3) were run under an CF87+I+ $\Gamma$  model, with n3 and hb1 run under a GTR+ $\Gamma$  model, and hb2 and hb3 run under a HKY+ $\Gamma$  model (Table S5.9.16). Tracer v1.6.1 was used to examine the convergence of Bayesian runs (Rambaut 2003; S5.9.11).

#### 5.4.12 MOLECULAR DATING

Molecular dating was performed using MCMCtree (Yang 2006) within PAML v.4.8 (Yang 2007) on the mitochondrial, nuclear, and combined data with both independent and autocorrelated models of rate variation across the respective tree topologies described above (Figure 5.2; S5.9.12). In each case, seven fossil-based age priors were used for calibration: avian root 66.5-124.1 Ma, Galloanserae node >66.5<83.8 Ma, Penguin/tube-nose node >60.5<72.3 Ma, Core land/ water birds node >60.5<72.3 Ma, non-ostrich palaeognaths node >56.0<72.3 Ma, Parrots/ Passeriformes node >53.5 Ma, Emu/kiwi node >24.5 Ma (S5.9.12, Table S5.9.18). For each analysis the GTR+ $\Gamma_4$  substitution model was employed (and collapsed to effectively 2-state F81+ $\Gamma_4$  for RY partitions). Default parameters were generally used, although sigma2\_gamma scale (=1) reflects the weak prior of 100 Ma for the avian root (cf., Jarvis et al. 2014; Ksepka and Phillips 2015), and rgene\_gamma scale (=2) is derived as the inverse of the approximately 1/2 substitutions per 100 Ma averaged across the data (S5.9.12).



**FIGURE 5.1** | DETAILS OF THE EGGSHELL SPECIMEN **A** Map of Madagascar showing the collection site of the eggshell sample; **B** photograph of *Aepyornis* sp. eggshell specimen M06-M028 showing it **i** as collected (not to scale), **ii** as viewed from above, and **iii** in cross-section; **C** annotated schematic of the mitochondrial genome of *Aepyornis* sp.: grey peaks represent the coverage across the genome with the average coverage depicted, and a pie chart displaying the percentage of endogenous (avian) DNA versus exogenous DNA from shotgun sequencing is shown left; **D** fragment length distribution of total avian and mapped mitochondrial reads obtained from both shotgun and capture enrichment.

## 5.5 RESULTS

---

### 5.5.1 SAMPLE SCREENING VIA AMPLICON SEQUENCING

Seven out of 13 eggshells (approximately 54%) screened for endogenous aDNA via sequencing of a *12S rRNA* barcoding region yielded elephant bird aDNA. For the most promising sample, M06-M028, 29 unique sequences were generated post quality and abundance filtering, all of which aligned closest with the published *Aepyornis hildebrandti* (KJ749824) and *Mullerornis agilis* (KJ749825) *12S rRNA* gene sequences, suggesting that the sample contained endogenous DNA and was largely free of contamination by other birds.

### 5.5.2 SHOTGUN SEQUENCING

Shotgun sequencing of aDNA extracted from the eggshell sample above (S5.9.5) yielded a total of 6,628,956 quality filtered, unique sequences above 30 bp, with 97.2% of reads having an average quality above Q20, and 52.5% of reads having an average quality above Q30 (Table S5.9.8). 2.82% (187,044; Table S5.9.7) of reads were taxonomically assigned to class Aves (Figure 5.1c), with about one third assigned to bacterial taxa (36.09%, Figure 5.1c), and less than one percent (0.63%, Figure 5.1c) assigned to human. The mean length of avian shotgun reads was 50.8 bp (St Dev 17.4 bp; Figure 5.1d). Of these avian reads (where over 80% assigned to Palaeognathae; S5.9.7), 1.87% (3,505; 0.05% of the total reads) mapped onto a ratite consensus reference mitochondrial genome (Table S5.9.7). Mitochondrial mapped reads had a mean length of 49.3 bp (St Dev 15.40 bp), with 97.6% of reads having an average quality over Q20, and 64.3% of reads having an average quality over Q30. 34,706 sequences were obtained from the extraction and negative control libraries, seven of which mapped to the ratite mitochondrial reference genome. However, only one of these sequences had an identical counterpart in the dataset, and this sequence was removed.

### 5.5.3 BAIT ENRICHMENT

Together, enrichment of shotgun libraries by hybridisation capture with avian mitochondrial and tinamou genomic baits (S5.9.6) yielded a further 557,074 quality filtered unique sequences above 30 bp with a mean length of 50.7 bp (St Dev 15.6 bp), 99.8% of sequences having an average quality above Q20, and 98.7% of sequences having an average quality above Q30. These sequences were composed of 8.12% (45,236; Table S5.9.7) avian reads, while 1.08% of the total captured reads (6,021; Table S5.9.7) mapped onto the ratite consensus mitochondrial reference genome, with an average length of 60.2 bp (St Dev 16.3 bp; Figure 5.1d). Although fewer reads were obtained by capture compared with direct shotgun sequencing, nearly four times as many were target (avian) DNA, with an approximately 20-fold increase in mitochondrial DNA reads. Furthermore, a greater proportion of captured avian sequences were long (i.e., above 60 bp) in comparison to non-captured reads (Figure 5.1d).

### 5.5.4 MITOCHONDRIAL GENOME RECONSTRUCTION *and* NUCLEAR MAPPING

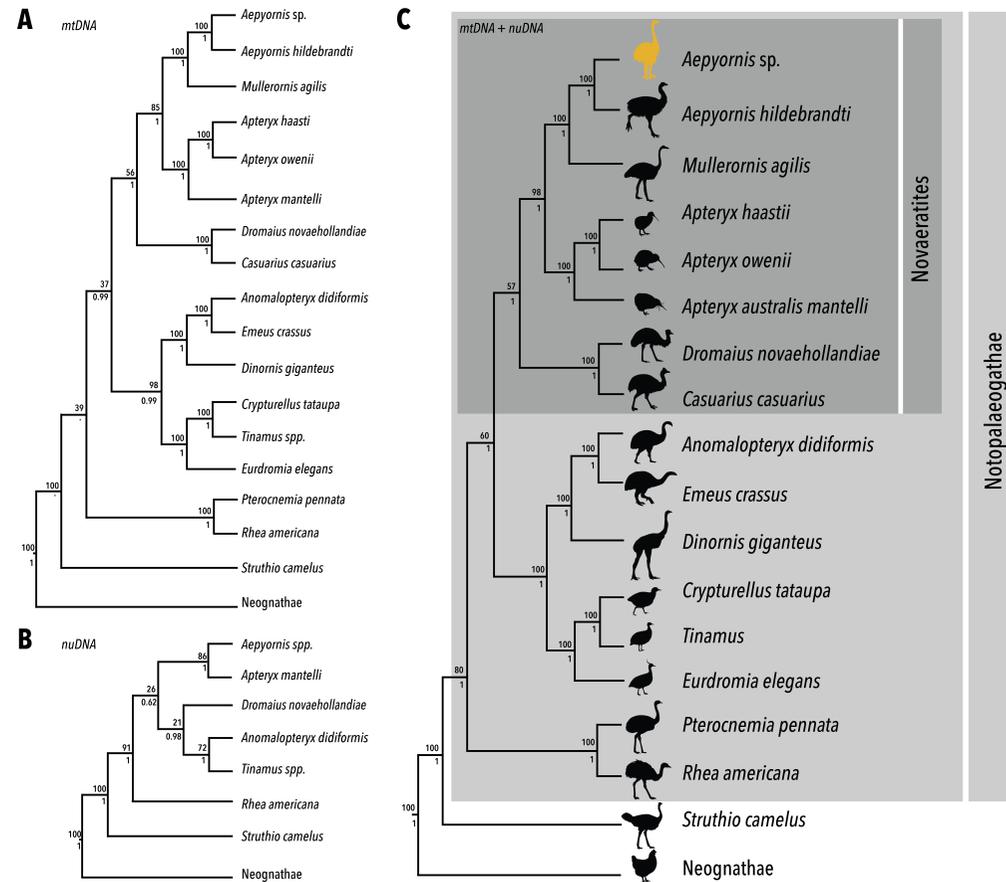
In total (combined shotgun and bait-enriched datasets), 9,478 unique avian reads above 30 bp mapped onto a ratite consensus mitochondrial reference genome (S5.9.7) with 641 sequences assembled across the control region (D-loop). This allowed us to independently reconstruct 16,668 bp of the *Aepyornis* sp. mitochondrial genome with an average coverage of 33.5X (Figure 5.1c). Regions with lower coverage largely correspond to the hyper-variable control region, although 1,077 bp of this region was recovered. This is consistent with the length of the control region in other ratites, which ranges from 1,034 bp in ostrich (*Struthio camelus*) to 1,362 bp in kiwi (*Apteryx owenii*). The characteristic pattern of a greater proportion of C-to-T misincorporations at both ends of the read that is expected from libraries built from single stranded aDNA (Schroeder et al. 2015), was observed in the mitochondrial mapped reads and suggests that the DNA is likely to be of ancient origin (S5.9.8, Figure S5.9.2). The influence of aDNA damage and errors on the final consensus sequence is likely to be minimal given the high coverage obtained, as well as the use of both *Life Technologies* and *Illumina* platforms to sequence several independent libraries.

In addition to a complete mitochondrial genome, we retrieved sequence information from 154 protein-coding regions (Table S5.9.10) of *Aepyornis* sp. that totaled 12,519 bp of phylogenetically informative nuclear loci with an average coverage of approximately 2.36X (with 52% of bases having a coverage of 1X and the remainder having a coverage of 2X or more), and average quality of Q24.8 (S5.9.9).

### 5.5.5 PHYLOGENY RECONSTRUCTION

Relative composition variability (RCV) and stemminess tests (Phillips 2008) respectively indicated base composition bias and phylogenetic signal erosion in the mitochondrial third codon positions that could lead to inaccurate phylogeny reconstruction (S5.9.11). RY-coding the mitochondrial third codon positions saw RCV fall from 0.1 to 0.04 (Table S5.9.11), while stemminess increased from 0.17 to 0.25 (Table S5.9.14), suggesting that there is both improved base compositional homogeneity and less substitution saturation when mitochondrial third codon positions are RY-coded. Nevertheless, we additionally performed phylogenetic analyses with no mitochondrial RY-coding (standard coding), as well as full mitochondrial RY-coding with mitochondrial third codon positions excluded entirely. Regardless of the level of mitochondrial RY-coding, the relationships among the neognath outgroups largely agree with extant avian phylogenies derived from genomic analyses (Jarvis et al. 2014; S5.9.11, Table S5.9.17); in particular, the outgroup relationships inferred from the combined mitochondrial and nuclear analyses fully agree with those of Jarvis et al. (2014), which adds credibility to the topology of the palaeognath phylogeny that we determined hereafter.

We confirm, with high statistical support, that elephant birds are sister to kiwi (Figure 5.2). Mitochondrial and nuclear DNA support this relationship, both independently and together: maximum likelihood bootstrap estimates range between 86% and 100%, while posterior probabilities are all 1, regardless of whether the mitochondrial third codon positions were RY-coded or not (Figure 5.2; Table S5.9.17). However, support decreases as more mitochondrial positions are



**FIGURE 5.2** | MITOCHONDRIAL, NUCLEAR, AND COMBINED PALAEOGNATH PHYLOGENETIC TREES **A** mitochondrial (with third codon positions RY-coded; where a node is not supported by Bayesian analyses, an ‘-’ symbol is displayed—note these unsupported nodes are favoured under ML and Bayesian inference with the most conservative coding scheme that has third codon positions excluded and all other sites RY-coded); **B** nuclear; and **C** combined mitochondrial (with third codon positions RY-coded) and nuclear palaeognath phylogenetic trees. Maximum Likelihood bootstrap and Bayesian posterior probability estimates are displayed above each node. Neognath outgroup relationships are not shown for simplicity (see Table S5.9.17).

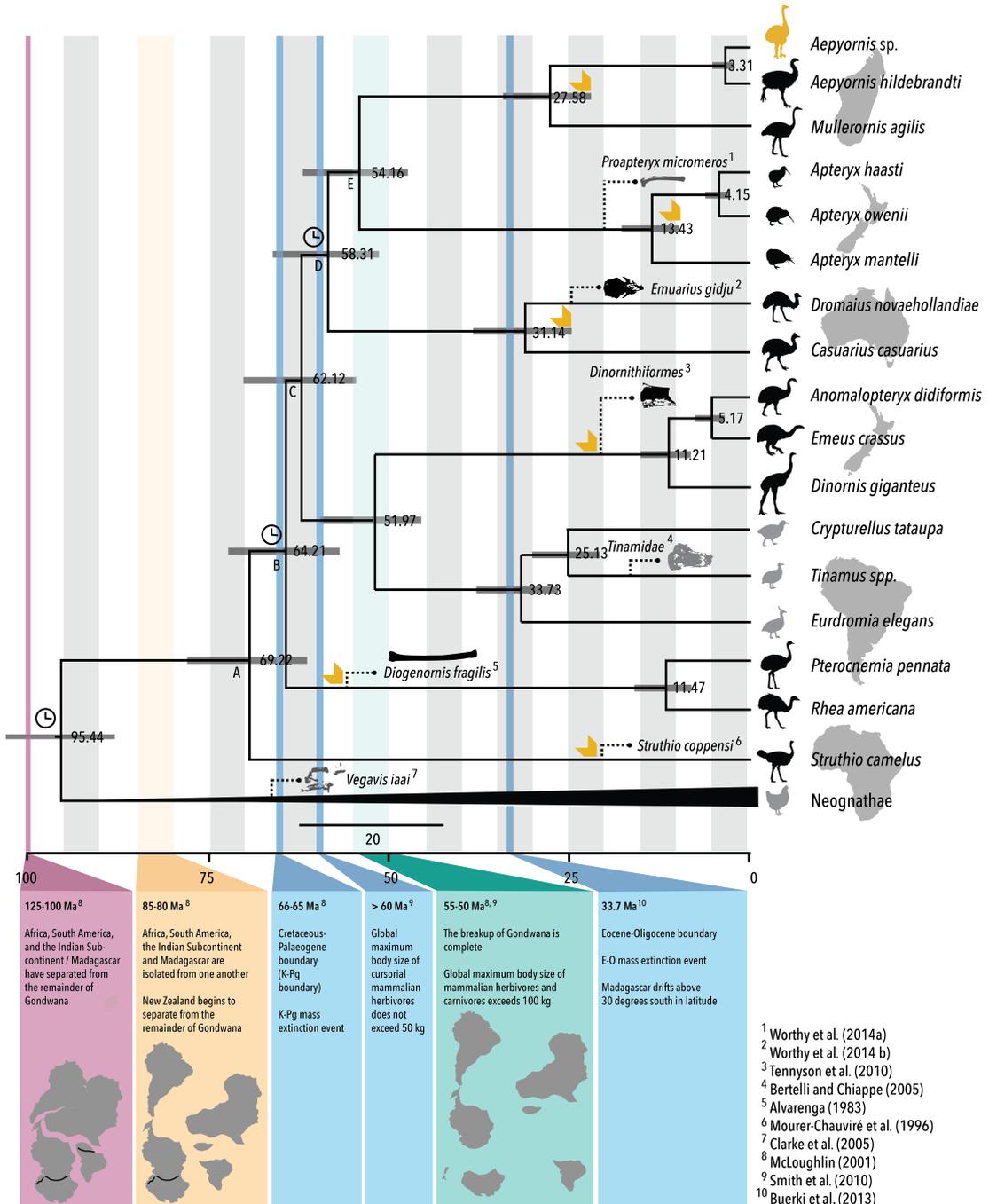
RY-coded, most likely due to loss of phylogenetic signal: when all mitochondrial positions are RY-coded with third codon positions excluded, bootstrap support decreases to 51% (Table S5.9.17) and posterior probability decreases to 0.93. It should be noted that this is not the case in the combined dataset; the addition of nuclear information restores support for this grouping, even with the highest level of mitochondrial RY-coding (i.e., full mitochondrial RY-coding with third codon positions excluded entirely; Table S5.9.17).

While most analyses placed ostrich as sister to all other palaeognaths with high support (Figure 5.2; Table S5.9.17), the monophyly of the “Notopalaeognathae”—all palaeognaths to the exclusion of ostrich—is best supported by the nuclear data (Figure 5.2b), with a combination of both nuclear and mitochondrial DNA resulting in 80% bootstrap support for this node (Figure 5.2c). Although Notopalaeognathae monophyly was unsupported by standard nucleotide mitochondrial data, fully RY-coding the mitochondrial genome increased bootstrap support for this grouping, up to 98% in combination with the nuclear data (Table S5.9.17). Bayesian analyses also show high support for this topology, but only when nuclear data is included (adding nuclear data increases support for every mitochondrial RY-coded grouping that was not already at 100%; Figure 5.2), or with moderate to high levels of mitochondrial RY-coding (Table S5.9.17). In fact, mitochondrial incongruence with the nuclear phylogeny is largely eliminated by RY-coding; ‘approximately unbiased’ (AU) testing on the nuclear data finds significant incongruence with the standard nucleotide mitochondrial topology ( $p=0.034$ ), but not with the RY-coded mitochondrial topology ( $p=0.589$ , S5.9.11).

The branching order of the emu/cassowary, rhea, and moa/tinamou clades within Notopalaeognathae are less robust. The clade grouping moa/tinamous with the “Novaeratitae” to the exclusion of the rheas does not garner high bootstrap support from either mitochondrial or nuclear analyses (Figure 5.2a and b); rather, this relationship gains the most support from a combination of both nuclear and mitochondrial DNA (Figure 5.2c). Similarly, bootstrap support for novaeratite monophyly (i.e., Casuariiformes sister to the elephant birds/kiwi) to the exclusion of moa/tinamous is highest when both mitochondrial and nuclear data are combined (Figure 5.2c). Both bootstrap support and posterior probability estimates for this

**TABLE 5.1** | SUMMARY OF PALAEOGNATH DIVERGENCE DATES. Divergence dates and their uncertainty for palaeognath orders estimated through molecular clock dating using mitochondrial data (\*with third codon positions RY-coded), nuclear data, and combined mitochondrial and nuclear data. Nodes labeled A-E correspond to those on Figure 5.3.

Node	Divergence time (Ma) $\pm$ 95% HPD		
	Mitochondrial*	Nuclear	Mitochondrial* + Nuclear
<b>A</b> Palaeognathae	69.0 (77.2-60.5)	72.1 (82.1-62.4)	69.2 (78.0-61.3)
<b>B</b> Notopalaeognathae	65.0 (72.3-57.2)	64.9 (72.3-56.2)	64.2 (72.2-56.8)
<b>C</b> Moa/tinamou/Novaeratites	62.6 (70.0-55.1)	62.6 (70.2-53.8)	62.1 (70.0-55.0)
<b>D</b> Casuariidae/elephant bird/kiwi	58.1 (65.7-51.1)		58.3 (66.0-51.3)
OR			
Casuariidae/moa/tinamou		60.0 (67.8-50.9)	
<b>E</b> Elephant bird/Kiwi	53.5 (61.0-46.6)	59.2 (66.6-49.8)	54.2 (61.7-47.5)
Moa/tinamou	52.7 (60.1-45.7)	48.8 (58.5-39.1)	51.9 (59.4-45.5)



---

**FIGURE 5.3 (LEFT) | COMBINED MITOCHONDRIAL AND NUCLEAR CHRONOGRAM.** Combined mitochondrial (with third codon positions RY-coded) and nuclear molecular clock dating tree with palaeognath divergence times shown next to each node with time (Ma) along the x-axis. Grey bars represent 95% highest posterior density intervals. The timings of major geological events are plotted along the time line. Volant taxa are represented by a grey silhouette as opposed to flightless taxa in black. Fossils used to calibrate the molecular clock are designated by an adjacent clock symbol (n.b., some among neognath calibrations are not depicted). Yellow arrows indicate minimum possible estimates for the evolution of flightlessness along each major palaeognath lineage. The continental distribution of each taxa is depicted by the adjacent transparency (n.b., the distribution for neognath outgroups is global). Nodes labeled A-E refer to those named in Table 5.1.

grouping tend to decrease as more mitochondrial positions are RY-coded (Table S5.9.17). Although there is no support for this topology from nuclear data alone, there is extremely little bootstrap support for the alternative grouping of emus sister to moa/tinamous (21%; Figure 5.2b). Although some clades in the nuclear tree are poorly resolved, there is good complementarity between the mitochondrial and nuclear datasets, as shown by the results of the AU test (above).

### 5.5.6 MOLECULAR DATING

In general, divergence dates estimated from alternative analyses (S5.9.12, Table S5.9.19) are in broad agreement with one another, suggesting that the node ages determined below remain largely unaffected by alternative analyses.

With mitochondrial third codon positions RY-coded, divergence estimates for all palaeognath orders span 16.3 Ma, between about 69 Ma and 52.7 Ma (Table 5.1; Figure S5.9.5). Without RY-coding, divergence dates differ from these estimates by less than 1 Ma (Table S5.9.19). In contrast, stem divergences of all palaeognath orders (except Dinornithiformes and Tinamiformes) estimated by nuclear data are older than the mitochondrial estimates (occurring between 72.1 Ma and 48.8 Ma; Table 5.1, Figure S5.9.5), and span 23.3 Ma—7 Ma slower than mitochondrial estimates. However, the nuclear data indicates that divergence from the base of Notopalaeognathae to the elephant bird/kiwi clade spans only 5.7 Ma, as opposed to 11.5 Ma estimated from the mitochondrial data—5.8 Ma faster. The divergence between elephant birds and kiwi occurs approximately 59.2 Ma based on the nuclear data, and approximately 53.5 Ma based on the mitochondrial data (Table 5.1). In contrast to the mitochondrial estimate, the nuclear estimate for the elephant bird/kiwi split is closer to the K-Pg boundary, occurring approximately 10.4 Ma *before* the divergence between moa and tinamous (rather than contemporaneously), and comes a mere 0.8 Ma after the Casuariiformes split from their common ancestor (as opposed to 4.5 Ma; Table 5.1).

Nevertheless, the dating estimates for the mitochondrial and nuclear phylogenies are consistent; in fact, the combined analysis estimates divergence between elephant birds and kiwi to be between these, approximately 54.2 Ma, and total

divergence time of all palaeognath orders is approximately 17.3 Ma, between 69.2 and 51.9 Ma (Table 5.1). The divergence between the two genera of elephant birds (*Aepyornis* and *Mullerornis*) is estimated to have occurred approximately 27.6 Ma (95% HPD 33.9-21.9 Ma).

## 5.6 DISCUSSION

---

### 5.6.1 MODE *and* TEMPO OF PALAEOGNATH EVOLUTION

We present the first palaeognath phylogeny based on nuclear and mitochondrial DNA data for all recent orders, and our analyses differ from other recent studies in several ways, which we detail below. Hence, nuclear and mitochondrial aDNA from the extinct Madagascan elephant bird, *Aepyornis* sp. has important implications for understanding palaeognath evolution.

Both nuclear and mitochondrial DNA unequivocally place elephant birds sister to kiwi, confirming the mtDNA findings of Mitchell et al. (2014); this is even despite the low coverage of the nuclear loci. Although the palaeognath root is ubiquitously placed between the ostrich and notopalaeognaths, its position is stabilised by the addition of nuclear DNA. Similarly, relative to the nuclear data alone, including mitochondrial DNA substantially improved the phylogenetic resolution (i.e., bootstrap support and Bayesian posterior probability), and brings the topology in agreement with most other recent studies (e.g., Mitchell et al. 2014).

However, the branching order of major clades within Notopalaeognathae (moa/tinamous, rheas, and emus/cassowaries) is less certain. In no analysis are moa/tinamous placed basal among Notopalaeognathae as suggested by Haddrath and Baker (2012) and Baker et al. (2014) (Table S5.9.17). While the position of the rheas as deepest among the notopalaeognaths is in agreement with Mitchell et al. (2014) and Phillips et al. (2010), their placement has very low statistical support that is only restored by the inclusion of nuclear DNA. Low statistical support and short internal branches at the base of Notopalaeognathae (Table S5.9.17) indicate that its constituent clades may have diverged near-simultaneously. The apparent polytomy may potentially be better resolved with the addition of data from more

nuclear loci, particularly for taxa with little nuclear data (e.g., cassowary; though it is worth noting that our nuclear topology is similar to that of Prum et al. (2015) whose tree was based upon a much larger genomic data set, albeit lacking the extinct ratites). However, even with many thousands of DNA base pairs of across hundreds of loci, statistical support for these internal branches remains poor (Harshman et al. 2008; Smith et al. 2013; Baker et al. 2014). In order to achieve fine scale resolution of these nodes, DNA from retroelements may be beneficial. Retroelements are generally not homoplastic and analysis of insertion events tends to be immune to certain biases (Haddrath and Baker 2012) that potentially arise when using protein-coding sequence (Jarvis et al. 2014), such as GC content among taxa in third codon positions (although here, GC content in nuclear third codon positions is unlikely to generate bias, being similar among palaeognaths; Table S5.9.12, Figure S5.9.3). Loci that have a faster rate of evolution potentially contain greater phylogenetic signal and have better resolved the relationships among neognaths (Jarvis et al. 2014). However, more data is not guaranteed to robustly resolve all palaeognath relationships; some neognath polytomies persist despite the inclusion of data from over 10,000 orthologous nuclear sequences (Jarvis et al. 2014). Resolving the basal divergence within Notopalaeognathae may also be confounded by incomplete lineage sorting; accordingly, multispecies coalescent models may benefit phylogenetic reconstructions. However, the short, low information aDNA sequences obtained here make implementing such approaches inappropriate (Mirarab et al. 2014; Tonini et al. 2015). Nevertheless, the elephant bird/kiwi grouping would be far less likely conflated with deep coalescence issues, given the high support values and a longer inferred stem lineage.

The combined analysis is currently our best estimate for the tempo of the palaeognath radiation, being based on our best-resolved phylogeny and most precise divergence estimates (i.e., smallest Bayesian 95% HPDs on average). This mirrors the findings of Ksepka and Phillips (2015) that nuclear data and RY-coded mitochondrial genomes provide closely comparable divergence estimates for birds, and together provide a more precise estimate of divergence time than either alone, which advocates for the combined approach taken here. The total divergence time among palaeognaths estimated from the combined mitochondrial and nuclear data were about 6 Ma more rapid than previously suggested by Mitchell et al. (2014),

who estimated the divergence time to occur across approximately 23 Ma based on mitochondrial DNA only. Similarly, divergence across the notopalaeognath polytomy spans 5.9 Ma—4.3 Ma faster than estimated by Mitchell et al. (2014) and occurring closer to the K-Pg boundary. This is consistent with the rapid radiation of notopalaeognaths implied by the lack of phylogenetic resolution at the base of this group. Furthermore, our data date the first three continental divergences within Notopalaeognathae (rheas from the remaining notopalaeognaths, moa/tinamous from novaeratites, and Casuariiformes from elephant birds/kiwi) to the Paleocene, coincident with the larger Neoaves radiation (Jarvis et al. 2014; Prum et al. 2015). In comparison, this diversification was inferred to take twice as long by mtDNA alone (Phillips et al. 2010; Mitchell et al. 2014), or to have occurred approximately 20 Ma earlier (Baker et al. 2014; Haddrath and Baker 2012) or 20 Ma later (Prum et al. 2015) with nuclear genomic data. The stem divergences of several palaeognath orders also occur after 66 Ma: the stem lineage of Tinamiformes, Dinornithiformes, Casuariiformes, Aepyornithiformes, and Apterygiformes all have origins in the Late Palaeocene-Early Eocene. This strengthens the argument that the Cretaceous-Palaeogene extinction event was the catalyst for a rapid radiation of palaeognaths, as well as for the neognaths, due to ‘ecological opportunity’, where an abundance of resources (Myers 2012) that were available to surviving organisms in the wake of the K-Pg mass extinction triggered a selective shift that led to adaptive divergence.

The divergence time we estimate between elephant birds and kiwi from the combined nuclear and mitochondrial data (Table 5.1; Figure 5.3) is not significantly different from that estimated by Mitchell et al. (2014) (50 Ma, 95% HPD 61.5-40.1 Ma); however, our estimate is more precise (i.e., Bayesian 95% HPDs are smaller), and occurs closer to the K-Pg boundary, further supporting the idea of a rapid radiation of palaeognaths. In addition, we confirm that the divergence between elephant birds and kiwi comes well after Madagascar and New Zealand were isolated (ca. 30 Ma later), suggesting that vicariance cannot explain the current distribution of ratites. Although our results reject vicariance to explain the distribution of elephant birds, dispersal may have occurred by a volant common ancestor that evolved flightlessness multiple times (Mitchell et al. 2014; Haddrath and Baker 2012; Phillips et al. 2010), and/or by ‘island hopping’ via land bridges

that may have existed at this time. While the latter hypothesis is unlikely (Ali 2011), new fossil evidence would be required to definitively distinguish between these scenarios. The idea of dispersal by a volant common ancestor is consistent with the fact that no flightless bird species that have regained the ability to fly have ever been documented, as well as the finding of a potentially volant “proto” kiwi from the early Miocene of New Zealand (Worthy 2013), in addition to Eocene fossils of other volant palaeognaths (e.g., *Lithornis*) found in Europe and the Northern Hemisphere (Houde, 1988; Mayr, 2008). If this is the case, flightlessness evolved at least six times among palaeognaths (Figure 5.3).

Thus, our results suggest that palaeognaths rapidly diversified but converged upon flightlessness independently in response to similar environmental pressures; for example, the sudden, simultaneous generation of novel ecospace across multiple landmasses that would have occurred with the extinction of non-avian dinosaurs at the K-Pg boundary 66 Ma. This mass extinction eliminated all known large terrestrial predators, and new vegetative habitats would have been readily exploitable by palaeognath ancestors with little competition from mammals. Our data is consistent with this hypothesis as all major stem notopalaeognath clades (rheas, moa/tinamous, and novaeratites) diverged before 55 Ma; only after this time did the maximum body size for cursorial mammalian herbivores reach a comparable size, and mammalian carnivores remained under 100 kg until ca. 50 Ma (Smith et al. 2010; Saarinen 2014). This scenario may be analogous to the repeated evolution of flightlessness among Gruiformes, such as rails, that have also undergone multiple rapid losses-of-flight in the recent past in response to island habitats lacking mammalian predators and competitors. Thus, the derivation of flightlessness in ratites appears to be another exemplary case of parallel evolution (Mitchell et al. 2014; Harshman et al. 2008).

Finally, the diversification of elephant bird genera is estimated to be about 10 Ma earlier than previous estimates (Mitchell et al. 2014). The divergence between *Aepyornis* and *Mullerornis* seems to coincide with the Eocene-Oligocene (E-O) boundary around 33.9 Ma—a time characterized by global cooling, widespread extinction, and floral turnover (also known as the ‘*Grande Coupure* biotic turnover’). At this time, Madagascar drifted above 30°S in latitude into the warm

and humid subtropics (Yoder 2004), establishing its major biomes as well as many of Madagascar's endemic flora (Buerki 2013). A change in plant communities brought about by this environmental upset is hypothesized to have prompted the diversification of lemurs on Madagascar (Yoder 2004), and similarly, may have triggered the divergence between elephant bird genera. The E-O boundary also may have punctuated speciation within other notopalaeognath lineages, as this time seems to also correspond with a diversification within the Casuariiformes (i.e., divergence between emu and cassowary) and Tinamiformes. However, testing this hypothesis would require more precise dating of these crown divergences. Shallow divergences that are poorly informed by mainly deeper calibrations ages tend to be more susceptible to error and different coding strategies than deep-divergences (Ksepka et al. 2014).

### 5.6.2 EGGHELL PALAEOGENOMICS *and* FUTURE DIRECTIONS

DNA preservation in fossils from tropical-temperate environments like Madagascar will always be challenging. For the first time, we explore the viability of eggshell as a substrate for the retrieval of whole genomes. In comparison to the partial mitogenome published by Mitchell et al. (2014), we retrieved a mitochondrial genome that is about 2271 bp more complete with roughly twice the average coverage and less than 100 ambiguous sites. With almost 3% avian DNA recovered from shotgun data alone, a higher percentage of endogenous DNA was recovered from eggshell than has been recovered from bone, which typically results in less than 1% endogenous DNA for other taxa (Shapiro and Hofreiter 2014). The mapped M06-M028 eggshell reads are likely conservative estimates due to requirements to map to a distant relative. Furthermore, capture could be optimised by enrichment with genomic baits from a closer relative (as opposed to tinamou whole-genome baits). Considering this, it is not beyond the realm of possibility to obtain a full genome of *Aepyornis*: based on reads mapping to aves (0.0012 Gb from 0.4 Gb) and assuming the size of the elephant bird genome is about the same as the kiwi (1.6 Gb; Le Duc et al. 2015), 1300 times the amount of sequence data generated here would be needed to obtain a draft 1X genome if the library does not sequence to saturation (estimated to be the equivalent of two full HiSeq runs).

Furthermore, eggshell offers probably the only opportunity to study the genetic diversity within and between species of elephant birds, and examine how this relates to their geographic distribution within Madagascar, because it is readily available in many places where elephant birds once nested, while fossil bone is rare. In addition, over half the eggshell samples tested yielded amplifiable elephant bird aDNA, which is a high proportion for ancient samples, particularly in hot environments; these samples, particularly those that were found *in situ*, may also yield genomic information in the future. Whole mitochondrial genomes of elephant birds from across their former range will allow us to address many unsolved questions about elephant bird evolution. For example, we could potentially predict how many ‘molecular species’ of elephant bird actually existed given the paucity of fossil bone. Considering that the biodiversity of other ratites (notably moa) has been grossly overestimated based on morphology, it is entirely possible that fewer species of elephant birds existed than have been named (approximately eight; Brodkorb 1963; Hume and Walters 2012). Alternatively, more species or genera of elephant birds may exist than have been described. In fact, studies on other organisms, such as hominids, have described new species based solely on aDNA (e.g., Krause et al. 2010). The integration of eggshell aDNA, 14C dating and eggshell stable isotope data provides a means by which to study species boundaries and extinction timelines.

The ability we now have with fossil eggshell to characterise parts of the elephant bird nuclear genome will not only improve our ability to further resolve tree topologies, but also will provide insights into elephant bird evolution that are not possible to glean from maternally inherited mitochondrial DNA alone: that is, nuclear DNA (including non-coding DNA) will allow us to examine the forces driving evolution, such natural selection and genetic drift by examining microsatellites (as has been done in moa; Allentoft et al. 2011) and genes that show signatures of selective sweeps or population bottlenecks. A better understanding of the population history of elephant birds will be import for determining the factors responsible for their speciation and extinction. There is also the potential to uncover functional mutations, including those responsible for island gigantism, flightlessness, and egg development, which may give insight into behavior and life-history strategies. For example, a genome-wide study of the kiwi found mutations

in opsin genes that could be responsible for nocturnal adaptations (Le Duc et al. 2015), and other studies have uncovered similar adaptive mutations in extinct organisms using aDNA (Rompler et al. 2006; Lalueza-Fox et al. 2007; Lalueza-Fox et al. 2008; Lalueza-Fox et al. 2009). Unique mutations among ratites in flightlessness genes may lend further support for the independent evolution of flightlessness from a volant common ancestor.

## 5.7 CONCLUSION

---

The evolutionary history of palaeognaths has historically been contentious but is finally reaching a consensus as our results affirm those of recent studies. We provide independent evidence that elephant birds are the closest relatives of the kiwi, and have refined the dating of the divergences among palaeognaths, through the addition of nuclear aDNA that is essential for clarifying deep evolutionary relationships. This study has also shown that eggshell can be a valuable source of high-quality aDNA and is an excellent substrate for studying the palaeogenomics of extinct organisms that lived in climates that are suboptimal for the preservation of aDNA. For elephant birds, aDNA from eggshell provides an avenue for future research into their phylogeography and functional genomics that will further shed light on the evolution of these understudied birds.

## 5.8 REFERENCES

---

- Ali JR and Krause DW. 2011. Late Cretaceous bioconnections between Indo-Madagascar and Antarctica: refutation of the Gunnerus Ridge causeway hypothesis. *Journal of Biogeography* **38**: 1855-1872.
- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, et al. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B-Biological Sciences* **279**: 4724-4733.
- Allentoft ME, Oskam C, Houston J, Hale ML, Gilbert MT, et al. 2011. Profiling the Dead: Generating Microsatellite Data from Fossil Bones of Extinct Megafauna-Protocols, Problems, and Prospects. *PLOS One* **6**.

- Allentoft ME and Rawlence NJ. 2012. Moa's Ark or volant ghosts of Gondwana? Insights from nineteen years of ancient DNA research on the extinct moa (Aves: Dinornithiformes) of New Zealand. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 36-51.
- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Baker AJ, Haddrath O, McPherson JD, Cloutier A. 2014. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Molecular Biology and Evolution* **31**: 1686-1696.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**.
- Blankenberg DVK, Coraro N, Ananda G, Lazarus R, Mangan M, et al. 2010. Galaxy: a web-based genome analysis tool for experimentalists. In: al. FMAe, editor. *Current Protocols in Molecular Biology*. Unit 19.10.11-21.
- Brodkorb P. 1963. Catalogue of fossil birds. *Bulletin of the Florida State Museum Biological Sciences* **7**: 206-207.
- Buerki S, Devey DS, Callmander MW, Phillipson PB, Forest F. 2013. Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society* **171**: 304-329.
- Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, et al. 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**: 172-175.
- Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillette LJ, et al. 2007. Patterns of vertebrate isochore evolution revealed by comparison of expressed mammalian, avian, and crocodylian genes. *Journal of Molecular Evolution* **65**: 259-266.

- Claramunt S and Cracraft J. 2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances* **1**.
- Clarke JA, Tambussi CP, Noriega JI, Erickson GM, Ketcham RA. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature* **433**: 305-308.
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. 2001. Complete mitochondrial genome sequences of two extinct moa clarify ratite evolution. *Nature* **409**: 704-707.
- Cracraft J. 1973. Continental drift, palaeoclimatology, and evolution and biogeography of birds. *Journal of Zoology* **169**: 455-545.
- Cracraft J. 1974. Phylogeny and evolution of ratite birds. *Ibis* **116**: 494-521.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15758-15763.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**.
- Dewar R. 1984. Extinctions in Madagascar: the loss of the subfossil fauna. In. *Quaternary extinctions: A prehistoric revolution*. Tuscon: University of Arizona Press.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.
- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.

- Giardine BR, Hardison RC, Burhans R, Elnitski L, Shah P, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451-1455.
- Gilbert MTP, Tomsho LP, Rendulic S, Packard M, Drautz DI, et al. 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**: 1927-1930.
- Goecks J, Nekrutenko A, Taylor J and The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biology* **11**.
- Grealy A, Macken A, Allentoft ME, Rawlence NJ, Reed E, et al. 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *Journal of Quaternary Science* **31**: 33-45.
- Guindon S and Gascuel O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systems Biology* **52**: 696-704.
- Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**: 1763-1768.
- Haddrath O and Baker AJ. 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proceedings of the Royal Society B-Biological Sciences* **279**: 4617-4625.
- Hahn C, Bachmann L, Chevreur B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* **41**.

- Harshman JB, Edward L, Braun MJ, Huddleston CJ, Bowie RCK, et al. 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. *Proceedings of the National Academy of Sciences* **105**: 13462-12467.
- Hoss M, Jaruga P, Zastawny TH, Dizdaroglu M, Paabo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24**: 1304-1307.
- Houde P. 1986. Ostrich ancestors found in the northern-hemisphere suggest new hypothesis of ratite origins. *Nature* **324**: 563-565.
- Huelsenbeck JP and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Hume JP and Walters M. 2012. *Extinct Birds*: Bloomsbury Publishing.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **345**: 1320-1331.
- Johnston P. 2011. New morphological evidence supports congruent phylogenies and Gondwana vicariance for palaeognathous birds. *Zoological Journal of the Linnean Society* **163**: 959-982.
- Jonsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682-1684.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

- Krause J, Fu Q, Good JM, Viola B, Shunkov MV, et al. 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**: 894-897.
- Ksepka DT and Phillips MJ. 2015. Avian diversification patterns across the K-Pg boundary: Influence of calibrations, datasets, and model misspecification. *Annals of the Missouri Botanical Garden* **100**: 300-328.
- Ksepka DT, Ware JL, Lamm KS. 2014. Flying rocks and flying clocks: disparity in fossil and molecular dates for birds. *Proceedings of the Royal Society B-Biological Sciences* **281**.
- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, et al. 2009. Bitter taste perception in Neanderthals through the analysis of the TAS2R38 gene. *Biological Letters* **5**: 809-811.
- Lalueza-Fox C, Gigli E, de la Rasilla M, Fortea J, et al. 2008. Genetic characterization of the ABO blood group in Neandertals. *BMC Evolutionary Biology* **8**: 342.
- Lalueza-Fox C, Römpler H, Caramelli D, Stäubert C, Catalano G, et al. 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* **318**: 1453-1455.
- Le Duc D, Renaud G, Krishnan A, Almén MS, Huynen L, et al. 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology* **16**.
- Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJP. 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques* **54**: 321-326.
- Mayr G. 2008. First substantial middle Eocene record of the Lithornithidae (Aves): a postcranial skeleton from Messel (Germany). *Annales de Paléontologie* **94**: 29-37.

- Mayr G. 2009. *Palaeogene Fossil Birds*. Berlin: Springer-Verlag.
- McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* **49**: 271-300.
- Mirarab S, Bayzid MdS, Boussau B, Warnow T. 2014. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* **346**: 1250463.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Myers EA. 2012. Ecological opportunity: trigger of adaptive radiation. *Nature Education Knowledge* **3**.
- Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, et al. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **277**: 1991-2000.
- Oskam CL, Jacomb C, Allentoft ME, Walter R, Scofield RP, et al. 2011. Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *Journal of Archaeological Science* **38**: 2589-2595.
- Phillips MJ, Gibb GC, Crimp EA, Penny D. 2010. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. *Systems Biology* **59**: 90-107.
- Phillips MJ and Pratt RC. 2008. Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums). *Molecular Phylogenetics and Evolution* **46**: 594-605.

- Prum RO, Berv JS, Dornburg A, Field DJ, et al. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**: 569- 578.
- Rambaut A. 1996. Sequence Alignment (Se-AL) Program. Version 2.0a. Oxford: University of Oxford.
- Rambaut A, Suchard MA, Xie W, Drummond AJ. 2003. Tracer: MCMC Trace Analysis Tool. Version 1.6.1pre.
- Ramírez O, Gigli E, Bover P, Alcover JA, Bertanpetit J, et al. 2009. Palaeogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine. *PLOS One* **4**: e5670.
- Rompler H, Rohland N, Lalueza-Fox C, Willerslev E, Kuznetsova T, et al. 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* **313**: 62-62.
- Saarinen JJ, Boyer AG, Brown JH, Costa DP, Morgan Ernest SK, et al. 2014. Patterns of maximum body size evolution in Cenozoic land mammals: eco-evolutionary processes and abiotic forcing. *Proceedings of the Royal Society B-Biological Sciences* **281**.
- Schroeder H, Ávila-Arcos MC, Malaspinas AS, Poznik GD, Sandoval-Velasco M, et al. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 3669-3673.
- Shapiro B and Hofreiter M. 2012. *Ancient DNA: methods and protocols*. New York Humana Press (Springer Science).
- Shapiro B and Hofreiter M. 2014. A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA. *Science* **343**.

- Smith FA, Boyer AG, Brown JH, et al. 2010. The evolution of maximum body size of terrestrial mammals. *Science* **330**: 1216-1219.
- Smith JV, Braun EL, Kimball RT. 2013. Ratite Nonmonophyly: Independent Evidence from 40 Novel Loci. *Systems Biology* **62**: 35-49.
- Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, et al. 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Systems Biology* **56**: 673–684.
- Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **30**: 1312-1313.
- Silvestro D and Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution* **12**: 335-337.
- Swofford DL. 2003. PAUP\*. Version 4.0b. Sunderland, Massachusetts: Sinauer Associates.
- Tonini J, Moore A, Stern D, Shcheglovitova M, Ortí G. 2015. Concatentation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLOS Currents Tree of Life* **1**: doi: 10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be.
- Tovondrafale T, Razakamanana T, Hiroko K, Rasoamiaramanan A. 2014. Palaeoecological analysis of elephant bird (Aepyornithidae) remains from the Late Pleistocene and Holocene formation of southern Madagascar. *Malagasy Nature* **8**.
- Vallender EJ. 2009. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. *Methods* **49**: 50-55.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.

- Worthy TH and Scofield RP. 2012. Twenty-first century advances in knowledge of the biology of moa (Aves: Dinornithiformes): a new morphological analysis and moa diagnoses revised. *New Zealand Journal of Zoology* **39**: 87-153.
- Worthy TH, Worthy JP, Tennyson AJD, Salisbury SW, Hand SJ, et al. 2013. Miocene fossils show that kiwi (Apteryx, Apterygidae) are probably not phyletic dwarves. *Paleornithological Research: Proceedings of the 8<sup>th</sup> International Meeting of the Society of Avian Paleontology and Evolution*: 63-80.
- Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.
- Yang Z and Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212-226.
- Yoder AD and Nowak MD. 2006. Has vicariance or dispersal been the prominent biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics* **37**: 405-431.
- Yoder AD and Yang Z. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Molecular Ecology* **13**: 757-773.
- Yonezawa T, Segawa T, Mori H, Campos PF, Hongoh Y, et al. 2016. Phylogenomics and morphology of extinct palaeognaths reveal the origin and evolution of ratites. *Current Biology* **27**: 1-10.
- Zhang G, Li C, Li Q, Li B, Larkin DM, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**: 1311-1320.
- Zhang G, Rahbek C, Graves GR, Lei F, Jarvis ED, et al. 2015. Genomics: Bird sequencing project takes off. *Nature*: **522**.

## 5.9 SUPPLEMENTARY INFORMATION

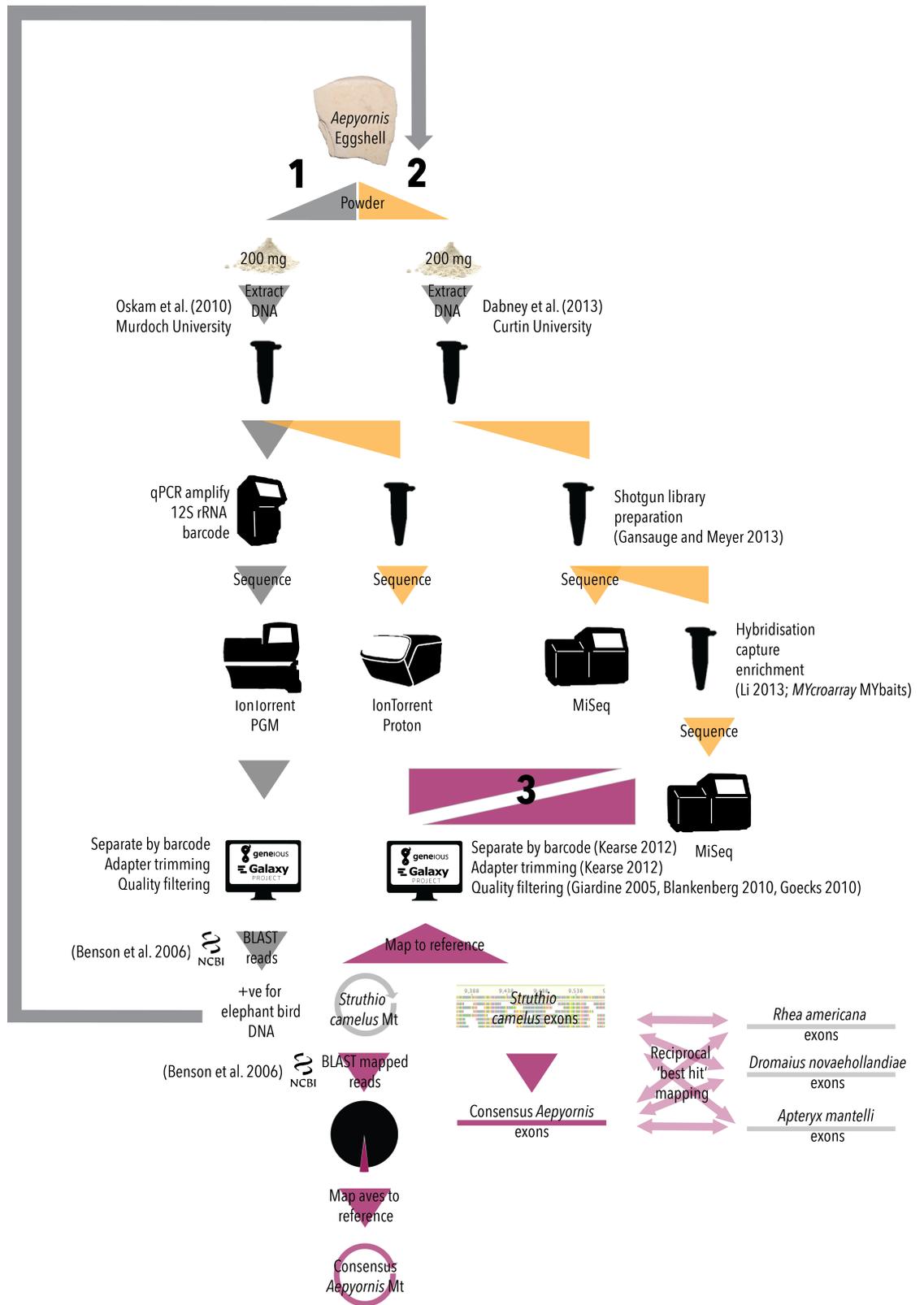
---

### S5.9.1 DATING *and* MORPHOLOGICAL IDENTIFICATION *of* EGGSHELL

M06-A028 was one of 50 collections containing common *Aepyornis* eggshell recovered from rich sites in the vicinity of Talaky on the south coast of Madagascar. The eggshell were identified as *Aepyornis* by their thickness, between 3 and 4 mm. Although M06-A028 was not directly dated by radiocarbon, the age of the sample is tightly constrained by sample M06-A025, collected only 150 m ENE of M06-M028, and from the same stratigraphic unit. Both samples have been "dated" by amino acid racemisation geochronology, which provides a powerful correlation tool. The two *Aepyornis* samples have indistinguishable racemisation values, confirming they are age correlatives. Sample M06-A025 has a conventional radiocarbon age of  $1305 \pm 15$  years BP (CURL- 19110) and a calibrated age of  $730 \pm 40$  years AD using Calib 7.1 and SHCal13 (Hogg et al. 2013). The samples were prepared at the University of Colorado's Laboratory for AMS Radiocarbon Preparation and Research and analysed at the University of California Irvine's WM Keck Carbon Cycle Accelerator Mass Spectrometry Laboratory. Preparation included a 50% leach of the sample to limit uncertainties due to carbon diffusion. Racemisation values were measured at the University of Colorado's Amino Acid Geochronology Laboratory using High-Pressure Liquid Chromatography and fluorescence detection.

### S5.9.2 ADNA EXTRACTION

Eggshell samples were prepared in a designated clean facility at Murdoch University, WA, Australia. Coveralls, double gloves, mask, eyewear, and boots were worn as personal protective equipment, and to minimise contamination with exogenous DNA. Before beginning, all surfaces (including tray, table, chair, *Dremel* tool, drill bits, electronic balance, mortar and pestles, and gloves) were cleaned with a solution of 10% bleach, followed by a solution of 70% ethanol (Willerslev and Cooper 2005; Knapp et al. 2012). The surface of eggshell was also cleaned with a solution of 10% bleach, followed by 70% ethanol.



**Figure S5.9.1** | SUMMARY OF THE STEPS TAKEN TO ISOLATE AND RECONSTRUCT THE MITOCHONDRIAL GENOME AND RETRIEVE NUCLEAR LOCI FROM *AEPYORNIS* EGG SHELL. Step one involved screening eggshell samples for elephant bird DNA by amplifying and sequencing a *12S rRNA* barcoding region. Shotgun libraries were then generated for the best sample and sequenced on two different NGS platforms (step two). Reads were mapped onto published ostrich reference sequences (step three).

The surface of the eggshell was removed using a *Dremel* drill and discarded. A clean drill bit was used to grind the eggshell into 200 mg of fine powder, which was placed into a sterile 2.0 ml *Eppendorf* tube and stored at -20°C until extraction. All extractions and subsequent qPCR reactions were prepared in a separate ultra-clean room to avoid contamination. All post-PCR methods were performed in a physically separated laboratory in keeping with standard aDNA practice (Willerslev and Cooper 2005; Shapiro and Hofreiter 2012).

Eggshell powder was extracted as per Oskam et al. (2010) and Shapiro and Hofreiter (2012). Digest buffer was prepared containing final concentrations of 1% Triton-X-100 (70 µl 10% Triton-X-100), 0.02 M Tris-HCl (14 µl 1 M Tris-HCl pH 8), 1 mg/ml Proteinase K powder (0.7 mg), 0.01 M DTT (7 µl 1M DTT), topped up to a final volume of 700 µl with 0.5 M EDTA (595 µl, final concentration 0.425 M). EDTA, Tris, DTT and Proteinase K were combined first and mixed, followed by the Triton-X-100. The buffer was incubated at 55°C with rotation for five min to activate the enzyme and dissolve the Triton X-100. 700 µl of digestion buffer was then added to 200 mg of eggshell powder. The tube was sealed with Parafilm M® (*Bemis NA*), and incubated in a rotating oven at 55°C for 24 hours. Samples were then centrifuged for 10 min at 13,000 rpm in a bench top centrifuge to collect undigested eggshell and cell debris. The supernatant was then transferred to a Vivaspin 500 column (MWCO 30 000; *Sartorius Stedim Biotech*, Germany), 500 µl at a time, and centrifuged at 13,000 rpm with the membrane facing outward for 10 min, or until concentrated to 50 µl. The concentrate was then transferred to a clean 1.5 ml microcentrifuge tube. Five volumes (250 µl) of Buffer PBI (*QIAGEN*; cat. No. 19066) was then added at to the concentrate, and mixed by pipetting up and down. 3 µl of sodium acetate was added to the mixture, if the pH was too high (purple). The 300-µl mixture was then transferred to a *QIAGEN* quick spin column (cat. No. 28106), allowed to incubate for two min at room temperature, and then centrifuged for one min at 13,000 rpm in a bench top centrifuge. After discarding the flow-through, 500 µl of Buffer AW1 (*QIAGEN*; cat. No. 19081) was added to the column, and the column was allowed to incubate for two min at room temperature, then centrifuged for one min at 13,000 rpm to wash the membrane. After discarding the flow-through, 500 µl of Buffer AW2 (*QIAGEN*; cat. No. 19072) was added to the column, and the column was allowed to incubate for two min at room temperature, and centrifuged for one min at

13,000 rpm to further wash the membrane. After discarding the flow-through, the column was centrifuged for a further one min at 13,000 rpm to ensure all buffer had passed through the membrane. The column was then placed into a clean 1.5 ml microcentrifuge tube, and 60 µl of Buffer EB (*QIAGEN*; cat. No. 19086) was added to the centre of the membrane. This was allowed to sit at room temperature for five min, and then the column was centrifuged for one min at 10,000 rpm to elute the DNA. The flow-through was then passed the column one more time (after incubating in the column for an additional five min at room temperature) to ensure all DNA was eluted. Eluted DNA was transferred to a clean, labelled 1.5 ml Lo-bind microcentrifuge tube (*Eppendorf*), and stored at -20°C.

Eggshell was independently powdered and extracted in a designated clean facility at Curtin University, WA, Australia, following Dabney et al. (2013) with minor changes. Eggshell digest buffer contained final concentrations of 0.45 M EDTA (1.66 ml of 0.5 M stock) and 0.25 mg/mL Proteinase K pH 8 (0.42 mg). 1.66 ml of digest buffer was added to 200 mg of eggshell powder. Eggshell powder was resuspended by vortexing. Samples were sealed with Parafilm M® and incubated with rotation at 37°C for 24 hr. Undigested powder was pelleted by centrifugation for 10 min at maximum speed in a bench top centrifuge. The supernatant was removed, and added to 21.66 ml of binding buffer, which contained final concentrations of 5 M guanidine hydrochloride (10.34 g powder), 40% isopropanol (8.66 ml of 100% isopropanol), 0.05% Tween-20 (10.83 µl of 100% Tween-20), and 90 mM sodium acetate pH 5.2 (649.8 µl of a 3M stock), topped up to 21.66 ml with Ultrapure water. A 15 ml Zymo-Spin V column (*Zymo Research*) was attached to a *QIAGEN* MinElute silica spin column (cat. No. 28004), secured in place with Parafilm M®, and placed into a 50 ml Falcon tube. 10 mL of binding buffer-supernatant solution was then transferred into the extension reservoir, capped, and sealed with Parafilm M®. This was centrifuged for four minutes at 1,500 x g. Flow-through was discarded and the remaining binding buffer-supernatant solution was passed through the column. Next, the filter assembly was removed from the falcon tube, disassembled, and the spin column was placed in a clean 2 ml collection tube, which was then centrifuged for one min at 13,000 rpm. The filter was washed twice with 750 µl of PE buffer (*QIAGEN*; cat. No. 28004) by centrifuging at 13,000 rpm for one min and discarding the flow through. The column was spun for a further one

min at maximum speed, and the placed in a clean 1.5 ml collection tube. To elute, 30 µl of EB buffer was pipetted on to the membrane and incubated for five minutes at room temperature. Eluate was collected by centrifugation for one min at maximum speed. Another 30 µl of EB buffer was added to the column membrane and incubated for a further five min at room temperature, and eluted by centrifugation for one min at maximum speed. The eluate (60 µl total) was transferred to a clean 1.5 ml Lo-bind *Eppendorf* tube with 1.5 µl of 1% TE-Tween added, and stored at -20°C.

### **S5.9.3 SCREEN SAMPLES *via* AMPLICON SEQUENCING**

The neat (i.e., undiluted) and a 1/10 dilution of each extract were amplified using primers targeting a 95 bp (53 bp without primers) region of the mitochondrial *12S rRNA* gene in birds (Table S5.9.1; Cooper 1994). The PCR reaction contained reagents in final concentrations of: 1.2 mg/ml *Fisher* BSA (3 µl of a 10 mg/ml stock), 1X *ABI* GeneAmp PCR buffer (2.5 µl of a 10X stock), 2.5 mM *ABI* MgCl<sub>2</sub> (2.5 µl of a 25 mM stock), 0.4 µM *IDT* forward primer (12SA; 1 µl of a 10 µM stock), 0.4 µM *IDT* reverse primer (12SC; 1 µl of a 10 µM stock), 0.25 mM *Bioline* dNTPs (0.25 µl of a 25 mM stock), 2 U *ABI* AmpliTaq Gold DNA polymerase, 0.6 µl of a 1:2000 dilution of 10 000X SYBR green dye, and 2 µl of DNA in a total reaction of volume of 25 µl (including 14.2 µl *GIBCO* HPLC-grade water). Thermalcycling conditions were: 95°C for five min, 50 cycles of 95°C for 30 sec, 54°C for 30 sec, and 72°C for 45 sec, followed by a melt curve of 95°C for 15 sec, 60°C for one min, 95°C for 15 sec, 72°C for 10 min. C<sub>T</sub> (cycle threshold) values were recorded for each sample, and amplification and melt curves were analysed to give a rough indication of PCR efficiency. The dilution exhibiting the least inhibition and greatest PCR efficiency was selected for subsequent fusion-tagging (in general the neat).

Triplicate fusion-tag qPCRs were performed to generate sequencing amplicons of 12SAC flanked by unique indexing and sequencing adapter sequences (Table S5.9.1). Fusion-tag PCR reaction and cycling conditions were as above. A positive control, extraction control, and negative controls (water) were included. The replicate PCR products were combined and purified twice using an *Agencourt* AMPure XP PCR purification kit as per the manufacturer's instructions with minor

changes: 1.4  $\mu\text{l}$  of AMPure XP SPRI beads (*Beckman-Coulter*) were added to 1  $\mu\text{l}$  of PCR product (i.e., 1.4X the volume of PCR product). Cleaned PCR products were eluted in 30  $\mu\text{l}$  *QIAGEN* EB buffer). The amplicon pool was run on a 2% agarose gel electrophoresis (5  $\mu\text{l}$  of PCR product was combined with 0.5  $\mu\text{l}$  of 6X loading dye and run on a gel (2.2 g agarose, 110 mL 1X TAE buffer, 8  $\mu\text{l}$  GelRed) alongside 3  $\mu\text{l}$  of 50 bp DNA ladder (GeneRuler, *Fermentas*) for 40 minutes at 96 V; the gel was visualised and photographed using a *BioRad* transilluminator). Based on the relative band intensity of the products, was further combined in approximately equimolar concentrations to create a sequencing library.

The sequencing library was purified one more time by using an *Agencourt* AMPure XP PCR purification kit as above. The sequencing library was diluted 1/2000, 1/4000, 1/8000, 1/16000, 1/32000, 1/64000, 1/128000 in a total of 100  $\mu\text{l}$  of Ultrapure water. Each dilution was qPCR amplified in duplicate along side a standard of known concentration using primers complementary to the sequencing adapters, in order to quantify the number of template molecules in the final sequencing library. PCR ‘no-template controls’ were included. The PCR reaction contained reagents in final concentrations of: 1X *ABI* Power SYBR Master Mix (12.5  $\mu\text{l}$  of a 2X stock), 0.4  $\mu\text{M}$  *IDT* forward primer IT\_A (Table S2; 1  $\mu\text{l}$  of a 10  $\mu\text{M}$  stock), 0.4  $\mu\text{M}$  *IDT* reverse primer IT\_P1 (Table S2; 1  $\mu\text{l}$  of a 10  $\mu\text{M}$  stock), and 2  $\mu\text{l}$  of library in a total reaction of volume of 25  $\mu\text{l}$  (including 8.5  $\mu\text{l}$  *GIBCO* HPLC-grade water). Thermalcycling conditions were: 95°C for five min, 40 cycles of 95°C for 30 sec, 60°C for 45 sec, followed by a melt curve of 95°C for 15 sec, 60°C for one min, and 95°C for 15 sec.  $C_T$ -values were recorded and compared to the standard in order to calculate the number of copies in each dilution, and how much DNA to add to the sequencing reaction to achieve a desired bead-to-template ratio during emulsion PCR using the Ion Torrent OneTouch2 (*Life Technologies*). Sequencing was performed using the Ion PGM Sequencing 200 kit v.2 following the manufacturer’s instructions.

**TABLE S5.9.1** | METABARCODING PRIMERS USED, AND ION TORRENT ADAPTER SEQUENCES THAT FLANK THE METABARCODING PRIMERS FOR FUSION TAGGING. NNNNNNNN refers to the index sequence (any bases).

Name	Sequence (5'-3')	Target Taxa	Gene	Reference	Annealing Temp (°C)
12SA (Forward)	CTGGGATTAGATACCCCACTAT	Bird (universal)	12S rRNA	(Cooper 1994)	54
12SC (Reverse)	GTTTTAAGCGTTTGTGCTCG	Bird (universal)	12S rRNA	(Cooper 1994)	54
Forward fusion tag adapter	CCATCTCATCCCTGCGTGTCTCC GACTCAGNNNNNNN	-			
Forward fusion tag adapter	CCTCTCTATGGGCAGTCGGTGAT NNNNNNN	-			

**TABLE S5.9.2 | PRIMERS USED FOR QUANTIFY THE FINAL SEQUENCE LIBRARY BY QPCR.**

<b>Name</b>	<b>Sequence (5'-3')</b>	<b>Function</b>	<b>Annealing Temp (°C)</b>
IT_A	CCATCTCATCCCTGCGTGTC	Forward qPCR quant primer (IT)	60
IT_P1	CCTCTCTATGGGCAGTCGGTGAT	Reverse qPCR quant primer (IT)	60
P5	AATGATACGGCGACCACCGAGATCTACAC	Forward qPCR quant primer (MiSeq)	60
P7	CAAGCAGAAGACGGCATAACGAGAT	Reverse qPCR quant primer (MiSeq)	60

#### S5.9.4 QUALITY CONTROL

Sequence data in FASTQ format that had passed the default filtering parameters was downloaded directly from the Ion Torrent server or BaseSpace, and imported into Geneious v.7.1.2 (<http://www.geneious.com>; Kearse 2012). Sequences were sorted by barcode the forward barcode with no mismatches by selecting the “separate reads by barcode” option from the drop-down menu “Sequence”. Reverse sequencing adapters were trimmed from the 3’ end by selecting “Trim ends” under the “Annotate and Predict” drop down menu. Next, a file was created for each of the primer sequences (or adapter sequences for shotgun libraries) by selecting the “new sequence” option from the drop-down menu “Sequence” and pasting in the sequence and name. Gene specific primer (GSP)/shotgun adapter sequences were annotated on each read by selecting the “find annotations” option from the drop-down menu “Annotate and Predict”. Reads containing the forward GSP/shotgun adapter sequence allowing no mis-match were extracted by selecting “extract annotations” from the drop-down menu “Tools”. The primer name was inserted into the space “annotation name” “contains”, and selecting “OK”. Adapter sequences were removed from the reads by selecting the “trim ends” option under the drop-down menu “Annotate and Predict”. Reads then containing the reverse GSP/shotgun adapter (with no mismatches) were annotated, extracted, and trimmed as above. For shotgun and enriched data (not amplicon), reads were sorted by length by right clicking and selecting “sort by length”; sequences under 30 bp were deleted. Files were re-named by barcode and the word “trimmed”, and sequences were batch-renamed to include the file name by selecting “batch rename” from the drop-down menu “Edit”. Each file was then exported in FASTQ format.

The trimmed FASTQ files were then imported into the open-source, web-based bioinformatics platform Galaxy ([usegalaxy.org](http://usegalaxy.org); Giardine et al. 2005; Blankenberg et al. 2010; Goecks et al. 2010). “FASTQ groomer” was used to convert the quality score encoding from Sanger to ASCII. Sequences with an average quality score less than 25 were removed using “Generic FASTQ manipulation” -> “Filter FASTQ by quality score and length” by selecting “add new quality filter on a range of bases”, then “define base offsets as a percentage of read length”, followed by “aggregate read score for specified range” as “mean of scores”, and finally “keep read length

when aggregate score is greater than or equal to 25". Next, sequences were removed by applying quality cut-offs for a certain percentage of bases in the sequence using "FASTX-toolkit for FASTQ data": first, the "quality cut-off value" was set to 10 under "filter by quality", followed by "percent of bases in sequence that must have quality equal to/higher than cut-off value" as 100 to keep only those sequences that contained all bases with a quality score of higher than 10; next, the cut-off value was set to 15 and the percentage of bases to 95 to keep only those sequences where at least 95% of bases had a quality higher than 15; and finally, the cut-off value was set to 20 and the percentage of bases to 90 to keep only those sequences where at least 90% of bases had a quality higher than 20. These filtered sequences were exported from Galaxy and imported in Geneious, grouped ("Sequences"-> "Group sequences into a list"), and dereplicated ("Edit"-> "Find duplicates").

#### **S5.9.5 SHOTGUN SEQUENCING**

Shotgun sequencing libraries were prepared by following (Gansauge and Meyer 2013) with minor changes. Modifications to the adapters used are listed in Table S5.9.3. An extraction control, no-template (water) control, and CL104 positive control were also included in the library building process. At step 1, reactions were performed in 0.2 ml 8-well PCR strip tubes. 12 µl of DNA extract was used, and *Afu* UDG was replaced by Ultrapure water. We elected to not use uracil-DNA glycosylase because "many aDNA researchers are reassured of the authenticity of the resulting ancient sequences when random C-T transitions are observed in cloned products of a PCR, as this form of damage is common in ancient samples" (Shapiro and Hofreiter 2012). At step 7, ligation products were stored overnight at -20°C. At step 13, tubes were incubated in a rotating hybridisation oven for two min at 65°C as opposed to a thermal shaker. At step 13, tubes were transferred to a thermal shaker pre-cooled to 15°C as opposed to a thermocycler. Steps 14, 15, 18, 19, 23, and 25 were performed in a thermal shaker. Step 22 was performed in a rotating hybridisation oven. At step 25, the supernatant was stored in a 1.5 mL Lo-Bind *Eppendorf* tube at -20°C. The qPCR standard used in step 27 is listed in Table S5.9.3. After step 28, the PCR products were run on a 2% agarose gel electrophoresis (as above) in order to confirm the library building process.

**TABLE S5.9.3** | LIGATION ADAPTERS, FUSION PRIMERS, AND SEQUENCING PRIMERS USED FOR SHOTGUN SEQUENCING ON BOTH THE ION TORRENT AND MISEQ NGS PLATFORMS (modifications from Gansauge and Meyer 2013). Unique indexes are represented by NNNNNNN (any base) below.

Name	Sequence (5'-3')	Function	Annealing Temp (°C)
CL53	ACACGACGCTCTC-ddC	Double-stranded adapter, strand 1	RT
CL78	[Phosphate]AGATCGGAAG[C9Spacer]3[TEG-biotin]	Single-stranded adapter	60
CL105_106_Std	ACACTCTTCCCTACACGACGCTCTTCCTCG TCGTTGGTATGGCTTCTATCGUATCGATCG ATCGACGATCAAGGCGAGTTACATGAAGAT CGGAAGAGCACACGTCTGAACTCCAGTCA C	Synthetic qPCR standard	-
IT_A_Index_Fwd adapter	CCATCTCATCCCTGCGTGTCTCCGACTCAG NNNNNNNACACTCTTCCCTACACGACGC TCTT	Ion Torrent forward indexing primer	60
IT_P1-Index-Rev adapter	CCTCTCTATGGGCGAGTCGGTGATNNNNNN NGTGACTGGAGTTCAGACGTGT	Ion Torrent reverse indexing primer	60
P5-RD1-Index-Fwd adapter	AATGATACGGCGACCACCGAGATCTACAC- TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAG-NNNNNNN- ACACTCTTCCCTACACGACGCTCTT	MiSeq forward indexing primer	60
P7-RD2-Index-Rev adapter	CAAGCAGAAGACGGCATAACGAGAT- GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAG-NNNNNNN- GTGACTGGAGTTCAGACGTGT	MiSeq reverse indexing primer	60
RD1	TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAG	MiSeq standard sequencing primer	-

At step 30 (Gansauge and Meyer 2013), the libraries were amplified in quintuplicate with unique fusion-tag indexing primers either for the IonTorrent platform or the MiSeq platform (Table S5.9.3). The PCR reaction contained reagents in final concentrations of: 1X AccuPrime *Pfx* reaction mix (2.5 µl of a 10X stock), 0.4 µM *IDT* forward indexing primer (1 µl of a 10 µM stock), 0.4 µM *IDT* reverse indexing primer (1 µl of a 10 µM stock), 0.025 U/µl AccuPrime *Pfx* polymerase (0.25 µl of a 2.5 U/µl stock), and 1 µl of neat library, made up to a total of 25 µl final volume with HPLC-grade water. Thermalcycling conditions were: 95°C for two min, 20 cycles of 95°C for 15 sec, 60°C for 30 sec, 68°C for one min. Replicate reactions were combined and purified twice using the AMPure XP PCR purification kit as above (*Beckman-Coulter*). Step 33 was not performed. Libraries were quantified via qPCR as above. Libraries were pooled in equal amounts and the final shotgun-sequencing library was quantified once again via qPCR as above. Note that for libraries run on the MiSeq platform, the qPCR primers used were P5 and P7 (Table S5.9.2), and the library dilution series followed 1/10, 1/100, 1/1000, 1/5000, 1/25000, 1/125000, 1/625000 in EB buffer (*QIAGEN*), which were run alongside a with a standard dilution series of known concentration ( $10^8$ ,  $10^6$ ,  $10^5$ ,  $10^4$ ,  $10^3$ ,  $10^2$  molecules).

At step 34 (Gansauge and Meyer 2013), a standard sequencing primer was used (RD1 for the MiSeq; Table S5.9.3). Sequencing was performed using the IonTorrent PGM (*Life Technologies*; Murdoch University) and IonTorrent Proton (Lottery West) Sequencing 200 kit v.2 following the manufacturer's instructions. For the MiSeq (*Illumina*; Curtin University), 4.0 µl of the 1/100 dilution of the library was combined with 12 µl of EB buffer (*QIAGEN*), and 2 µl of 1M molecular biology-grade NaOH and incubated for five min at 25°C, then placed on ice. 10 µl of this mixture was then added to 990 µl of HT1 buffer, and placed on ice. 550 µl of this mixture was then combined with 50 µl of 20 pM denatured PhiX, and placed on ice. 600 µl of this mixture was added to the reagent cartridge in slot 17. The options selected when creating the sample sheet were: Other/ FASTQ only/ Sample preparation kit = TruSeq LT/ No index reads/ Single end/ 175 cycles/ No custom primer/ No trimming. The library was run using a MiSeq 150 v.3 kit.

The sequencing output was trimmed, filtered, and dereplicated as above (S5.9.4).

## S5.9.6 TARGETED ENRICHMENT *through* HYBRIDISATION CAPTURE

**TABLE S5.9.4** | GENBANK ACCESSION OF THE SEQUENCES THAT COMPRISED THE AVIAN BAITS.

<b>Species</b>	<b>Accession</b>
<i>Anas platyrhynchos</i>	EF521425
<i>Anas platyrhynchos</i>	EU737158
<i>Anas platyrhynchos</i>	EU737316
<i>Anas platyrhynchos</i>	EU737485
<i>Anas platyrhynchos</i>	EU737631
<i>Anas platyrhynchos</i>	EU737784
<i>Anas platyrhynchos</i>	EU737945
<i>Anas platyrhynchos</i>	EU738108
<i>Anas platyrhynchos</i>	EU738236
<i>Anas platyrhynchos</i>	EU738401
<i>Anas platyrhynchos</i>	EU738565
<i>Anas platyrhynchos</i>	EU738730
<i>Anas platyrhynchos</i>	EU738887
<i>Anas platyrhynchos</i>	EU739054
<i>Anas platyrhynchos</i>	EU739196
<i>Anas platyrhynchos</i>	EU739352
<i>Anas platyrhynchos</i>	EU739516
<i>Anas platyrhynchos</i>	EU739590
<i>Anas platyrhynchos</i>	EU739751
<i>Anas platyrhynchos</i>	EU739906
<i>Anas platyrhynchos</i>	EU805796
<i>Cacatua sulphurea</i>	EF521452
<i>Cacatua sulphurea</i>	EU737184
<i>Cacatua sulphurea</i>	EU737343
<i>Cacatua sulphurea</i>	EU737657
<i>Cacatua sulphurea</i>	EU737810
<i>Cacatua sulphurea</i>	EU737972
<i>Cacatua sulphurea</i>	EU738126
<i>Cacatua sulphurea</i>	EU738261
<i>Cacatua sulphurea</i>	EU738428
<i>Cacatua sulphurea</i>	EU738590
<i>Cacatua sulphurea</i>	EU738755
<i>Cacatua sulphurea</i>	EU738913
<i>Cacatua sulphurea</i>	EU739076
<i>Cacatua sulphurea</i>	EU739221

<i>Cacatua sulphurea</i>	EU739379
<i>Cacatua sulphurea</i>	EU739615
<i>Cacatua sulphurea</i>	EU739776
<i>Cacatua sulphurea</i>	EU739932
<i>Cacatua sulphurea</i>	JF414404
<i>Himantornis haematopus</i>	EF521497
<i>Himantornis haematopus</i>	EU737229
<i>Himantornis haematopus</i>	EU737392
<i>Himantornis haematopus</i>	EU737701
<i>Himantornis haematopus</i>	EU738021
<i>Himantornis haematopus</i>	EU738161
<i>Himantornis haematopus</i>	EU738308
<i>Himantornis haematopus</i>	EU738475
<i>Himantornis haematopus</i>	EU738636
<i>Himantornis haematopus</i>	EU738803
<i>Himantornis haematopus</i>	EU738962
<i>Himantornis haematopus</i>	EU739119
<i>Himantornis haematopus</i>	EU739267
<i>Himantornis haematopus</i>	EU739428
<i>Himantornis haematopus</i>	EU739664
<i>Himantornis haematopus</i>	EU739822
<i>Himantornis haematopus</i>	EU739977
<i>Himantornis haematopus</i>	EU740141
<i>Himantornis haematopus</i>	EU740305
<i>Himantornis haematopus</i>	JF496968
<i>Rallus limicola</i>	EF521546
<i>Rallus limicola</i>	EU302737
<i>Rallus limicola</i>	EU302780
<i>Rallus limicola</i>	EU737277
<i>Rallus limicola</i>	EU737445
<i>Rallus limicola</i>	EU737747
<i>Rallus limicola</i>	EU737905
<i>Rallus limicola</i>	EU738074
<i>Rallus limicola</i>	EU738361
<i>Rallus limicola</i>	EU738527
<i>Rallus limicola</i>	EU738685
<i>Rallus limicola</i>	EU738852
<i>Rallus limicola</i>	EU739015
<i>Rallus limicola</i>	EU739314
<i>Rallus limicola</i>	EU739479
<i>Rallus limicola</i>	EU739570

<i>Rallus limicola</i>	EU739713
<i>Rallus limicola</i>	EU740030
<i>Rallus limicola</i>	EU740193
<i>Sarothrura elegans</i>	EF521554
<i>Sarothrura elegans</i>	EU737285
<i>Sarothrura elegans</i>	EU737453
<i>Sarothrura elegans</i>	EU737596
<i>Sarothrura elegans</i>	EU737754
<i>Sarothrura elegans</i>	EU737913
<i>Sarothrura elegans</i>	EU738082
<i>Sarothrura elegans</i>	EU738212
<i>Sarothrura elegans</i>	EU738369
<i>Sarothrura elegans</i>	EU738535
<i>Sarothrura elegans</i>	EU738693
<i>Sarothrura elegans</i>	EU738859
<i>Sarothrura elegans</i>	EU739023
<i>Sarothrura elegans</i>	EU739172
<i>Sarothrura elegans</i>	EU739322
<i>Sarothrura elegans</i>	EU739487
<i>Sarothrura elegans</i>	EU739720
<i>Sarothrura elegans</i>	EU739875
<i>Sarothrura elegans</i>	EU740038
<i>Sarothrura elegans</i>	EU740201
<i>Sarothrura elegans</i>	EU740364
<i>Sarothrura elegans</i>	JF497012
<i>Struthio camelus</i>	AY277496
<i>Struthio camelus</i>	EF521423
<i>Struthio camelus</i>	EU302743
<i>Struthio camelus</i>	EU302786
<i>Struthio camelus</i>	EU737156
<i>Struthio camelus</i>	EU737314
<i>Struthio camelus</i>	EU737629
<i>Struthio camelus</i>	EU737782
<i>Struthio camelus</i>	EU737943
<i>Struthio camelus</i>	EU738399
<i>Struthio camelus</i>	EU738563
<i>Struthio camelus</i>	EU738886
<i>Struthio camelus</i>	EU739052
<i>Struthio camelus</i>	EU739194
<i>Struthio camelus</i>	EU739350
<i>Struthio camelus</i>	EU739514

<i>Struthio camelus</i>	EU739588
<i>Struthio camelus</i>	EU739749
<i>Struthio camelus</i>	EU739904
<i>Anas platyrhynchos</i>	NC 009684
<i>Balearica regulorum</i>	FJ769841
<i>Calyptorhynchus baudinii</i>	JF414242
<i>Casuarius casuarius</i>	NC 002778
<i>Dinornis giganteus</i>	NC 002672
<i>Eudyptula minor</i>	NC 004538
<i>Melopsittacus undulatus</i>	NC 009134
<i>Ninox novaeseelandiae</i>	NC 005932
<i>Porphyrio hochstetteri</i>	EF532934
<i>Struthio camelus</i>	NC 002785

---

All reagents and preparation for the capture took place in a designated ultraclean environment at Curtin University (TRACE facility).

100mer mitochondrial and nuclear avian baits with 50 bp tiling were previously designed using the sequences in Table S5.9.4 and manufactured by MYBaits (*MYcroarray*). Whole genome enrichment (WGE) baits (<http://www.mycroarray.com/mybaits/mybaits-WGE.html>) were previously synthesised from tinamou (*Nothocercus bonapartei*; ID ZMUC146563) genomic DNA (11 µg at 116 ng/µl) by Jean-Marie Rouillard (MYbaits, *MYcroarray*, Ann Arbor, MI), following their proprietary procedure. This specimen was a fresh skin sample from Zamora-Chinchipe, Ecuador (GPS coordinates -4.26667, -79.05) collected in 2000 by Jon Fjeldså. Genomic DNA was sheared, prepared, and globally transcribed into RNA baits using biotinylated rUTP (Enk et al. 2014). Baits were stored at -20°C. Tinamou baits were chosen because these were already synthesised and available for use at the time of the study.

Relaxed hybridisation captures were performed following Li (2013) and the *MYcroarray* MYbaits Sequence Enrichment for Targeted Sequencing kit as below.

In a 0.2 ml PCR tube (round capped), a solution was prepared containing final concentrations of 0.27 µg/µl Chicken Cot-1 (2.5 µl of a 1 µg/µl stock), 1.74 µM *IDT* forward blocking primer (MiSeq; Table S5.9.5; 0.08 µl of a 200 µM stock), 1.74 µM *IDT* reverse blocking primer (MiSeq; Table S5.9.5; 0.08 µl of a 200 µM stock), and added to 6.5 µl of library for a total of 9.16 µl. In a 0.2 ml, a solution was prepared containing final concentrations of 10.86X Hyb#1 (i.e., 20 µl of a 20X stock of Hyb#1 which is 20X SSPE), 0.01 M Hyb#2 (i.e., 0.8 µl of a 0.5 M stock of Hyb#2 which is 0.5M EDTA, pH 8.0), 10.86X Hyb#3 (i.e., 8 µl of a 50X stock of Hyb#3 which is Denhardt's solution), and 0.22% Hyb#4 (i.e., 8 µl of a 1% stock of Hyb#4, which is 10% SDS), for a final volume of 36.8 µl. In a 0.2 ml PCR tube, a solution was prepared containing final concentrations of 16.66X capture baits (5 µl of a 20X stock) and 3.33 U/µl RNase block (*SUPERase*; 1 µl of a 20 U/µl stock) for a final volume of 6 µl. Solutions were vortexed to mix and briefly spun in a bench-top microcentrifuge.

**TABLE S5.9.5** | CUSTOM BLOCKING PRIMERS (FOR SEQUENCING ON BOTH IONTORRENT AND MiSEQ PLATFORMS) USED IN THE ENRICHMENT PROTOCOL.

Name	Sequence (5'-3')	Function
IT_Blocking_F	CCATCTCATCCCTGCGTGTCTCCGACTCAGIIIIIIACACTCTTCCCT ACACGACGCTCTCC/3InvdT/	IonTorrent forward capture blocking primer
IT_Blocking_R	CCTCTCTATGGGCAGTCGGTGATIIIIIIIGTGACTGGAGTTCAGACGT GTGTGCTCTCCGATCT/3InvdT/	IonTorrent reverse capture blocking primer
MiSeq_Blocking_F	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCA GATGTGTATAAGAGACAGIIIIIIACACTCTTCCCTACACGACGCTCT T/3InvdT/	MiSeq forward capture blocking primer
MiSeq_Blocking_R	CAAGCAGAAGACGGCATAACGAGATGTCTCGTGGGCTCGGAGATG TGTATAAGAGACAGIIIIII GTGACTGGAGTTCAGACGTGTGCTCT/3InvdT/	MiSeq reverse capture blocking primer

The first solution was placed in a thermalcycler and incubated for five min at 95°C. The second solution was then placed in the thermalcycler and both solutions were incubated for three min at 65°C. The third solution was then placed in the thermalcycler and all solutions were incubated for two min at 65°C, after which 9 µl of the first solution and 13 µl of the second solution were transferred into the third solution to create a final volume of 28 µl. 10 µl of mineral oil was added on top of the reaction to prevent evaporation. The solution was incubated in the thermalcycler at 65°C for 11 hr, 60°C for 11 hr, 55°C for 11 hr, and finally 50°C for 11 hr.

Wash buffer 2 (0.1X SSC, 0.1% SDS) was pre-heated in a heat block to 45°C. 20 µl of M-270 Dynabeads (*ThermoFisher*) per reaction were pelleted with a magnetic rack and the supernatant was discarded. 40 µl per reaction of room-temperature Binding buffer (1 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA) was added to the beads, vortexed, and pelleted for two min on a magnetic rack, and the supernatant was discarded. This step was repeated two more times. The beads were resuspended in 40 µl per reaction of Binding buffer. 1 µl of 10% Tween-20 was added to the total bead suspension. 180 µl of Binding buffer was then added to a 1.5 ml *Eppendorf* tube (one tube per reaction), and 20 µl of the bead resuspension was added to this. The hybridisation solution was then added to the bead resuspension-Binding buffer solution and incubated at room temperature for 30 min with rotation.

The bead solution was pelleted with a magnetic rack for two min and the supernatant was discarded. 186 µl of Wash buffer 1 (1X SSC, 0.1% SDS) was added to the beads, pipetted up and down to resuspend, and incubated for 10 min at room temperature. Beads were pelleted and the supernatant was discarded. This step was repeated. 186 µl of pre-warmed Wash buffer 2 was added to the beads, pipetted up and down to resuspend, and incubated at 45°C for 10 min. This step was repeated two more times. Finally, 50 µl of freshly prepared elution buffer (100 mM NaOH) was added to beads, which were vortexed to mix, and incubated at room temperature for 20 min. Beads were pelleted and the supernatant was transferred to a clean 1.5 ml *Eppendorf* tube. Finally, 70 µl of Neutralisation buffer (1 M Tris-HCl pH 7) was added.

The final capture solution was purified and concentrated using *Agencourt* AMPure XP PCR purification kit (as above). The captured library was then PCR amplified. The PCR reaction contained final concentrations of 1X *AccuPrime Pfx* Reaction mix (5  $\mu$ l of a 10X stock), 0.025 U/ $\mu$ l *AccuPrime Pfx* polymerase (0.5  $\mu$ l of a 2.5 U/ $\mu$ l stock), 0.2  $\mu$ M forward primer (P5; Table S5.9.2; 1  $\mu$ l of a 10  $\mu$ M stock), 0.2  $\mu$ M reverse primer (P7; Table S5.9.2; 1  $\mu$ l of a 10  $\mu$ M stock), 37.5  $\mu$ l of Ultrapure water, and 5  $\mu$ l of capture library, for a final volume of 50  $\mu$ l. The PCR cycling conditions were: 98°C for 30 seconds, followed by 40 cycles of 98°C for 20 seconds, 60°C for 30 seconds, 72°C for 30, with a final extension of 72°C for five min.

The amplified captured library was purified using *Agencourt* AMPure XP PCR purification kit (as above), diluted, and quantified by qPCR for sequencing on the MiSeq platform as above. The sequencing output was trimmed, filtered, and dereplicated as above.

The number of sequences obtained after trimming for shotgun, enriched, and control libraries sequenced on both the Ion Torrent and MiSeq platforms are recorded in Table S5.9.6. The number of sequences obtained after quality filtering for shotgun, enriched, and control libraries sequenced on both the Ion Torrent and MiSeq platforms are recorded in Table S5.9.6. The number of final unique sequences obtained for shotgun, enriched, and control libraries sequenced on both the Ion Torrent and MiSeq platforms are recorded in Table S5.9.6. The FASTQ file for all filtered sequences can be downloaded from DataDryad. All non-enriched sequences were BLAST (see below) to assign taxonomy and determine the percentage of endogenous DNA obtained.

**TABLE S5.9.6** | THE NUMBER OF READS ABOVE 30 BP OBTAINED AFTER TRIMMING, QUALITY FILTERING, AND DEREPLICATION FOR VARIOUS LIBRARIES SEQUENCED ON BOTH THE ION TORRENT OR MISEQ PLATFORMS.

Dataset	Number of reads > 30 bp		
	After trimming	After quality filtering	Unique
Ion Torrent	24022737	8861650	4048430
Ion Torrent Controls	543	370	320
MiSeq	8601070	7764488	2857399
MiSeq-Mt Enriched	1103089	1078193	202412
MiSeq-Nu Enriched	1417989	1379820	355536
MiSeq Controls	37028	36006	34537

### S5.9.7 MITOCHONDRIAL GENOME ASSEMBLY

In Geneious, mitochondrial genomes of kiwi, emu, cassowary, ostrich, rhea, moa, and tinamou were downloaded from GenBank were aligned with default parameters by selecting “pairwise/multiple align” from the “Align/assemble” drop down menu. The consensus sequence was extracted from the alignment and used as reference upon which to map all reads. All trimmed, quality filtered, and unique reads above 30 bp were mapped iteratively (10 iterations) onto this consensus ratite mitochondrial genome using the Geneious Mapper under “Align/assemble” → “Map to reference”. A medium-low sensitivity/fast (word length 18) with default parameters was used. A strict consensus sequence (50% majority-ruled base calling) was generated. A list of mapped reads were saved in a sub-folder.

From Geneious, mapped sequences were batch renamed, exported in FASTA format, and BLAST through the Pawsey Centre supercomputing facilities, in order to obtain taxonomic assignments for the sequences by comparison to a custom aggregate database of NCBI’s GenBank database (Benson et al. 2006; non-redundant nucleotide; April 2015) and 32,000 unpublished *Struthio camelus* genomic contigs. The blastn program within Blast-2.2.30+ (Altschul et al. 1990) was evoked through a terminal window with the optional parameters: `-num_alignments 3 – num_descriptions 3 –reward 1 –evaluate 10`.

Next, blast files and FASTA files were imported into the software MEGAN v4.70.4 ([ab.inf.uni-tuebingen.de/data/software/megan4](http://ab.inf.uni-tuebingen.de/data/software/megan4); Huson et al. 2007) by selecting “import from BLAST” in the drop-down menu “File”. From the browser, the blast file created above was selected, and under the tab “LCA parameters”, 1 was chosen for minimum support, 35 was chosen for minimum score, and 10 was chosen for top percent. No minimum complexity filter was selected, or percent identity filter. Once the tree was created, the tree was collapsed at the order taxonomic level by selecting the option “collapse at taxonomic level” under the drop-down menu “Tree”.

Avian reads were extracted in FASTA format from the “Aves” node using the “Extract reads” option, as were “unassigned” reads and reads with “no hit”. The number of unique reads above 30 bp that were assigned to the order Aves for both

shotgun and enriched libraries are recorded in Table S5.9.7. Over 80% of shotgun avian reads assigned to Palaeognathae, with almost all of the remainder having a best hit to neognaths but most not aligning across the entire length of the query (i.e., they are probably avian, but do not match any known reference with high confidence). These reads were then re-mapped onto the consensus sequence from above as before, and a consensus sequence was generated as before for bases with a coverage greater than two. The number of unique reads above 30 bp from both shotgun and enriched libraries that mapped onto the mitochondrial genome are recorded in Table S5.9.7. Statistics of the read distributions can be found in Table S5.9.8. Mapped tRNA-Phe was used as a seed to complete assembly of the control region in MITObim (Hahn et al. 2013), using all unique sequences as the read pool. The final mitochondrial DNA sequence can be found on GenBank (accession number: KY412176) or downloaded from DataDryad (doi:10.5061/dryad.6h3q7).

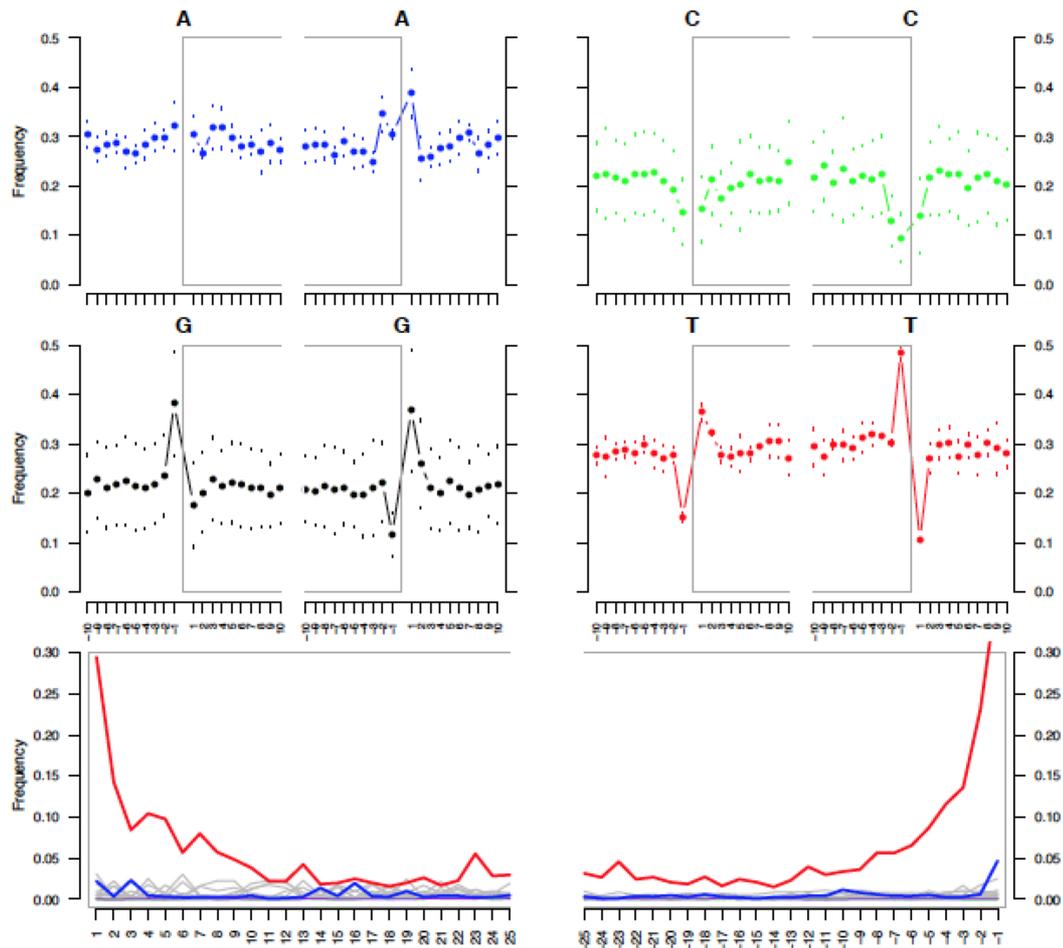
**TABLE S5.9.7** | THE NUMBER OF UNIQUE READS ABOVE 30 BP FOR SHOTGUN AND ENRICHED LIBRARIES THAT WERE ENDOGENOUS (AVIAN) AND MITOCHONDRIAL (MAPPED TO A MITOCHONDRIAL GENOME).

Dataset	Number of reads > 30 bp			
	Unique	Avian	Mt Mapped	% mtDNA
Not enriched	6628596	187044	3505	0.05
Enriched	557074	45236	6021	1.08
Total	7131929	228399	9478	0.13
Controls	34706	829	7	0.02

**TABLE S5.9.8** | STATISTICS OF THE READ DISTRIBUTIONS OF SEVERAL DATASETS INCLUDING THE MEAN FRAGMENT LENGTH AND PERCENTAGE OF HIGH QUALITY BASES.

<b>Dataset</b>	<b>Maximum length (bp)</b>	<b>Mean length (bp)</b>	<b>St Dev (bp)</b>	<b>% &gt; Q20</b>	<b>% &gt; Q30</b>
Enriched	100	50.70	15.60	99.80	98.70
Enriched mapped	100	59.10	16.40	99.70	98.40
Not enriched	185	50.80	17.40	97.20	52.50
Not enriched mapped	143	49.30	15.40	97.60	64.30
Total	185	50.90	17.30	97.30	55.80
Total mapped	143	55.20	16.70	98.90	86.10
Controls	99	50.60	16.00	99.90	98.80
Controls mapped	77	54.10	19.40	100.00	99.70

## S5.9.8 AUTHENTICITY



**FIGURE S5.9.2** | NUCLEOTIDE MISINCORPORATION PLOTS for avian reads mapped against the consensus sequence generated from the final round of mapping all reads against a ratite consensus mitochondrial genome. **Top panels:** the average frequency of bases across the ends of the read (enclosed by grey box). **Bottom panel:** the frequency of various substitutions across 25 bp from both the 5' and 3' ends are shown: C-to-T (red), G-to-A (blue), other (grey).

mapDamage 2.0.6 (Ginolhac et al. 2011; Jonsson et al. 2013) was used to chart the frequency of nucleotide substitutions across reads as a way of assessing aDNA authenticity. Mitochondrial mapped reads were exported from Geneious in SAM format, and the final consensus mitochondrial genome generated was exported in FASTA format. The files were placed into the same folder as the mapDamage program. From the mapDamage folder directory, mapDamage was evoked by typing the following command into a Terminal window: `mapDamage -i [input .sam file] -r [input .fasta reference file] -l 140`. Libraries built from single stranded DNA (Gansauge and Meyer 2013) are expected to show a characteristic accumulation of C-to-T misincorporations at both the 5' and 3' ends that is typical of authentic aDNA (Schroeder et al. 2015) as well as a higher frequency of purines immediately flanking the read. This pattern of greater damage at the ends of fragments was observed in the mitochondrial mapped reads (Figure S5.9.2), where the frequency of C-to-T misincorporations ranges between about 5%-30% within the first 10 bp from either end. This suggests that the DNA is likely to be of ancient origin.

#### **S5.9.9 IDENTIFICATION of NUCLEAR SEQUENCES**

Because there are fewer copies of nuclear loci compared with mitochondrial loci, less stringent criteria were used to find nuclear sequences. In order to identify orthologous, phylogenetically informative nuclear loci, unique reads were mapped, in Geneious v.7.1.2 to a custom database of *Struthio camelus* (ostrich) reference sequences containing 7976 nuclear protein-coding exons from Jarvis et al. (2014). Ostrich was used as a reference for mapping reads because more nuclear data is available for the ostrich than any other palaeognath. All trimmed, unique reads above 30 bp were mapped (no iterations) using the Geneious Mapper under “Align/assemble” → “Map to reference”. Custom sensitivity was used with the following parameters: word length 15 bp, allow gaps, minimum overlap 80 bp, ignore words repeated more than 12 times, maximum mismatches per read 30%, map multiple best matches randomly, maximum gap size 15 bp, index word length 13 bp, maximum ambiguity 4, and minimum mapping quality of 30 (which specifies a minimum confidence that the mapping is correct in order to map reads; a quality of 30 means 99.9% confidence the mapping is correct). The minimum coverage accepted was one. Hits were accepted as orthologues only if: (1) the read was 80 bp

**TABLE S5.9.9** | TAXON SAMPLING: species for which mitochondrial and nuclear data (various loci) are available. Where data were available only for a close relative, the name of that species is included.

Two-letter code	Four-letter code	Species name	Common name
CA	CASCA	<i>Casuarus casuarius</i>	Cassowary
DR	DRONO	<i>Dromaius novaehollandiae</i>	Emu
EM	EMECR	<i>Emeus crassus</i>	Eastern moa
AN	ANODI	<i>Anomalopteryx didiformis</i>	Little bush moa
DI	DINGI	<i>Dinornis giganteus</i>	Dinornis
CY	CRYTA	<i>Crypturellus tataupa (mt)/Crypturellus soui (nu)/Nothoprocta perdicaria (nu)</i>	Crypturellus
TI	TINAM	<i>Tinamus major (mt)/ Tinamus guttatus</i>	WTinamou/Gtinamou
EU	EUDEL	<i>Eudromia elegans</i>	Elegant crested tinamou
AH	APTHA	<i>Apteryx haastii</i>	Great spotted kiwi
AM	APTMA	<i>Apteryx australis mantelli</i>	Brown kiwi
AO	APTOW	<i>Apteryx owenii</i>	Little spotted kiwi
RH	RHEAM	<i>Rhea americana</i>	Great Rhea
PT	PTEPE	<i>Rhea pennata (Pterocnemia pennata)</i>	Lesser Rhea
ST	STRCA	<i>Struthio camelus</i>	Ostrich
EB	AEPMA	<i>Aepyornis maximus</i>	Elephant bird
AE	AEPHI	<i>Aepyornis hildebrandti</i>	Elephant bird
MU	MULAG	<i>Mullerornis agilis</i>	Elephant bird
AL	ALELA	<i>Alectura lathamii</i>	Brush turkey
GA	GALGA	<i>Gallus gallus</i>	Chicken
AS	ANSSE	<i>Anseranas semipalmata</i>	Magpie goose
AP	ANAPL	<i>Anas platyrhynchos</i>	Pekin duck
PY	PYGAD	<i>Pygoscelis adeliae</i>	Adelie penguin
ME	MELUN	<i>Melopsittacus undulatus</i>	Budgerigar
NE	NESNO	<i>Nestor notabilis/Strigops habroptilus (mt)</i>	Kea
PE	PELCR	<i>Pelicanus crispus/Eurypharanyx pelecanoides (mt)</i>	Dalmatian pelican
AC	ACACH	<i>Acanthisitta chloris</i>	Rifleman
TA	TAEGU	<i>Taeniopygia guttata</i>	Zebra finch
FU	FULGL	<i>Fulmarus glacialis/ Procellaria cinerea (mt)</i>	Fulmar

or longer and aligned across the entire length of the read; (2) the percent similarity for the alignment was higher than that of ostrich and chicken for the same locus; and (3) reciprocal comparison of the ostrich genome to elephant bird sequence set identifies the same read as its closest match (Vallender 2009). Although accepting a minimum read length of 50 bp or even 30 bp resulted in more loci being recovered (but not a significant increase in coverage per locus), an accepted length of 80 bp was chosen for stringency; such a conservative approach would have resulted in a rejection of a large number of potential nuclear loci, but also ensures that the loci chosen for further analysis are have the highest likelihood of being true orthologues. 112 hits were accepted. (Note that the majority of sequences were not mapped to nuclear loci because they fell below the 80 bp cut-off for accepting nuclear loci for analysis. In addition, a large proportion of sequences are expected to non-coding nuclear DNA, which would not have mapped to the reference coding sequences that were used. As such, we do not report the number of mapped nuclear reads because it would not be a true reflection of the amount of nuclear DNA obtained. Presumably all avian reads that did not map to the mitochondrial genome would be nuclear).

Unique reads were also aligned to a custom database of 44 nuclear protein-coding exons from Baker et al. (2014) and 67 nuclear protein-coding exons from Chojnowski et al. (2007), Hackett et al. (2008), Harshman (2008), Haddrath and Baker (2012), Smith et al. (2013) using blastn as implemented in Geneious with the following parameters: word size 7, gap cost 5, extension 2, low complexity filter, match scoring 2, mismatch scoring -3, maximum hits 5. A smaller 'minimum read length' was accepted for these loci because they are well-characterised, with the orthologous loci from the other palaeognaths also being described; as such, we were more confident that the elephant bird sequences retrieved for these loci were truly orthologous. 42 hits were accepted.

All the ostrich orthologues were then mapped against unpublished kiwi (Le Duc et al. 2015), emu, and rhea whole genome contigs to find the corresponding loci in these ratites. These sequences were then mapped back the elephant bird data set to establish orthology. Additionally, this 'reciprocal best hits' method helps overcome ascertainment bias that may have been inadvertently introduced by initial mapping to ostrich as opposed to another ratite; that is, without employing this method, we may

find only those elephant bird sequences that are most similar to ostrich, thereby biasing the dataset toward orthologues that more closely resemble ostrich over other ratites.

The species (at least one representative of every palaeognath family plus 11 neognath outgroups) for which mitochondrial and nuclear data were obtained are listed in Table S5.9.9. Loci/ predicted genes for which elephant bird orthologues were found are listed in Table S5.9.10, along with a data matrix showing the completeness of data sampling across taxa.

Coverage of the nuclear loci averaged 2.36X, with 52% of bases having a coverage of 1X (average quality Q31.1 with 94.3% of bases above Q20), 21% of bases having a coverage of 2X (average quality Q25 with 87.4% of bases above Q20), 15% of bases having a coverage of 3X (average quality Q23.2 with 80.8% of bases above Q20), and 12% of bases having a coverage of 4X or greater (average quality Q23.4 with 78.5% of bases above Q20). As such, the coverage was too low to accurately map the damage patterns. Of the sites with 1X coverage, 22.3% fall within 10 bp of fragment ends that typically contain a higher proportion of damaged bases; however, the average quality for these bases was above Q30.

#### **S5.9.10 ALIGNMENTS *and* PARTITIONING**

Elephant bird sequences for each locus were aligned in Geneious with the corresponding locus from 27 other birds (above) using the default parameters (Geneious Alignment, global alignment with free end gaps, 51% similarity (5.0/-3.0) cost matrix, gap open penalty 12, gap extension penalty 3). Nuclear multiple sequence alignments (MSA) were concatenated by dataset source (i.e., Jarvis et al. (2014) genes in another dataset).

MSAs were translated in MUSCLE (Edgar 2004) to find ORFs and refine the alignments. The final multiple sequence alignment was then imported into Se-AL v.2.0a (Rambaut 1996) for manual checking including codon positioning. Codon partitioning was performed in PAUP v.4.0b (Swofford 2003). Both mitochondrial

**TABLE S5.9.10 | DATA MATRIX:** mitochondrial and nuclear loci sampled for each taxon (other than the *Aepyornis* sp. sequenced in this study). Grey cells indicate data were available, white cell indicate no data were available. The superscript within each cell refers to the source of the sequence (see footnotes).

Locus/ Predicted Gene	Length (bp)	Taxon																											
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS	
Mitochondria	14596	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
ALDOB	46									2	2	2	2	2	2	2	2	2	1					1			1	1	
BACH1	42						1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	
IRBP	52						1	1	2	2	2	2	2	2	2	2	2	2		1			1	1	1		1	1	
PTPN12	78						1	1	2	2	2	2	2	2	2	2	2	2	1	1			1	1	1	1	1	1	
TRAF6	53						1	1	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	
ATMIN	48											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
BRS3	63											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
CCDC53	46											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
CHRM5	37											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
DACT1	58											4	4			6		4	1	1	4		1		4				
DCAF17	47											4		5	3		5	4	1		4	1	1	1	1	4	1		
DHX35	38											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
DTHD1	75											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		
E2F5	66											4	4	5	3	6	5	4	1	1	4	1	1	1	1	4	1		

Locus/ Predicted Gene	Length (bp)	Taxon																										
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS
ELFN1	58											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
EXOG	48											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
FAM131B	37											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
FUT8	80											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
HRH2	75											4	4		3	6	5	4	1	1	4	1		1	4	1		
HTR1D	62											4	4		3	6	5	4	1	1	4	1	1	1	4	1		
IFT80	38											4	4	5		6	5	4	1	1	4	1		1	4	1		
KCNMA1	47											4	4	5		6	5	4	1	1	4				4			
KNTC1	39											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
LANCL1	42											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
LNX1	39											4	4	5	3	6	5	4	1	1	4	1	1		4	1		
MAPKAPK5	52											4	4	5	3	6	5	4	1		4	1	1	1	4	1		
NDUFAF1	91											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
PDS5B	85											4	4		3	6	5	4	1	1	4			1	4			
POLE2	81											4	4	5	3	6	5	4	1	1	4	1	1		4	1		
PRELP	52											4	4		3	6	5	4	1	1	4	1	1	1	4	1		
PRPF4B	44											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1		
PUM2	85											4	4	5	3		5	4	1	1	4	1	1		4			
REV1	32											4	4	5	3		5	4	1	1	4	1	1	1	4	1		

Locus/ Predicted Gene	Length (bp)	Taxon																								
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU
SLU7	65											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1
SNAP47	37											4	4	5	3	6	5	4	1	1		1	1	1	4	1
TECRL	50											4		5	3		5	4	1	1	4	1	1	1	4	1
THSD1	66											4	4	5	3	6	5	4	1	1	4	1	1	1	4	1
TMEM181	41											4	4		3	6	5	4	1	1	4	1	1		4	
TREM2	50											4	4	5	3	6	5	4	1				1		4	
WDR17	59											4	4	5	3	6	5	4	1	1	4	1	1	5	4	1
ZMAT3	34											4	4	5		6	5	4	1	1	4	1	1	5	4	1
ZNF217	46											4	4	5	3	6	5	4	1	1	4	1	1	5	4	1
LOC104146248	111											3		5	3	6	5	5	5	5	5	5		5	5	5
NEO1	96											3		5	3	6	5	5	5	5	5	5	5	5	5	5
ANKLE2	103											3		5	3	6	5	5	5	5	5	5	5	5	5	5
ASPM	95											3		5	3	6	5	5	5	5	5	5	5		5	5
DUOX2	97											3		5	3	6	5	5	5	5	5	5	5	5	5	5
DNAH10	90											3		5	3	6	5	5	5	5	5	5	5	5		5
FAN1	106											3			3	6	5	5	5	5	5	5	5	5	5	5
CEP112	100											3		5	3	6	5	5	5	5	5	5		5	5	5
PHKB	93											3		5	3	6	5	5	5	5	5	5	5	5	5	5
GTF3C1	104											3			3	6	5	5	5	5	5	5	5	5	5	5

Locus/ Predicted Gene	Length (bp)	Taxon																											
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS	
ZDHHC16	93											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
HGF	91											3		5	3	6	5	5	5	5	5	5				5	5		
NDUFAF1	115											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
CCDC186	100											3			3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
SLC6A18	98											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
DHX29	94											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
TGFBR3	112											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
SPDL1	112											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
KLHL6	103											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
SUZ12	98											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
FAM20B	91											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
RASAL2	96											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
SUSD3	81											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ZGPAT	117											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
LOC106483813	94											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
TMEFF2	123											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
NDUFS1	117											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ZFYVE1	92											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
INTS7	97											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5

Locus/ Predicted Gene	Length (bp)	Taxon																										
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS
GLRA3	91											3			3	6	5	5	5	5	5	5	5	5	5	5	5	5
PTPRQ	99											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
DICER1	95											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TRMT61A	100											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TRAK1	90											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
ZEB2	92											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
JARID2	70											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TMF1	98											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
MRPL47	95											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
KIAA2018	101											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
ZNF521	95											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
CHD7	89											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
LOC104137995	74											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TRPA1	90											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
LOC104137995	101											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TBC1D8	92											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TNFRSF11B	97											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
AIM1	95											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
FN1	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5

Locus/ Predicted Gene	Length (bp)	Taxon																											
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS	
PAPPA2	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5		
DCAF12	88											3		5	3	6	5	5	5	5	5	5	5			5	5	5	
FAM69A	84											3		5	3	6	5	5	5	5	5	5	5	5				5	
LOC104138152	82											3		5	3	6	5	5	5	5	5	5	5					5	
STAG2	85											3		5	3	6	5	5	5	5	5	5	5	5			5		
LOC104153980	81											3		5	3	6	5	5	5	5	5	5	5	5				5	
CRLS1	85												5			6	5	5	5	5	5	5	5	5	5	5	5	5	5
PEX1	80											3		5	3	6	5	5	5	5	5	5	5				5	5	
LOC104153847	81											3		5	3	6	5	5	5	5	5	5				5			
STRA8	83											3			3	6	5	5	5	5	5	5			5	5	5	5	5
POLE2	81											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
PLA2G6	81											3		5	3	6	5	5			5	5	5	5	5	5	5	5	5
PDE7B	89											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
LOC104149548	80											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ATP7B	89											3		5	3	6	5	5	5	5	5	5	5	5	5	5			5
CEP57	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
CLN8	94											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ROS1	80											3			3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ODF2L	83											3		5	3	6	5	5	5	5	5	5			5	5	5	5	5

Locus/ Predicted Gene	Length (bp)	Taxon																										
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS
GPR56	88											3			3	6	5	5	5	5	5	5	5	5	5	5	5	5
ARHGAP25	84											3		5	3	6	5	5	5	5				5		5	5	5
SLC23A2	84											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
VPS26B	107											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
PUSL1	118											3		5	3	6	5	5	5	5	5			5		5	5	
NAB1	87											3		5	3	6	5	5	5	5	5	5	5		5		5	
HEMK1	84											3		5	3	6	5	5			5	5	5	5	5	5	5	5
GOT2	88											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
RUSC2	97											3			3	6	5	5	5	5	5	5	5	5	5	5	5	5
PPP1R26	86											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
AACS	82											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
BRD8	87											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
TET1	99											3		5	3	6	5	5	5	5	5					5	5	
CCAR1	82											3		5	3	6	5	5	5	5	5	5	5		5		5	
TTPAL	80											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
RCAN3	81											3			3	6	5	5	5	5	5	5	5	5	5		5	
CARD11	84											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
FAM214A	90											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5
COPB1	89											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5

Locus/ Predicted Gene	Length (bp)	Taxon																												
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS		
FRA10AC1	83											3		5	3	6	5	5	5	5	5	5		5	5	5	5			
PLSCR5	86											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5		
LOC104152342	83											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5		
ANKDD1A	81											3		5	3	6	5	5	5	5	5	5	5	5		5		5		
AOX1	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
SP3	80											3			3	6	5	5	5	5	5	5	5	5		5	5	5		
PDZRN4	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
GPR176	97											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
GIMD1	90											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
CEP170	80											3		5	3	6	5	5	5	5	5	5		5		5		5	5	
GRIN2B	83											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
LATS1	84											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
TNS3	85											3		5	3	6	5	5	5	5	5	5	5			5		5		
CCDC53	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
FAM65B	81											3		5	3	6	5	5	5	5	5	5	5	5	5		5	5		
FAM65B	80											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
CEP76	84											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
PTPN6	88											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	
BACH2	85											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5	

Locus/ Predicted Gene	Length (bp)	Taxon																											
		AE	MU	EM	DI	CY	AH	AO	PT	EU	CA	DR	AN	TI	RH	AM	ST	GA	AP	PY	ME	NE	PE	AC	TA	FU	AL	AS	
ZFH4	80											3		5	3	6	5	5	5	5	5	5	5		5		5		
IL10RB	87											3		5	3	6	5	5	5	5	5	5	5	5		5		5	
LRP12	88											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
ZRSR2	89											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
CD14	86											3		5	3	6	5	5	5	5	5	5		5	5	5	5	5	
ZBTB1	84											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5
EDARADD	81											3		5	3	6	5	5	5	5	5	5	5	5	5	5	5	5	5

- <sup>1</sup> Sequence available in GenBank
- <sup>2</sup> Genes listed in Haddrath and Baker (2012); available in GenBank
- <sup>3</sup> Sequences obtained from Gilbert (et al.); unpublished
- <sup>4</sup> Genes listed in Baker and Haddrath (2014); available through DataDryad
- <sup>5</sup> Genes listed in Jarvis et al. (2014); available through GenBank or through GigaScience
- <sup>6</sup> Genes listed in Le Duc et al. (2015); available through GenBank

and nuclear protein-coding genes were partitioned into 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> codon positions, while mitochondria were further partitioned into stems and loops, resulting in 11 partitions in total. The differently-sourced nuclear sequences were partitioned separately because they differ in taxonomic composition (Table S5.9.10). That is, Jarvis et al. (2014) genes were partitioned in first, second, and third codon positions (n1, n2, n3, respectively; Table S5.9.11), and Haddrath and Baker (2012) and Baker et al. (2014) genes were partitioned into first, second, and third codon positions (hb1, hb2, hb3, respectively; Table S5.9.11). Multiple sequence alignments can be downloaded from DataDryad (doi:10.5061/dryad.6h3q7) in fasta format.

### **S5.9.11 PHYLOGENETIC ANALYSIS**

**RCV ANALYSIS.** The partitioned MSA was imported into PAUP v.4.0b (Swofford 2003). Relative composition variability (RCV; Phillips et al. 2010) for each partition (both nucleotide and RY-coded) was estimated by exporting base frequencies from PAUP and performing calculations in *Microsoft's* Excel (Table S5.9.11). RCV is calculated as the sum of the absolute deviation of each base frequency for each taxon from the average across all taxa, divided by the number of taxa (Phillips et al. 2010). For mitochondrial partitions, data from all 28 taxa were included, while for nuclear partitions, only the 16 taxa with nuclear information were included (note that mitochondrial partitions were also reanalysed with only the 16 taxa for which nuclear data were also available).  $\chi^2$  P-values below 0.05 indicate significant base composition bias for that partition; however, this may not be particularly meaningful as the power of this test is highly influenced by the number of sites included (Phillips et al. 2010). Rather, a reduction in RCV estimates with RY-coding indicates that base composition bias is somewhat mitigated by RY-coding. RCV estimates fall for M1, M3 and loops mitochondrial partitions upon RY-coding, whether data from 28 taxa or 16 taxa were used to generate estimates (Table S5.9.11), indicating that these positions may benefit from being RY-coded. No nuclear partitions exhibited significant base composition bias, indicating that RY-coding these partitions is unnecessary.

**TABLE S5.9.11** | RCV CALCULATIONS for both nucleotide and RY-coded partitions for when all taxa are included versus when only 16 taxa (for which both nuclear and mitochondrial data are available) are included.

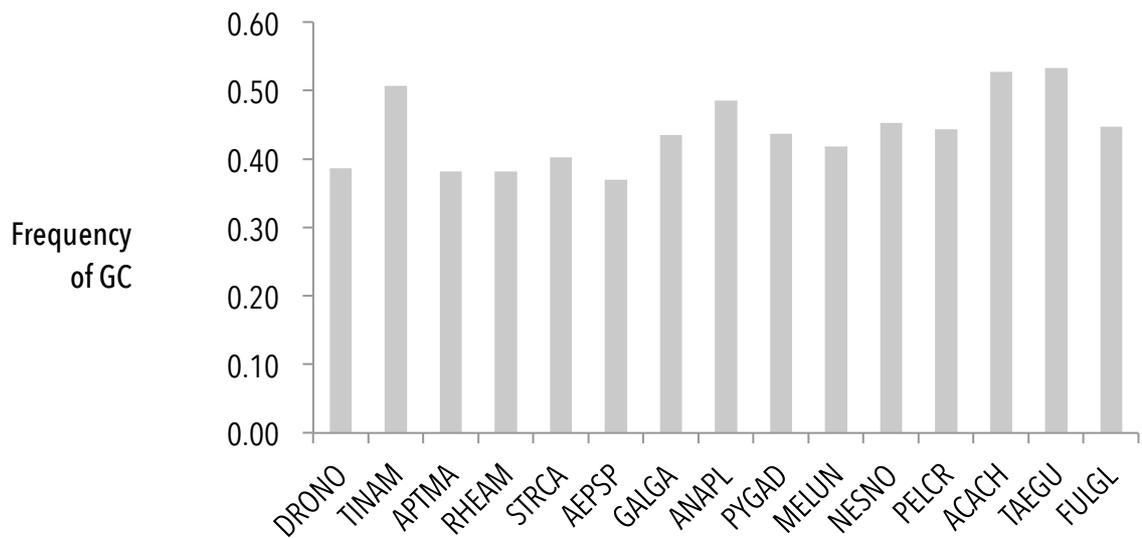
Partition	All taxa				16 taxa			
	RCV	$\chi^2$ P-value	RCV-RY	$\chi^2$ P-value	RCV	$\chi^2$ P-value	RCV-RY	$\chi^2$ P-value
m1	0.06	0.00	0.02	0.61	0.07	0.00	0.03	0.41
m2	0.03	1.00	0.08	1.00	0.04	0.97	0.03	0.98
m3	0.10	0.00	0.05	0.00	0.10	0.00	0.04	0.00
stem	0.06	0.87	0.07	0.98	0.05	0.75	0.06	0.97
loop	0.08	0.00	0.07	0.02	0.07	0.00	0.06	0.28
n1	NA	NA	NA	NA	0.10	1.00	0.08	0.96
n2	NA	NA	NA	NA	0.10	1.00	0.09	0.94
n3	NA	NA	NA	NA	0.09	0.15	0.05	0.99
hb1	NA	NA	NA	NA	0.08	1.00	0.07	1.00
hb2	NA	NA	NA	NA	0.07	1.00	0.07	1.00
hb3	NA	NA	NA	NA	0.07	1.00	0.09	1.00
all nuc	NA	NA	NA	NA	0.06	0.19	0.04	0.86
all mt	0.07	0.00	0.04	0.00	0.07	0.00	0.03	0.00
all	NA	NA	NA	NA	0.07	0.00	0.03	0.00

The nuclear GC content is very similar (and low) among palaeognaths, except for tinamous, for which it is higher, like for many neognaths (Figure S5.9.3). This compositional bias is clearly overwhelmed by phylogenetic signal for tinamous grouping within palaeognaths. GC content often obscures the primary compositional bias in mitochondrial DNA, which instead is the relative composition of T vs C or A vs G.

Unsurprisingly, given the lower phylogenetic signal retention (stemminess), the mitochondrial DNA is more susceptible to composition bias misleading phylogeny. This is already covered for palaeognaths by base frequency distance tree analyses in Phillips et al. (2010). Nevertheless, it is notable that the mitochondrial DNA C/T variation pattern (Table S5.9.13, Figure S5.9.4) is consistent with the standard nucleotide incongruence with all other analyses, with rheas and moa (plus tinamous) being attracted towards the neognaths. RY coding the mitochondrial third codon positions eliminates that C/T variation and allows the ostrich to fall deepest among palaeognaths in agreement with the nuclear results and retroposon data (see Baker et al. 2014).

**TABLE S5.9.12** | NUCLEAR THIRD CODON POSITION (N3, HB3) BASE COMPOSITIONS ACROSS TAXA.

		Base Frequency					
	Taxon	A	C	G	T	GC	Number of sites (bp)
1	DRONO	0.30	0.21	0.18	0.32	0.39	884
2	TINAM	0.24	0.28	0.23	0.25	0.51	884
3	APTMA	0.30	0.21	0.17	0.32	0.38	884
4	RHEAM	0.30	0.21	0.18	0.32	0.38	884
5	STRCA	0.29	0.21	0.19	0.31	0.40	884
6	AEPSP	0.31	0.21	0.16	0.32	0.37	884
7	GALGA	0.26	0.24	0.20	0.30	0.44	884
8	ANAPL	0.26	0.27	0.21	0.26	0.49	884
9	PYGAD	0.27	0.24	0.20	0.30	0.44	884
10	MELUN	0.27	0.22	0.20	0.31	0.42	884
11	NESNO	0.25	0.23	0.22	0.30	0.45	884
12	PELCR	0.26	0.24	0.21	0.30	0.44	884
13	ACACH	0.21	0.29	0.24	0.26	0.53	884
14	TAEGU	0.21	0.29	0.25	0.26	0.53	884
15	FULGL	0.26	0.24	0.20	0.29	0.45	884

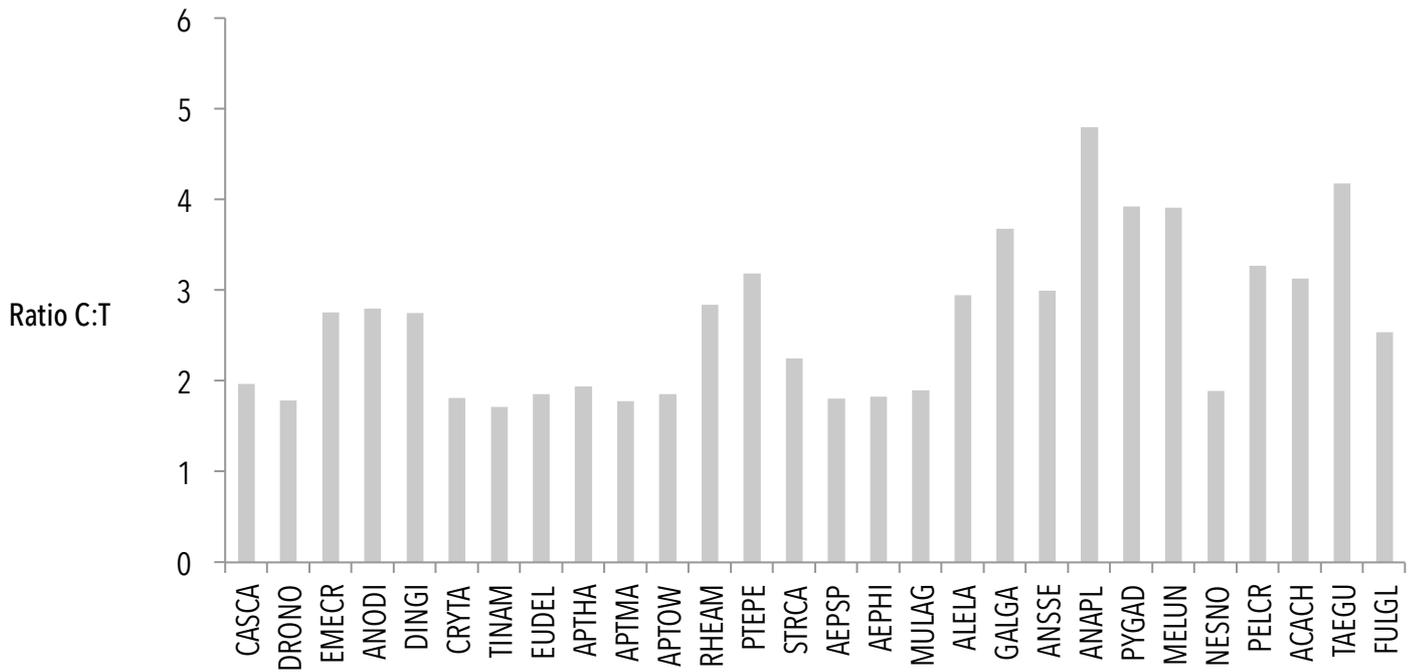


---

**FIGURE S5.9.3** | GC CONTENT OF NUCLEAR THIRD CODON POSITION (N3, HB3) ACROSS TAXA.

**TABLE S5.9.13** | MITOCHONDRIAL THIRD CODON POSITION (M3) BASE COMPOSITIONS.

		Base Frequency				Ratio of C:T	Number of sites (bp)
		A	C	G	T		
1	CASCA	0.39	0.37	0.05	0.19	1.97	3428
2	DRONO	0.41	0.35	0.04	0.20	1.79	3428
3	EMECR	0.39	0.42	0.04	0.15	2.76	3428
4	ANODI	0.39	0.42	0.04	0.15	2.80	3428
5	DINGI	0.39	0.41	0.05	0.15	2.75	3428
6	CRYTA	0.35	0.39	0.04	0.22	1.82	3428
7	TINAM	0.37	0.38	0.03	0.22	1.71	3428
8	EUDEL	0.35	0.41	0.02	0.22	1.86	3428
9	APTHA	0.41	0.37	0.03	0.19	1.94	3428
10	APTMA	0.41	0.36	0.03	0.20	1.78	3428
11	APTOW	0.40	0.36	0.04	0.20	1.85	3428
12	RHEAM	0.33	0.45	0.06	0.16	2.84	3428
13	PTEPE	0.33	0.47	0.06	0.15	3.18	3428
14	STRCA	0.38	0.39	0.06	0.17	2.25	3428
15	AEPSP	0.40	0.36	0.03	0.20	1.81	3428
16	AEPHI	0.39	0.37	0.04	0.20	1.82	3428
17	MULAG	0.41	0.37	0.03	0.19	1.90	3428
18	ALELA	0.34	0.44	0.08	0.15	2.95	3428
19	GALGA	0.39	0.46	0.04	0.12	3.68	3428
20	ANSSE	0.40	0.41	0.05	0.14	3.00	3428
21	ANAPL	0.36	0.46	0.08	0.10	4.80	3428
22	PYGAD	0.40	0.46	0.03	0.12	3.92	3428
23	MELUN	0.42	0.45	0.02	0.11	3.91	3428
24	NESNO	0.42	0.36	0.03	0.19	1.89	3428
25	PELCR	0.37	0.43	0.07	0.13	3.27	3428
26	ACACH	0.36	0.44	0.06	0.14	3.13	3428
27	TAEGU	0.43	0.43	0.04	0.10	4.18	3428
28	FULGL	0.40	0.40	0.04	0.16	2.54	3428



**FIGURE S5.9.4** | RATIO OF C:T FREQUENCY OF MITOCHONDRIAL THIRD CODON POSITION (M3) ACROSS TAXA.

**STEMMINESS.** Stemminess is a measure of the amount of phylogenetic signal erosion, and can be used alongside RCV to determine whether RY-coding certain partitions will mitigate biases in the data (i.e., “compositional heterogeneity”) that may lead to incorrect phylogenetic inference (Phillips 2008; Phillips et al. 2010). Stemminess is calculated as the sum of the internal branch lengths of a tree divided by the sum of internal and external branch lengths, and high stemminess indicates less saturation of phylogenetic signal that may alleviate some of the bias in base composition.

Branch lengths were generated for each partition using a null hypothesis tree and the MSA with PAUP and were imported into Excel in order to calculate uncorrected stemminess for each partition (both nucleotide and RY-coded). RY-coding all mitochondrial partitions improves the stemminess, but the stemminess of the mitochondrial third codon position (M3) is most greatly improved by RY-coding (stemminess increases from 0.17 to 0.25; Table S5.9.14). Although stemminess appears to increase with RY-coding nuclear partitions, too much phylogenetic information is lost by RY-coding nuclear partitions.

Uncorrected stemminess gives an indication of which partitions may require RY-coding; however, it is the corrected stemminess values (Table S5.9.15) that are more informative for making coding decisions. Corrected stemminess was calculated for each partition (both nucleotide and RY-coded) of the MSA after tree generation with MrBayes (Bayesian) in order to determine whether the model appropriately corrects substitution saturation. MrBayes parameters were as follows: rates=inv gamma, mcmc ngen = 10,000,000, sample freq = 5000, nchains = 3. MrBayes was run using the online bioinformatics platform CIPRES (Miller 2010). The stemminess was determined from the MrBayes output that was imported into TreeStat v.1.2 (<http://tree.bio.ed.ac.uk/software/treestat/>; Rambaut, 2007; Table S5.9.15). These corrected stemminess results show that RY coding still appears to enhance the signal on internal branches in Bayesian inference. However, only the mtDNA 1<sup>st</sup> and 3<sup>rd</sup> position corrected stemminess 95% HPDs fall outside (lower) than those ranges for the nuclear partitions. Among these two mtDNA partitions it is the 3<sup>rd</sup> positions that have by far the higher base compositional heterogeneity and hence, these present the strongest case for RY-coding. Furthermore, this scheme was favoured by Mitchell et

**TABLE S5.9.14** | UNCORRECTED STEMMINESS VALUES for both nucleotide and RY-coded partitions when all taxa are included versus when only 16 taxa (for which nuclear and mitochondrial data are available) are included.

Partition	All taxa		16 taxa	
	Stemminess	Stemminess-RY	Stemminess	Stemminess-RY
m1	0.21	0.28	0.07	0.14
m2	0.24	0.28	0.12	0.18
m3	0.17	0.25	0.04	0.10
stems	0.22	0.28	0.11	0.20
loops	0.20	0.28	0.08	0.16
n1	NA	NA	0.18	0.14
n2	NA	NA	0.18	0.32
n3	NA	NA	0.16	0.25
hb1	NA	NA	0.10	0.15
hb2	NA	NA	0.10	0.13
hb3	NA	NA	0.14	0.22
all mt	0.18	0.26	0.06	0.11
all nuc	NA	NA	0.15	0.21
all	NA	NA	0.06	0.12

**TABLE S5.9.15** | CORRECTED STEMMINESS VALUES FOR BOTH NUCLEOTIDE AND RY-CODED PARTITIONS.

Partition	Stemminess			Stemminess-RY		
	Average	Upper	Lower	Average	Upper	Lower
m1	0.13	0.14	0.11	0.18	0.21	0.15
m2	0.18	0.21	0.16	0.22	0.27	0.17
m3	0.15	0.16	0.13	0.16	0.18	0.14
stems	0.17	0.20	0.15	0.20	0.29	0.12
loops	0.17	0.20	0.14	0.22	0.27	0.18
n1	0.20	0.22	0.17	0.21	0.26	0.16
n2	0.20	0.23	0.17	0.29	0.35	0.23
n3	0.22	0.24	0.20	0.24	0.28	0.21

al. (2014), and thus our results are comparable. Phillips et al. (2010) additionally examined palaeognath trees based on base compositional frequencies and found that the major concern was mitochondrial third codon positions. However, all of the mitochondrial partitions show improved corrected stemminess with RY coding, although even with RY-coding the mtDNA 3<sup>rd</sup> positions HPD falls well below the nuclear 2<sup>nd</sup> and 3<sup>rd</sup> positions. In another treatment, we thus exclude mtDNA 3<sup>rd</sup> positions, but RY code all of the others. For each the nuclear partitions, the 95% HPDs for corrected stemminess overlap between the standard and RY-coded results, further justifying why nuclear partitions were not RY-coded.

**MODELTEST.** jModelTest v.2.1.7 (Guindon 2003; Darriba 2012) was run through PAUP v.4.0b (Swofford 2003) to find the best-fitting substitution model for each partition. Substitution models for each partition are shown in Table S5.9.16.

**RAXML PARAMETERS.** An input .phy file where the RY coded partitions were modified to binary characters (i.e., R = 0, Y = 1, ambiguous characters = -) was run in RaxML v.1.5 (Stamatakis 2014) with the following parameters: Data type = DNA or mixed; Analysis / Multiple outgroups; Import partitions.txt (DNA for nt coded, BIN for ry coded); Auto mre = 500; BS brL; GTRGammI or BinGammI (for only RY coded). The output was a 'bipartitions.tre' file.

**MRBAYES PARAMETERS.** MrBayes v.3.2.6 (Huelsenbeck and Ronquist 2001) was implemented via the online bioinformatic CIPRES science gateway (Miller 2010). Bayesian inference analyses were run with unlinked substitution models and branch-length rate multipliers among the partitions (Phillips 2008). Three MCMC chains for each of two independent runs proceeded for 10<sup>7</sup> generations with trees sampled every 2500 generations. The parameters specified in the input NEXUS file were as follows: Ntax = number of taxa (16 only for the nuclear dataset); Nchar = number of characters (this depends on the number of partitions included or excluded); Format datatype = dna; Gap = -; Interleave = yes; Partition by codon = list of the partition names included; Set partition = bycodon; Lset applyto= [partition #]; nst=6 or 2; rates=invgamma/gamma [determined from ModelTest]; Outgroup GALGA; Prset applyto = all; statefreqpr = fixed (empirical); Unlink statefreq=all; Unlink revmat=all; Unlink pinvar=all; Unlink brlens=all; Link brlens= [certain partitions];

**TABLE S5.9.16 | MODEL TEST RESULTS.**

Partition	16 taxa		All taxa		Model used
	hLRT	AIC	hLRT	AIC	
m1	GTR+I+G	GTR+I+G	GTR+I+G	GTR+I+G	GTR+I+G (CF87+I+G for RY)
m2	HKY+I+G	TVM+I+G	HKY+I+G	TVM+I+G	GTR+I+G (CF87+I+G for RY)
m3	HKY+I+G	GTR+I+G	TrN+I+G	GTR+I+G	GTR+I+G (CF87+I+G for RY)
stems	HKY+I+G	SYM+I+G	HKY+I+G	SYM+I+G	GTR+I+G (CF87+I+G for RY)
loops	TrN+I+G	GTR+I+G	HKY+I+G	SYM+I+G	GTR+I+G (CF87+I+G for RY)
n1	TIM+I+G	GTR+I+G	NA	NA	GTR+I+G
n2	TrN+I+G	GTR+I+G	NA	NA	GTR+I+G
n3	HKY+G	SYM+G	NA	NA	GTR+G
hb1	K81uf+I+G	K81uf+I+G	NA	NA	GTR+G
hb2	HKY+G	HKY+G	NA	NA	GTR+G
hb3	HKY+G	HKY+G	NA	NA	HKY+G

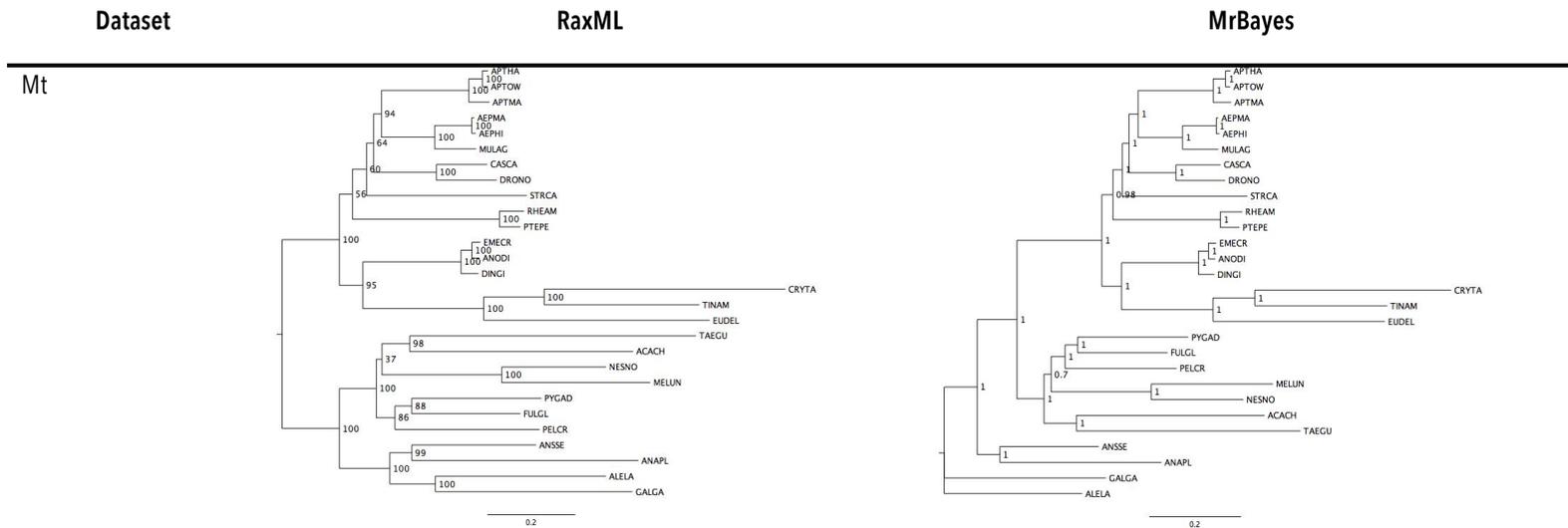
Link brlens= [other partitions]; Prset applyto = [some partitions]; ratepr= variable; Prset applyto = [other partitions]; ratepr= variable; Mcmc ngen = 10000000; printfreq=5000; samplefreq=2500; nruns=2; nchains=3; temp=0.1; diagnfreq=5000; savebrlens=yes; filename=[filename]; SUMT filename= [filename]; relburnin=yes; burnin=0.2; SUMP filename= [filename]; relburnin=yes; burnin=0.2. The output was a 'con.tre' file.

**CONVERGENCE.** The burn-in for each MrBayes run ensured that  $-\ln L$  had plateaued, clade frequencies had converged between runs (i.e., the average standard deviation of split frequencies was  $<0.02$ , approaching 0), and that PRSF (potential scale reduction factors, a convergence diagnostic) approached 1 (Phillips 2008). Output files were then opened in Tracer v.1.6.1 (Rambaut 2003) to ensure that estimated sample sizes (ESS) for substitution parameter estimates were above 200, and that parameter probability estimates had overlapping distributions, further confirming that runs had converged.

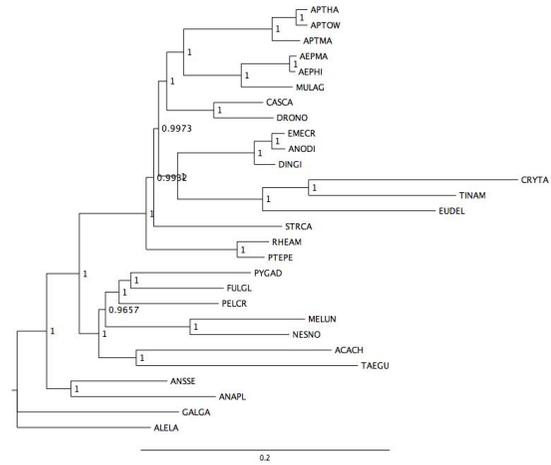
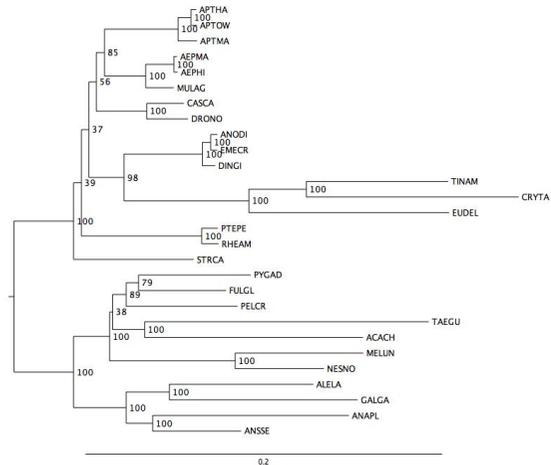
**CONGRUENCE.** Using the program CONSEL v.0.20 (Shimodaira 2002; Shimodaira and Hasegawa 2001) and the nuclear DNA, we statistically compared (using the approximately unbiased test or AU test) three palaeognath trees: the tree favoured by the nuclear data ( $-45506.9151 \ln L$  units), the tree favoured by the standard coded mitochondrial data ( $-45527.6659 \ln L$  units), and the tree favoured by the RY-coded mitochondrial data ( $-45506.9715 \ln L$  units). Note the mtDNA with either the third codon positions RY-coded or fully RY-coded favours the topology as the combined data analyses. The nuclear topology does not significantly differ from that of the tree favoured by the mitochondrial RY-coded data, differing by only  $0.0564 \ln L$  units ( $p = 0.589$ ), whereas the tree favoured by the standard-coded mitochondrial data is  $20.7518 \ln L$  units less likely than the nuclear tree, and is strongly rejected ( $p = 0.034$ ). Thus, the nuclear data is significantly incongruent with the standard-coded mitochondrial data, but is very closely congruent with the RY-coded mitochondrial data. This result suggests that combining the nuclear data with the RY-coded mitochondrial data in a partitioned analysis is appropriate.

**PHYLOGRAMS.**

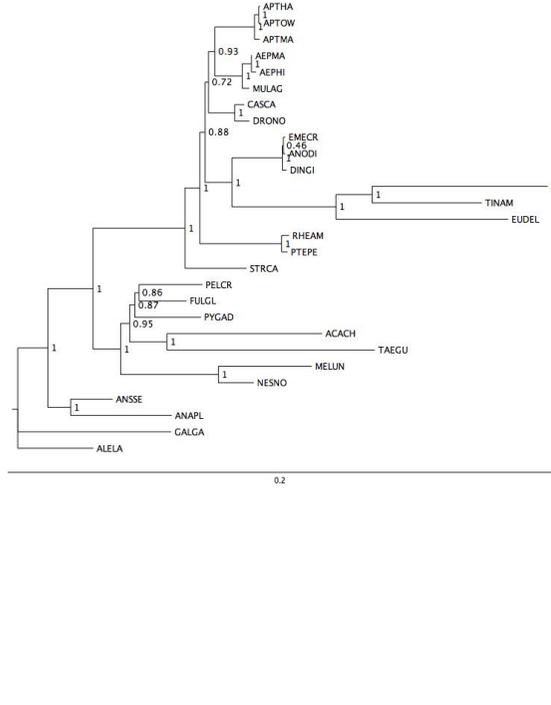
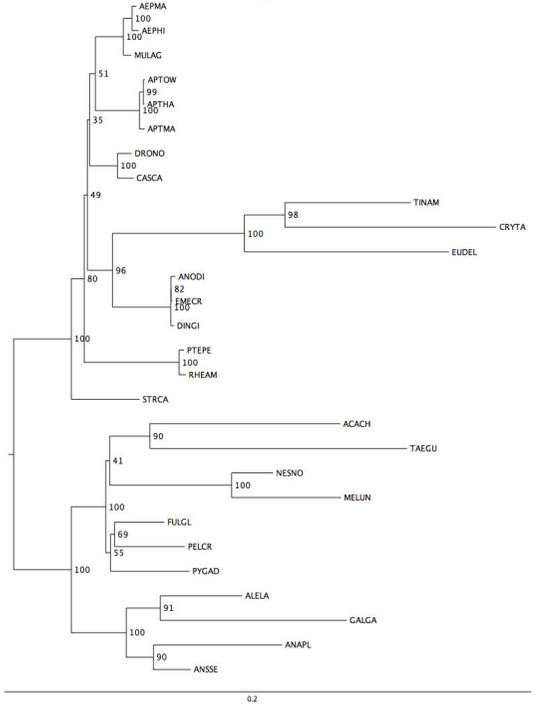
**TABLE S5.9.17** | PHYLOGRAMS FROM BOTH MAXIMUM LIKELIHOOD AND BAYESIAN ANALYSES OF VARIOUS NON-SITE-STRIPPED DATASETS. Bootstrap values (RaxML) or posterior probabilities (MrBayes) for each node appear as the number at the node. Scale refers to genetic distance. Blue highlighted rows are the three datasets believed to be the most accurate and, as such, are the focus of comparisons and discussion in the main text.



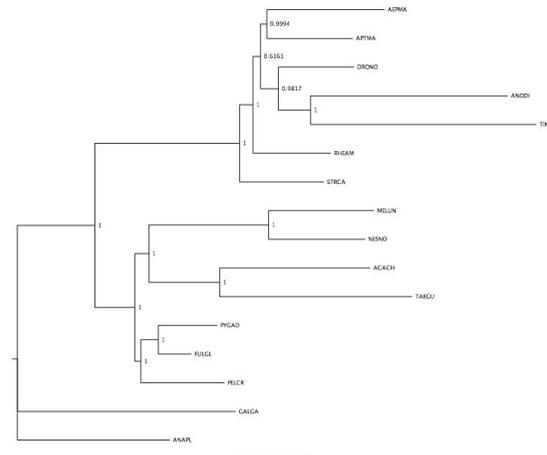
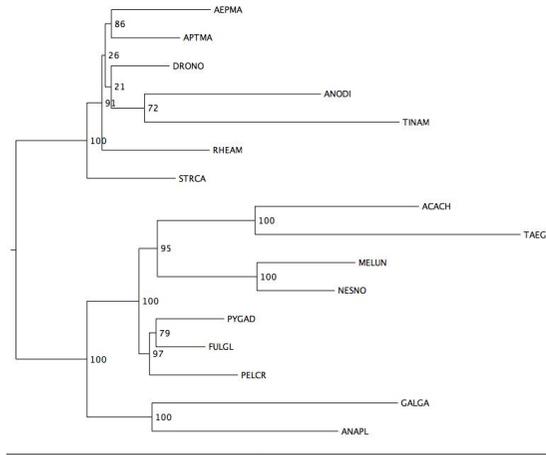
Mt  
(M3 RY)



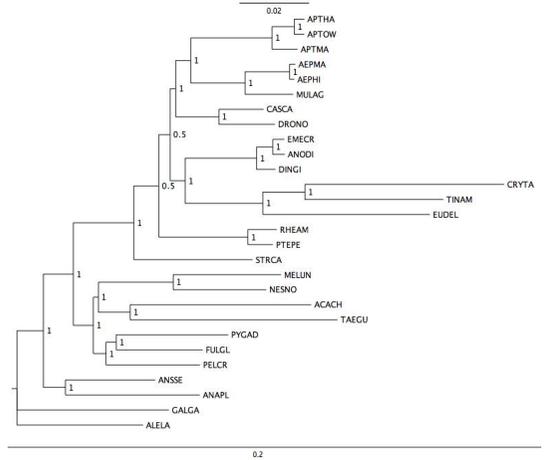
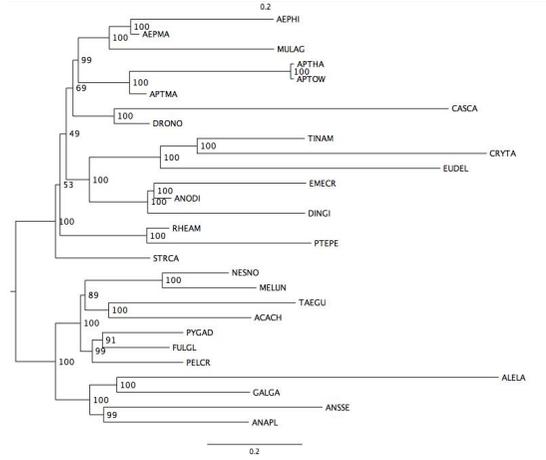
Mt  
(RY, M3 excluded)



Nuc



Nuc+Mt





## S5.9.12 MOLECULAR DATING

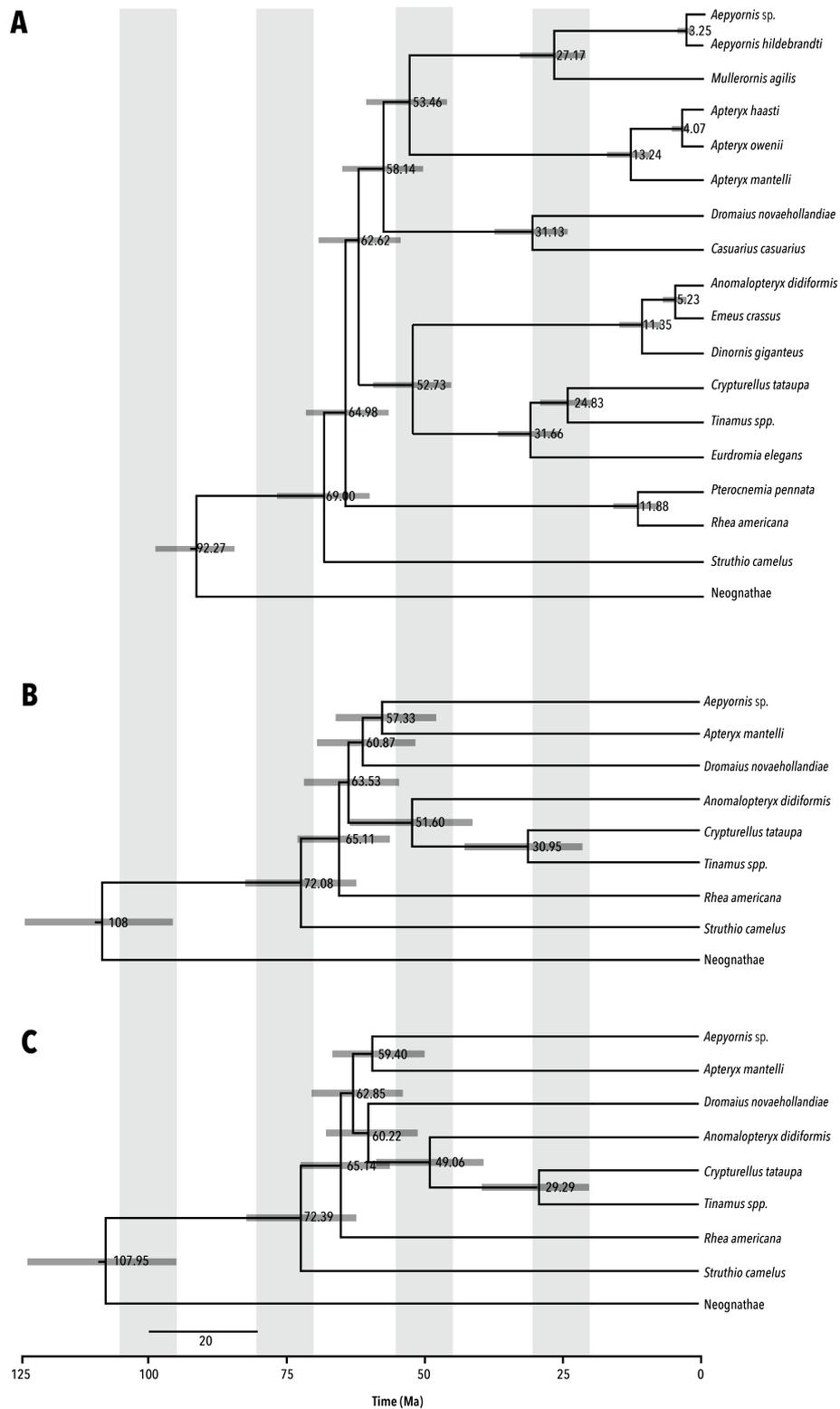
### FOSSIL CALIBRATIONS.

**TABLE S5.9.18** | FOSSIL CALIBRATIONS USED FOR MOLECULAR DATING. Minima are based on oldest generally agreed crown members. Maxima are based on absence of putative members from very well sampled avifaunas that do preserve stem or ecological equivalents and cover predicted geographic regions of origin. The stochastic importance of sampling is particularly important for clades that may have a long fuse of low diversity (e.g., aves), less so for clades that immediately radiated (as shown from very short molecular branch lengths, e.g., LTT plots), such that at least moderate diversity existed early for higher fossil sampling probability. Stratigraphic conversions, where required, are based on the International Chronostratigraphic Chart 2015/1.

Node	Minimum age (Ma)	Maximum age (Ma)	Justification	Reference
Avian root	66.5	124.1	Minimum = <i>Vegavis</i> ; maximum covers absence of putative crown birds from well sampled avifaunas as far back as the Jianshangou beds.	Clarke JA, Tambussi CP, Noriega JI, Erickson GM, Ketcham RA. 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. <i>Nature</i> <b>433</b> .
Galloanserae	66.5	83.8	Minimum = <i>Vegavis</i> ; maximum covers absence of Galloanserae from global Campanian avifaunas.	Clarke et al. (2005)

Penguin/tube-nose	60.5	72.3	Minimum = <i>Waimanu</i> ; maximum covers absence of penguin/tube-noses from global Maastrichtian avifaunas.	Slack KE, Jones CM, Ando T, Harrison GL, Fordyce RE, Arnason U, Penny D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. <i>Molecular Biology and Evolution</i> <b>23</b> (11444-1155).
Core land/water birds	60.5	72.3	Minimum = <i>Waimanu</i> ; maximum covers absence of putative members of this clade (or deeper Neoaves) from global middle-early Maastrichtian avifaunas. <i>Polarornis</i> at approx. 67 Ma is the oldest putative member, although its original designation as an early loon has been dismissed by many authors.	Slack et al. (2006)
Non-ostrich palaeognaths	56.0	72.3	Minimum = <i>Diogenornis</i> (variously found to be a stem rhea or cassuariiform); maximum covers absence of non-ostrich (or any) palaeognathae from global Maastrichtian avifaunas.	Alvarenga HME. 1983. Uma ave ratita do paleoceno brasileiro: bacia caladria de Itaborai, estado do rio de Janeiro, Brasil. Boletim do Museu Nacional (Rio de Janeiro). <i>Geologica</i> <b>37</b> (1-194). Phillips et al. (2011)
Parrots/Passeriformes	53.5	NA	Minimum = <i>Pulchrapollia gracilis</i>	Ksepka DT and Clarke JA. 2015. Phylogenetically vetted and stratigraphically constrained fossil calibrations within Aves. <i>Palaeontologia Electronica</i> <b>18.1.3FC</b> .
Emu/kiwi	24.5	NA	Minimum - <i>Emuarius</i>	Worthy TH, Hand SJ, Archer M. 2014. Phylogenetic relationship of the Australian Oligo-Miocene ratite <i>Emuarius gidju</i> Casuariidae. <i>Integr. Zool.</i> <b>9</b> (148-166).

**MCMCTREE PARAMETERS.** The input parameters for molecular dating analysis using MCMCtree (Yang, 2006) were: seed = -1, ndata = no. partitions (depending on the dataset), seqtype = 0, usedata = 3, RootAge = >66.5<124.1, model = 7, alpha = 1, ncatG = 4, cleandata = 0, BDparas = 1 1 0, kappa\_gamma = 6 2, alpha\_gamma = 1 1, rgene+gamma = 1 2, sigma2\_gamma = 1 1, finetune = 1, print = 1, burnin = 10,000, sampfreq = 50, nsample = 1200. The output .BV file was renamed to in.BV and used as the input to rerun the analysis (i.e., usedata = 2). Fossil calibrations described in Table S5.9.18 were used to constrain node ages. For each dataset described in Table S5.9.17 (nuclear, mitochondrial, etc.), analyses with both independent (clock = 2) and correlated (clock = 3) rates were performed with two different input phylogenetic trees: one with the moa/tinamou clade as the second-deepest lineage (as suggested by Haddrath and Baker 2012, and Baker et al. 2014), and the other with the rhea clade as the deepest lineage among the notopalaeognathae (as suggested by our phylogenetic analyses, as well as Mitchell et al. 2014). Each analysis was repeated a total of two times. Results are shown in Figure S5.9.5 and Table S5.9.19.



**FIGURE S5.9.5 | CHRONOGRAMS FOR ALTERNATIVE DATASETS.** **A** mitochondrial (with third codon positions RY-coded), **B** nuclear (with the mitochondrial topology), and **C** nuclear (with the nuclear topology) molecular clock dating trees with palaeognath divergence times shown adjacent to each node and time (Ma) along the x-axis. Grey bars represent 95% highest posterior density intervals.

**MOLECULAR DATING RESULTS SUMMARY.**

**TABLE S5.9.19** | DIVERGENCE TIMES (MYA) OF EACH MAJOR PALAEOGNATH CLADE FOR MULTIPLE DATASETS, INPUT PHYLOGENETIC TREES, AND RATES. Node numbers refer to Figure 3. Blue highlighted rows are the three datasets believed to be the most accurate and, as such, are the focus of comparisons and discussion in the main text.

Dataset	Deepest among Noto-palaeognathae	Rates	Run	Node age (mya)	Node					
					A Palaeognathae to the exclusion of Neognathae	B Noto-palaeognathae to the exclusion of Ostrich	C Novaeratites to the exclusion of Moas/Tinamous	C Moa/Tinamou-Novaeratites to the exclusion of Rheas	D Novaeratites to the exclusion of other Noto-palaeognaths	E Elephant birds-Kiwi to the exclusion of Casuariidae
Nuc	Moa/tinamous	Independent	1	Average	68.6	62.4	58.2		54.8	51.5
				Lower 95% CI	60.4	55.7	45.4	41.8	38.1	
				Upper 95% CI	79.7	71.8	69.6	67.1	64.3	
Nuc	Moa/tinamous	Independent	2	Average	68.3	62.1	57.8		54.5	51.1
				Lower 95% CI	60.3	55.7	45.1	41.4	37.6	
				Upper 95% CI	79.5	71.8	69.6	66.8	64.1	
Nuc	Moa/tinamous	Correlated	1	Average	71.9	65.2	62.2		59.1	55.6
				Lower 95% CI	61.9	56.2	52	48.6	45	
				Upper 95% CI	81.9	72.6	71	68.1	64.8	
Nuc	Moa/tinamous	Correlated	2	Average	71.7	65.1	62.1		58.9	55.4
				Lower 95% CI	61.9	56.2	52.2	48.6	45	
				Upper 95% CI	81.9	72.6		71	68	64.7

Nuc	Rheas	Independent	1	Average	69.4	61.6		63.1	58.4	55.1
				Lower 95% CI	60.8	53.6		55.8	47.4	43.1
				Upper 95% CI	80.2	70.7		72	68.2	65.4
Nuc	Rheas	Independent	2	Average	69.1	61.3		62.9	58.1	54.8
				Lower 95% CI	60.6	53.5		55.8	47.6	43.4
				Upper 95% CI	80	70.5		71.9	68	65.3
Nuc (Figure S5b) (mt topology)	Rheas	Correlated	1	Average	71.6	63		64.6	60.3	56.8
				Lower 95% CI	62.1	54.3		56.1	51.3	47.4
				Upper 95% CI	81.7	71.3		72.5	68.7	65.4
Nuc (Figure S5b) (mt topology)	Rheas	Correlated	2	Average	72.1	63.5		65.1	60.9	57.3
				Lower 95% CI	62.2	54.4		56.3	51.5	47.7
				Upper 95% CI	82	71.4		72.5	69	65.7
Nuc (Figure S5b) (mt topology)	Rheas	Correlated	3	Average	71.8	63.3		64.8	60.6	57
				Lower 95% CI	62	54.3		56.1	51.3	47.4
				Upper 95% CI	82.1	71.4		72.5	68.9	65.5
Mt	Moa/ tinamous	Independent	1	Average	68.8	64.7	58.6		49.3	44.8
				Lower 95% CI	60	56.8	45.1		37	32.8
				Upper 95% CI	77.7	71.9	68.6		61.1	56.8
Mt	Moa/ tinamous	Independent	2	Average	67.3	63.6	58.3		52.8	48.3
				Lower 95% CI	59.7	56.6	50.4		44.4	40.1
				Upper 95% CI	72.9	69.1	64.6		59.2	54.9
Mt	Moa/ tinamous	Correlated	1	Average	68.7	64.7	61.2		56.5	52.1
				Lower 95% CI	60.4	57.1	54.1		49.9	45.7
				Upper 95% CI	76.2	71.7	68.3		63.4	58.7

Mt	Moa/ tinamous	Correlated	2	Average	69.1	65.1	61.6	56.8	52.3
				Lower 95% CI	60.9	57.6	54.6	50.5	46.3
				Upper 95% CI	76.4	72	68.6	63.7	59
Mt	Rheas	Independent	1	Average	66.9	61.1	63.7	54.1	49.8
				Lower 95% CI	59.5	54.4	56.8	46.4	42
				Upper 95% CI	72.9	67.2	69.8	60.7	56.5
Mt	Rheas	Independent	2	Average	67.8	61.3	64.8	51.4	47.1
				Lower 95% CI	59.4	53.1	56.8	37.4	33.1
				Upper 95% CI	75.4	68.5	71.6	62.5	58.7
Mt	Rheas	Correlated	1	Average	68.4	62	64.7	57.3	53
				Lower 95% CI	60	54.5	56.9	50.2	46.2
				Upper 95% CI	76	69.2	72	64.2	59.6
Mt	Rheas	Correlated	2	Average	68.2	61.8	64.5	57.1	52.7
				Lower 95% CI	59.9	54.4	56.8	50.1	46.1
				Upper 95% CI	75.6	68.9	71.7	63.9	59.3
Mt (M3 RY- coded)	Moa/ tinamous	Independent	1	Average	67.2	63.4	57.2	51.5	46.7
				Lower 95% CI	59.5	56.4	47.9	41.9	37.3
				Upper 95% CI	73.5	69.7	64.8	59	54.4
Mt (M3 RY- coded)	Moa/ tinamous	Independent	2	Average	66.9	63.3	51.9	45.1	41.7
				Lower 95% CI	59	56.2	39.4	34.3	31.2
				Upper 95% CI	74.5	70.5	63.7	56.5	53
Mt (M3 RY- coded)	Moa/ tinamous	Correlated	1	Average	69.2	65.1	61.4	57	52.2
				Lower 95% CI	61.2	57.7	54.6	50.6	46.1
				Upper 95% CI	76.8	72	68.8	64.5	59.5

Mt (M3 RY-coded)	Moa/tinamous	Correlated	2	Average	69.2	65.1	61.4	56.9	52.1
				Lower 95% CI	60.9	57.4	54.5	50.4	46
				Upper 95% CI	77	72.1	68.9	64.4	59.4
Mt (M3 RY-coded)	Rheas	Independent	1	Average	66.9	61.1	63.5	53.7	49
				Lower 95% CI	59.4	54.2	56.6	45.2	40.5
				Upper 95% CI	73.3	67.6	69.8	61	56.6
Mt (M3 RY-coded)	Rheas	Independent	2	Average	66.8	61	63.4	53.7	49
				Lower 95% CI	59.3	54.1	56.4	45	40.2
				Upper 95% CI	73.7	67.8	70.1	61.5	56.9
Mt (M3 RY-coded) (Figure S5a)	Rheas	Correlated	1	Average	69	62.6	65	58.1	53.5
				Lower 95% CI	60.7	55.1	57.3	51.1	46.7
				Upper 95% CI	77.2	70	72.3	65.7	61
Mt (M3 RY-coded) (Figure S5a)	Rheas	Correlated	2	Average	68.9	62.5	64.9	58.1	53.4
				Lower 95% CI	60.4	55	57.1	51	46.5
				Upper 95% CI	77.2	70	72.3	65.6	61
Mt (RY, M3 excluded)	Moa/tinamous	Independent	1	Average	71.3	65.4	63.9	33.1	27.7
				Lower 95% CI	62.1	56.7	52.1	22.8	16.5
				Upper 95% CI	83.1	72.3	71.5	58.4	52.9
Mt (RY, M3 excluded)	Moa/tinamous	Independent	2	Average	71.4	65.4	62.5	34.3	28.8
				Lower 95% CI	62.3	56.7	39.7	23.5	17.4
				Upper 95% CI	82.8	72.3	71.5	59.2	52.6
Mt (RY, M3 excluded)	Moa/tinamous	Correlated	1	Average	70.4	64.2	62.2	57.5	52.6
				Lower 95% CI	61.9	56.5	53.4	47.4	41.4

				Upper 95% CI	81.4	72.3	71.2	67.7	63.8
Mt (RY, M3 excluded)	Moa/ tinamous	Correlated	2	Average	70.6	64.1	62.9	59.4	54.2
				Lower 95% CI	62.0	56.5	55.0	49.6	42.6
				Upper 95% CI	81.4	72.2	71.4	69.3	65.3
Mt (RY, M3 excluded)	Rheas	Independent	1	Average	70.5	62.9	64.9	36.6	30.9
				Lower 95% CI	61.7	53.7	56.6	23.7	18.0
				Upper 95% CI	81.6	70.6	72.1	60.6	54.7
Mt (RY, M3 excluded)	Rheas	Independent	2	Average	70.7	63.3	65.2	35.6	29.9
				Lower 95% CI	62	54.1	56.8	23.6	17.8
				Upper 95% CI	81.4	71	72.3	59.4	53
Mt (RY, M3 excluded)	Rheas	Correlated	1	Average	70.3	61.7	63.9	57.5	51.7
				Lower 95% CI	61.9	53.8	56.4	48.3	41.4
				Upper 95% CI	81.7	70.4	72.2	67.3	62.8
Mt (RY, M3 excluded)	Rheas	Correlated	2	Average	70.2	61.3	63.5	57.6	51.9
				Lower 95% CI	61.8	53.6	56.2	48.8	41.7
				Upper 95% CI	81.4	70.3	72.2	67.6	63.2
Nuc + Mt	Moa/ tinamous	Independent	1	Average	65.9	62	55.9	50.9	47.1
				Lower 95% CI	59.4	55.9	47.7	33.5	30.8
				Upper 95% CI	82	82	62.8	58	54.2
Nuc + Mt	Moa/ tinamous	Independent	2	Average	65.7	61.5	56.9	52.1	48.1
				Lower 95% CI	59.6	56	50.8	45.7	41.2
				Upper 95% CI	72.3	67.9	64.4	59	55.3
Nuc + Mt	Moa/ tinamous	Correlated	1	Average	68.2	63.6	60.5	56.4	52.4
				Lower 95% CI	60.5	56.5	53.6	49.9	46

				Upper 95% CI	77.1	72	68.7		64.5	60.3
Nuc + Mt	Moa/ tinamous	Correlated	2	Average	68.1	63.5	60.4		56.2	52.2
				Lower 95% CI	60.7	56.6	53.8		50	46.1
				Upper 95% CI	76.8	71.7	68.4		64	59.6
Nuc + Mt	Rheas	Independent	2	Average	65	58.9		61.1	53.5	49.7
				Lower 95% CI	59.2	53.8		56	47.6	43.7
				Upper 95% CI	71.9	65.2		67.6	60.1	56.5
Nuc + Mt	Rheas	Correlated	1	Average	68.4	61.3		63.7	57.2	53.2
				Lower 95% CI	60.6	54.4		56.6	50.7	46.9
				Upper 95% CI	77.1	69.2		71.8	64.9	60.7
Nuc + Mt	Rheas	Correlated	2	Average	68.3	61.3		63.6	57.2	53.2
				Lower 95% CI	60.7	54.5		56.6	50.7	47.1
				Upper 95% CI	76.9	69		71.6	64.7	60.4
Nuc + Mt (M3 RY-coded)	Moa/ tinamous	Independent	1	Average	64.7	60.4	55.4		50.5	46.6
				Lower 95% CI	59.1	55.8	48.1		42	38.3
				Upper 95% CI	71.9	67.6	63.2		58.1	54.2
Nuc + Mt (M3 RY-coded)	Moa/ tinamous	Independent	2	Average	69.6	64.7	60.9		56.6	53.1
				Lower 95% CI	60.7	56.3	50.9		46	42.4
				Upper 95% CI	78.2	72.2	69.7		66	63.1
Nuc + Mt (M3 RY-coded)	Moa/ tinamous	Correlated	2	Average	69.3	64.5	61.3		57.5	53.2
				Lower 95% CI	61	56.8	53.9		50.4	46.4
				Upper 95% CI	77.9	72.2	69.3		65.3	61.1
Nuc + Mt (M3 RY-coded)	Moa/ tinamous	Correlated	1	Average	69.4	64.6	61.4		57.5	53.3
				Lower 95% CI	61.2	56.9	54		50.5	46.5

				Upper 95% CI	78	72.3	69.4	65.4	61.2	
Nuc + Mt (M3 RY-coded)	Rheas	Independent	2	Average	64.7	58.6		60.5	53.3	49.4
				Lower 95% CI	59.1	53.6		55.8	47	42.9
				Upper 95% CI	71.9	65.3		67.4	60.4	56.5
Nuc + Mt (M3 RY-coded)	Rheas	Independent	1	Average	64.4	58.3		60.3	53.2	49.3
				Lower 95% CI	59	53.6		55.8	47.2	43.1
				Upper 95% CI	71.9	65.3		67.4	60.4	56.8
Nuc + Mt (M3 RY-coded) (Figure 3)	Rheas	Correlated	1	Average	69.2	62.1		64.2	58.3	54.2
				Lower 95% CI	61.2	54.9		56.7	51.4	47.4
				Upper 95% CI	78	70		72.2	66	61.8
Nuc + Mt (M3 RY-coded) (Figure 3)	Rheas	Correlated	2	Average	69.2	62.1		64.2	58.3	54.2
				Lower 95% CI	61.4	55.1		56.9	51.5	47.5
				Upper 95% CI	77.9	69.9		72.1	65.9	61.5
Nuc+Mt (RY, M3 excluded)	Moa/ tinamous	Independent	1	Average	67.3	61.3	59.0		56	52.6
				Lower 95% CI	60.4	55.8	50.1		45.2	41.5
				Upper 95% CI	76.8	70.5	68.9		66.5	63.8
Nuc+Mt (RY, M3 excluded)	Moa/ tinamous	Independent	2	Average	66.5	60.3	57.3		53.2	49.4
				Lower 95% CI	60.3	55.7	48.5		40.9	37
				Upper 95% CI	75.7	69.3	67		63.5	60.0
Nuc+Mt (RY, M3 excluded)	Moa/ tinamous	Correlated	1	Average	71.6	64.8	63.1		60.2	56.3
				Lower 95% CI	62.9	56.7	55		52.0	47.9
				Upper 95% CI	81.2	72.4	71.5		68.6	64.8

Nuc+Mt (RY, M3 excluded)	Moa/ tinamous	Correlated	2	Average	67.7	65.4	65	64.2	63.2
				Lower 95% CI	61.7	59.7	59.2	58.1	56.5
				Upper 95% CI	74.3	72.3	72.1	71.6	71
Nuc+Mt (RY, M3 excluded)	Rheas	Independent	1	Average	67.4	60	61.4	57.4	53.8
				Lower 95% CI	60.4	54	55.8	49.4	45
				Upper 95% CI	76.9	69.1	70.5	67	63.9
Nuc+Mt (RY, M3 excluded)	Rheas	Independent	2	Average	67.3	53.3	56.5	50.7	45.4
				Lower 95% CI	57.6	43.4	54.3	34.4	19.4
				Upper 95% CI	77.6	57.3	59.6	56.1	54.1
Nuc+Mt (RY, M3 excluded)	Rheas	Correlated	1	Average	70.4	62	63.6	59.3	55
				Lower 95% CI	62.3	54.7	56.3	51.9	47.6
				Upper 95% CI	80.6	72.2	72.2	68	63.9
Nuc+Mt (RY, M3 excluded)	Rheas	Correlated	2	Average	70.4	62	64	60	56.1
				Lower 95% CI	61.8	54.8	56.3	52.1	47.9
				Upper 95% CI	80.4	71.2	72.3	69.2	66.1
					<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
	<b>Deepest among Noto-palaeognathae</b>	<b>Rates</b>	<b>Run</b>	<b>Node age (mya)</b>	<b>Palaeognathae to the exclusion of Neognathae</b>	<b>Noto-palaeognathae to the exclusion of Ostrich</b>	<b>Moa/Tinamou-Novaeratites to the exclusion of Rheas</b>	<b>Moa/Tinamou-Emu to the exclusion of Elephant birds-Kiwi</b>	<b>Elephant birds-Kiwi to the exclusion of Moa/Tinamou-Emu</b>
<b>Dataset</b>									
Nuc	Rheas	Independent	1	Average	68.9	62.6	60.5	58.0	57.2

(Nuc topology)				Lower 95% CI	60.9	55.8	52.8	49.9	47.6
				Upper 95% CI	79.1	71.3	69.2	67.0	66.2
Nuc	Rheas	Independent	2	Average	69.2	62.8	60.8	58.3	57.5
(Nuc topology)				Lower 95% CI	60.8	55.9	52.8	49.8	47.4
				Upper 95% CI	79.7	71.5	69.6	67.4	66.4
Nuc	Rheas	Correlated	1	Average	71.9	64.6	62.3	59.7	58.9
(Figure S5c)				Lower 95% CI	62.4	56.1	53.7	50.7	49.6
(Nuc topology)				Upper 95% CI	81.8	72.1	70.0	67.7	66.4
Nuc	Rheas	Correlated	2	Average	72.4	65.1	62.9	60.2	59.4
(Figure S5c)				Lower 95% CI	62.4	56.2	53.9	51.1	49.9
(Nuc topology)				Upper 95% CI	82.3	72.4	70.4	67.8	66.7

### S5.9.13 SUPPLEMENTARY REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Baker AJ, Haddrath O, McPherson JD, Cloutier A. 2014. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Molecular Biology and Evolution* **31**: 1686-1696.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**.
- Blankenberg DVK, Coraro N, Ananda G, Lazarus R, Mangan M, et al. 2010. Galaxy: a web-based genome analysis tool for experimentalists. In: al. FMAe, editor. *Current Protocols in Molecular Biology*. Unit 19.10.11-21.
- Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillette LJ, et al. 2007. Patterns of vertebrate isochore evolution revealed by comparison of expressed mammalian, avian, and crocodylian genes. *Journal of Molecular Evolution* **65**: 259-266.
- Cooper A. 1994. Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthisittidae). *Experientia* **50**: 558-563.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, et al. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 15758-15763.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

- Enk JM, et al. 2014. Ancient whole genome enrichment using baits built from modern DNA. *Molecular Biology and Evolution* **31**: 1292-1294.
- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Giardine BR, Hardison RC, Burhans R, Elnitski L, Shah P, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451-1455.
- Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**: 2153-2155.
- Goecks J, Nekrutenko A, Taylor J and The Galaxy Team. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biology* **11**.
- Guindon S and Gascuel, O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systems Biology* **52**: 696-704.
- Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**: 1763-1768.
- Haddrath O and Baker AJ. 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proceedings of the Royal Society B-Biological Sciences* **279**: 4617-4625.
- Hahn C, Bachmann L, Chevreaux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* **41**.

- Harshman JB, Edward L, Braun MJ, Huddleston CJ, Bowie RCK, et al. 2008. *Phylogenomic evidence for multiple losses of flight in ratite birds. Proceedings of the National Academy of Sciences* **105**: 12462-12467.
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Huelsenbeck JP and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **345**: 1320-1331.
- Jonsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682-1684.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage - Building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Le Duc D, Renaud G, Krishnan A, Almén MS, Huynen L, et al. 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology* **16**.
- Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJP. 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques* **54**: 321-326.

- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*: 1-8.
- Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, et al. 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **277**: 1991-2000.
- Phillips MJ, Gibb GC, Crimp EA, Penny D. 2010. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. *Systems Biology* **59**: 90-107.
- Phillips MJ and Pratt RC. 2008. Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums). *Molecular Phylogenetics and Evolution* **46**: 594-605.
- Rambaut A. 1996. Sequence Alignment (Se-AL) Program. Version 2.0a. Oxford: University of Oxford.
- Rambaut A and Drummond AJ. 2007. TreeStat v1.2: Tree statistic calculation tool. Version 1.2.
- Rambaut A, Suchard MA, Xie W, Drummond AJ. 2003. Tracer: MCMC Trace Analysis Tool. Version 1.6.1pre.
- Schroeder H, Ávila-Arcos MC, Malaspinas AS, Poznik GD, Sandoval-Velasco M, et al. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 3669-3673.
- Shapiro B and Hofreiter M. 2012. *Ancient DNA: methods and protocols*. New York Humana Press (Springer Science).

- Shimodaira H and Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**: 1246-1247.
- Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**: 492-508.
- Smith JV, Braun EL, Kimball RT. 2013. Ratite Nonmonophyly: Independent Evidence from 40 Novel Loci. *Systems Biology* **62**: 35-49.
- Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **30**: 1312-1313.
- Swofford DL. 2003. PAUP\*. Version 4.0b. Sunderland, Massachusetts: Sinauer Associates.
- Vallender EJ. 2009. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. *Methods* **49**: 50-55.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Yang Z and Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212-226.

## 5.10 EPILOGUE

---

This chapter looked at using aDNA extracted from fossil eggshell to shed new light on the evolutionary history of elephant birds, including their relationship to other palaeognaths. Unlike previous studies, using elephant bird eggshell instead of bone allowed us to obtain both a full mitochondrial genome and nuclear loci that resulted in more precise estimates for the timing of the palaeognath radiation, as well as independent verification of New Zealand's kiwi as the closest living relative of Madagascar's elephant birds. The results of this study highlight how aDNA information can, in the words of TH Huxley, slay “a beautiful hypothesis”—that being the vicariance model of speciation described by Richard Dawkins in this chapter's epigraph, which, while an elegant theory that was the accepted view of ratite evolution for many years, turns out to have little foundation in reality.

As shown in Chapters 2, 3, and 4, utilising non-traditional substrates can be key when attempting to extract aDNA from warm, tropical climates. Here, we show that obtaining genomic information from extinct birds is made more successful by extracting DNA from avian eggshell as opposed to bone, as eggshell contains a higher proportion of endogenous DNA (i.e., it is notable that a whole mitogenome was recovered here without the need for enrichment through hybridisation capture). However, hybridisation capture was found to be a highly successful method of enriching for endogenous aDNA targets from eggshell. Thus, eggshell is the most promising substrate for future investigations into the genetic mechanisms underlying elephant bird speciation, adaptation, and extinction.

The next chapter (Chapter 6) moves in this direction, using eggshell aDNA from multiple samples across Madagascar to explore the genetic diversity within elephant birds for the first time. Doing so required further optimisation of the eggshell DNA extraction method and improvements to the hybridisation capture methods employed here, both of which are expanded upon in the following chapter.

— CHAPTER 6 —

---

**ANCIENT DNA FROM FOSSIL EGGSHELL MORPHOTYPES REVISES  
PHYLOGEOGRAPHY *and* SYSTEMATICS of EXTINCT MADAGASCAN  
ELEPHANT BIRDS (AVES: AEPYORNITHIDAE)**

---

*When they found an Aepyornis with a thigh a yard long, they thought they had reached the top of the scale, and called him Aepyornis maximus. Then someone turned up another thigh-bone four feet six or more, and that they called Aepyornis titan...if they get any more Aepyornises, he reckons some scientific swell will go and burst a blood-vessel.*

- HG Wells  
*Aepyornis* Island

## 6.1 PROLOGUE

---

Chapter 5 exemplified how aDNA extracted from fossil eggshell can be used to retrieve genomic information from extinct birds, particularly those who leave few skeletal remains with well-preserved aDNA, such as the elephant birds of Madagascar. Because the tropical climate of Madagascar is not conducive to long-term aDNA preservation, coupled with the fact that skeletal specimens are rare, eggshell aDNA may be the only opportunity for us to further our knowledge of the phylogeography and systematics of elephant birds.

In this chapter, we use aDNA extracted from many elephant bird eggshell specimens from around Madagascar in order to explore just that. To do this involved further optimisation of aDNA extraction from eggshell in order to improve the success rate of the method. We tested several different factors (both on their own and in combination with one another) that may have an effect on the recovery of aDNA from eggshell, including: (1) the presence or absence of a 95°C heat step after digestion; (2) choice of digest buffer (namely, Oskam et al. 2010 versus Dabney et al. 2014); (3) the presence or absence of a 30 min predigestion step prior to an overnight digestion; (4) the use of hydrochloric acid to decalcify the eggshell prior to overnight digestion; and (5) the volumetric ratio of chaotropic salt solution to digest buffer for binding DNA to silica (i.e., 5:1 versus 13:1). To evaluate the recovery of aDNA from each of these methods, we compared relative aDNA quantity obtained via qPCR, as well as the relative proportion of endogenous aDNA to contaminating bacterial DNA sequenced. In addition, there was an attempt to quantify the absolute amounts of endogenous DNA retrieved from each method using a digital PCR (QuantStudio 3D); however, these attempts, though promising, were unsuccessful and in the interest of time were not pursued further. Coupled with the design and use of elephant bird-specific hybridisation capture baits (based on the genome retrieved in Chapter 5) to better enrich for mitochondrial DNA, these experiments improved our ability to recover aDNA from eggshell, allowing us to retrieve an additional 16 near-complete mitochondrial genomes from across Madagascar.

Using this information, we reconstruct the evolutionary relationships between elephant bird taxa, and explore how these relationships are correlated with their

geographic distribution as well as eggshell morphology (both macro- and microstructure). We also reevaluate the current taxonomic classification of elephant birds in light of these data.

The study presented in this chapter is a manuscript in preparation for submission in a high-impact journal (*Proceedings of the National Academy of Sciences*), given that this is the first study to use genetic information as well as eggshell morphology (using micro X-ray computed tomography) to clarify the taxonomy of elephant birds, describe their phylogeography within Madagascar, and the explore factors that may have been responsible for their diversification. This is no easy feat considering that aDNA preservation in this environment is highly limited. We envision that this manuscript will be ready for submission in early 2017. Note that the author list may change with the addition of new data.

### **6.1.1 ACKNOWLEDGEMENTS**

This work was funded by a grant from the Australian Research Council (ARC) awarded to JH (DE120100107). MB was supported in this research by an ARC future fellowship (FT0991741). GF wishes to acknowledge the National Geography Society, as well as the Easterbrook Distinguished Scientist Award (from the Division of Quaternary Geology and Geomorphology Division of the Geological Society of America), for funding sample collecting trips in Madagascar. KD's collections were made possible with funding from the National Science Foundation Graduate Research Fellowship Program, the P.E.O. Scholar Award, the Yale Institute of Biospheric Studies, the Yale MacMillan Center for International and Area Studies and the Yale Council on Archaeological Studies. Research permissions were granted by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Autorisation Numéro 128/13-MESupReS/SG/DGRP and by the Centre de Documentation et de Recherche sur l'Art et les Traditions Orales Malgaches (CEDRATOM), under the auspices of the Memorandum of Understanding between the University of Toliara, under the direction of Dr. Barthélémy Manjakahery, Director of the CEDRATOM, and Yale University, under the direction of Dr. Roderick McIntosh, Professor of Anthropology. Local permission to carry out archaeological research was granted by the Office du Maire, Commune de Befandefa

and by the Chefs de Fokontany of Andavadoaka, Nosy Ve, Antsaragnangny, Lamboara, Ampasilava and Salary. Permits for the export of archaeological materials for the purposes of laboratory analysis were granted by the Secretariat Général of the Ministère de l'Artisanat de la Culture et des Patrimoines, Direction Régionale de la Culture et du Patrimoine Atsimo Andrefana, Visas de Sorties Numéro 09/06-MCP/SG/DRCP.AA; Numéro 05/14-MACP/SG/DRCP.AA; Numéro 08/14-MACP/SG/DRCP.AA in accordance with Avis Numéro 375, 02/02/1978. Computing resources at the Pawsey Supercomputing Centre (WA) were used to perform BLAST searches. We would like to thank Alison Devault at *MYcroarry* for assistance designing enrichment baits. We give special thanks to the Morombe Archaeological Project (MAP) team and to the people of Andavadoake, Madagascar for helping collect eggshell. The authors acknowledge the facilities, and the scientific and technical assistance of the National Imaging Facility at the Centre for Microscopy, Characterisation and Analysis, The University of Western Australia—a facility funded by the University, State and Commonwealth Governments. The authors would like to thank James Hansford for his expertise on elephant bird taxonomy and knowledge of the morphology of skeletal specimens.

### **6.1.2 AUTHOR CONTRIBUTIONS**

MB, GM, and JH conceived and supervised the study. AG and MB designed the experiments. AG performed the genetic experiments. AG analysed the data with assistance from MP. GM and KD collected and supplied eggshell for genetic analysis. GM performed the dating, amino acid racemization and isotope analysis. KD performed dating. DE performed the micro-CT and 2D/3D analyses. PR and AB performed confocal microscopy. AG wrote the manuscript with contributions from all authors.

### **6.1.3 AUTHOR DECLARATIONS**

The authors declare no competing interests.

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis, or to use any of my own published work (e.g., journal articles) in which the copyright is held

by another party (e.g. publisher, co-author). Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

All supplementary data related to this article can be found at in the file “Supplementary Information” published with the online version of the article, as well as section 6.8.

Mitochondrial and nuclear genome sequences for the studied specimens will be deposited within GenBank (available [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov)) upon publication. Data will be deposited on DataDryad upon publication. Correspondence and requests for materials should be address to AG ([alicia.grealy@uqconnect.edu.au](mailto:alicia.grealy@uqconnect.edu.au)).

ANCIENT DNA FROM FOSSIL EGGSHELL MORPHOTYPES REVISES  
PHYLOGEOGRAPHY AND SYSTEMATICS OF EXTINCT MADAGASCAN ELEPHANT  
BIRDS (AVES: AEPYORNITHIDAE)

—*manuscript in preparation*—

Alicia Grealy<sup>1\*</sup>, Gifford Miller<sup>2</sup>, Matthew Phillips<sup>3</sup>, Diana Engineer<sup>4</sup>, Paul Rigby<sup>4</sup>,  
Alysia Buckley<sup>4</sup>, Kristina Douglass<sup>5</sup>, James Haile<sup>6</sup>, Michael Bunce<sup>1</sup>

<sup>1</sup> Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment  
and Agriculture, Curtin University, Bentley, WA 6102, Australia

<sup>2</sup> Institute of Arctic and Alpine Research, Geological Sciences, University of  
Colorado, Boulder, Colorado 80303, USA

<sup>3</sup> Vertebrate Evolution Group, School of Earth, Environmental and Biological  
Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia

<sup>4</sup> Centre for Microscopy, Characterisation and Analysis, Harry Perkins Institute of  
Medical Research, The University of Western Australia, Crawley, WA 6009,  
Australia

<sup>5</sup> Department of Anthropology, National Museum of Natural History, Smithsonian  
Institution, USA

<sup>6</sup> Research Laboratory for Archaeology and the History of Art, Oxford University,  
Oxford OX12JD, UK.

\* Corresponding author: Trace and Environmental DNA Laboratory, Department of  
Environment and Agriculture, Curtin University, Perth, WA 6102, Australia. E-mail  
address: [alicia.grealy@uqconnect.edu.au](mailto:alicia.grealy@uqconnect.edu.au)

**Key words:** ancient DNA, eggshell, high-throughput sequencing, Madagascar,  
morphology, Palaeognath, phylogeography

## 6.2 ABSTRACT

---

The elephant birds of Madagascar (Aves: Aepyornithidae) were large, flightless ratites that became extinct almost a millennium ago. Although skeletal fossils are rare, at least seven species of elephant bird have been described. However, these identifications remain controversial due to a lack of information regarding species distributions, the limits of intra- versus inter-specific morphological variation, and the degree of sexual dimorphism exhibited by these birds. As such, little is known about the evolutionary and geographic relationships among elephant bird species. Here, we use fossil eggshell ( $n=20$ ) morphological and mitochondrial DNA information from North, Southwest, and South Madagascar to revisit elephant bird taxonomy and provide the first analysis of elephant bird phylogeography. High-throughput sequencing of 16 near-complete and four partial mitochondrial genomes from eggshell reveals genetic variation that is unambiguously correlated to eggshell thickness, macromolecular structure, and geographic distribution. Despite major skeletal differences within genera that support the classification of multiple species, low levels of within-genera genetic variation in barcoding genes suggest that, during the Holocene, no more than one species of *Aepyornis* and one species of *Mullerornis* existed in South-Southwest Madagascar; however, we describe a genetically and morphologically divergent taxon from Madagascar's North that may represent a novel oospecies of *Aepyornis*. Furthermore, phylogenetic and molecular dating indicates divergence in elephant bird genera occurred ca. 30 Ma, potentially prompted by changes in Madagascar's palaeoclimate and vegetation that occurred during the Eocene-Oligocene. Divergence within *Aepyornis* is estimated to have occurred 7-8 Ma around the onset of the monsoon season in Madagascar that could have ecologically isolated North and South populations as rainfall increased in the North. These results offer the first insight into elephant bird biodiversity and speciation, and advocate for a major revision of elephant bird systematics that integrates these new molecular and morphological data.

### 6.3 INTRODUCTION

---

THE extinct elephant birds of Madagascar (Aepyornithidae) were large, flightless ratites, whose evolutionary relationships to other birds were, until recently, relatively unknown. While several studies over the past few years have shown elephant birds to be sister to New Zealand's kiwi (Mitchell et al. 2014a; Yonezawa et al. 2016; Grealy et al. 2017), the biodiversity and taxonomy within elephant birds remains controversial as the relationships between species have received little attention since they were formally described for the first time over 150 years ago (Hume and Walters 2012). About seven or eight species of elephant birds across two genera are generally accepted based on morphological comparison of skeletal fossils (Table 6.1). However, the taxonomic classifications of these skeletal specimens are uncertain as “many of the forms lack proper diagnoses, the limits of species are ambiguous, aspects of potential sexual dimorphism are unknown, [and] synonyms are complicated” (Goodman and Patterson 1997). For instance, most elephant bird species have been described from incomplete Pleistocene-Holocene post-cranial material from South and Central Madagascar only (Balanoff and Row 2007). The use of ancient DNA (aDNA), on the other hand, has proven to be highly successful in assisting the delineation of extinct bird species and their geographic boundaries (e.g., Bunce et al. 2003; Huynen et al. 2003; Lambert 2005; Huynen and Lambert 2014; Mitchell et al. 2014b). For New Zealand's extinct moa, aDNA analysis of *Dinornis* bone specimens from a rich Quaternary fossil record revealed two species (one sexually dimorphic) as opposed to the three species described from skeletal morphology (Bunce et al. 2003; Huynen et al. 2003). Not only does this suggest that perhaps fewer species of elephant bird actually existed than have been described, but also that the morphology of some skeletal elements can be poor for resolving taxonomy in extinct ratites, particularly when the fossil record is sparse. As such, the systematics of elephant birds is “in need of major revision” (Goodman and Patterson 1997), and consequently, the phylogenetic and geographic relationships among elephant bird taxa remain unknown. Information from aDNA has yet to contribute to clarifying the taxonomy within elephant birds and their phylogeographic relationships because, unlike moa, bone specimens are rare, and the warm, humid environment of Madagascar is suboptimal for the preservation of aDNA, resulting in highly degraded aDNA even in relatively young fossil bone (Grealy et al. 2016).

**TABLE 6.1** | CURRENT ACCEPTED SPECIES OF ELEPHANT BIRD. Superscripts on references are cross-referenced with Figure 6.1.

Family	Genus	Species	Synonyms	Reference	
Aepyornithidae	<i>Aepyornis</i>	<i>A. gracilis</i>		Described by <sup>1</sup> Monnier in 1913 ( <sup>3</sup> Brodkorb 1963, <sup>4</sup> Hume and Walters 2012)	
		<i>A. hildebrandti</i>	<i>A. mulleri</i>	Described by Burckhardt in 1893 and <sup>2</sup> Milne-Edwards in 1894 ( <sup>3</sup> , <sup>4</sup> )	
		<i>A. maximus</i>	<i>A. modestus</i> <i>A. ignens</i> <i>A. titan</i>	Described by <sup>2</sup> Milne-Edwards in 1894, <sup>5</sup> Andrews in 1894, <sup>6</sup> Lambrecht in 1933, <sup>7</sup> Milne-Edwards in 1869, <sup>8</sup> Geoffroy-Saint-Hilaire in 1850, 1851 ( <sup>3</sup> , <sup>4</sup> )	
		<i>A. medius</i>	<i>A. grandidieri</i> <i>A. cursor</i> <i>A. lentus</i>	Described by <sup>2</sup> Milne-Edwards and Grandidier in 1894, 1866, <sup>9</sup> Rowley in 1867 ( <sup>3</sup> )	
		<i>Mullerornis</i>	<i>M. rudis</i>	<i>Flacourtia rudis</i>	<sup>2</sup> Milne-Edwards and Grandidier in 1894 ( <sup>3</sup> , <sup>4</sup> )
			<i>M. agilis</i>		<sup>2</sup> Milne-Edwards and Grandidier in 1894 ( <sup>3</sup> , <sup>4</sup> )
			<i>M. betsilei</i>		<sup>2</sup> Milne-Edwards and Grandidier in 1894 ( <sup>3</sup> , <sup>12</sup> Turvey 2009)
				<i>M. grandis</i>	Described by <sup>10</sup> Lamberton in 1934 ( <sup>11</sup> MacPhee et al. 1985)

Eggshell has been shown to be an excellent alternative substrate for the retrieval of aDNA (including both whole mitochondrial genomes and nuclear loci) from elephant birds (Oskam et al. 2010; Grealy et al. 2017) that does not require destruction of rare skeletal specimens. Eggshell can also be found in abundance (Dewar 1984) in the places that elephant birds nested, and therefore is a better option for a comprehensive investigation of their taxonomy and phylogeography than bone. Here, 20 new full and partial mitochondrial genomes reconstructed using aDNA extracted from elephant bird eggshell specimens collected from South, Southwest, and North Madagascar, combined with hybridisation enrichment and next-generation sequencing approaches, were used to verify and revise taxonomy through molecular means, as well as generate the first phylogeographic record for elephant birds. Furthermore, we investigated the ability of elephant bird eggshell morphological analysis, as well as eggshell carbon and oxygen isotope content, to demarcate genera or mitochondrial haplotypes. Together, these data provide insights into the processes underlying elephant bird speciation and contributes to our understanding of how these birds fit into the rich evolutionary history of Madagascar.

## **6.4 MATERIALS *and* METHODS**

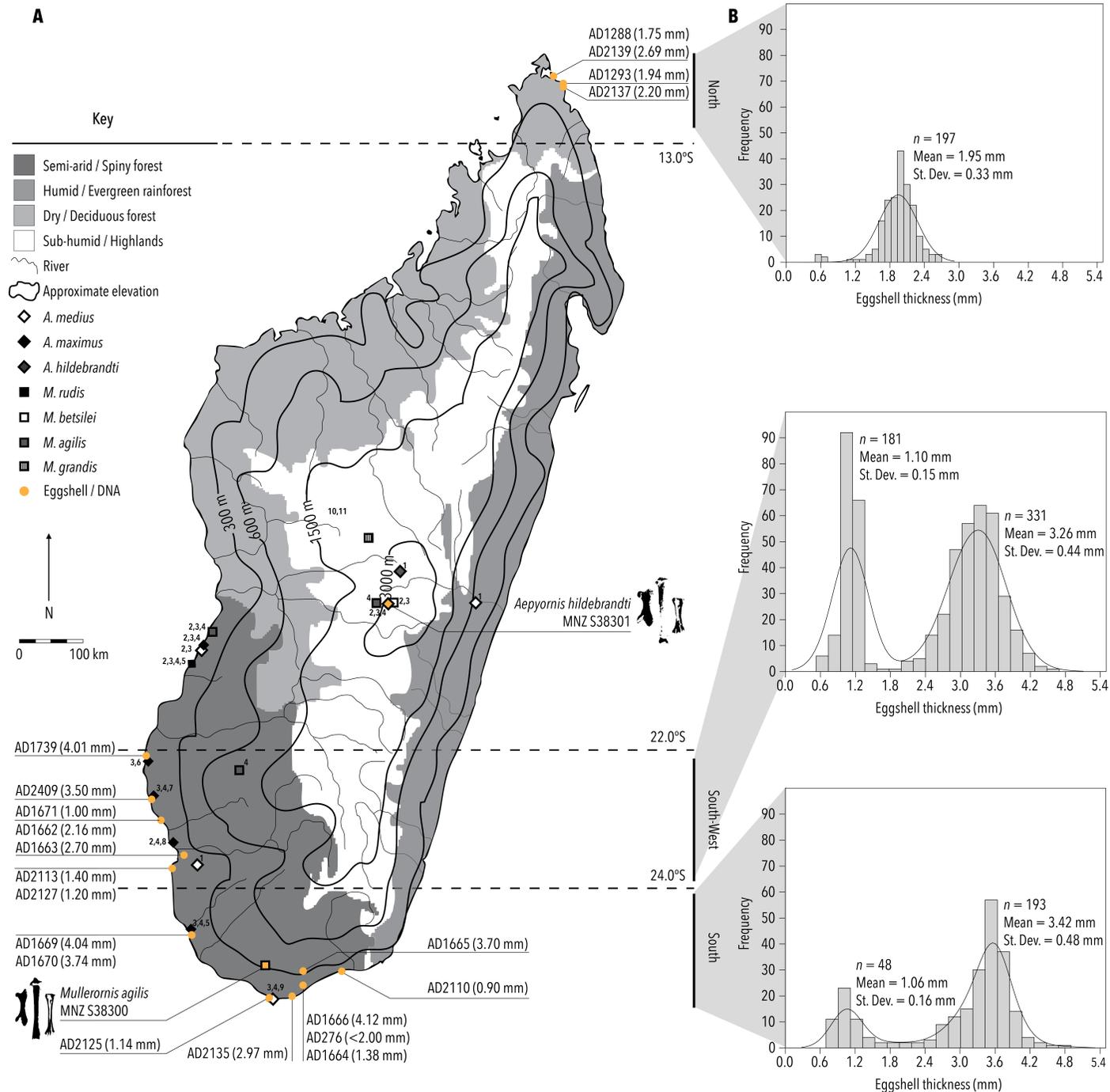
---

### **6.4.1 EGGSHELL SPECIMEN COLLECTION**

Eggshell specimens were collected over several field seasons from various locations in the North, South, and Southwest of Madagascar (Figure 6.1) by GM, AG, KD, and JH (S6.8.1). Eggshell was stored at room temperature.

### **6.4.2 CHARACTERISATION *of* EGGSHELL THICKNESS MORPHOTYPES**

The average thickness of each eggshell sample calculated as the mean of the thicknesses of four sizes measured with a digital caliper. The thickness of 197 elephant bird eggshells collected from the North, 512 eggshells randomly collected from the Southwest, and 241 eggshells collected from the South of Madagascar (Figure 6.1) were also measured to examine the distribution of eggshell thicknesses from each region. Summary statistics for these distributions were calculated in PAST v.3.11 (Hammer et al. 2001).



---

**FIGURE 6.1** | **A** MAP OF MADAGASCAR DEPICTING THE GEOGRAPHIC LOCATION OF EGGSHELL SAMPLES COLLECTED AND ANALYSED (circles). Samples are represented by their ID# and the thickness of the sample is given in mm in the brackets to the right of the sample number. The location of fossil specimens of *Aepyornis* (diamonds) and *Mullerornis* (squares) are shown. Superscripts next to the symbol are primary and secondary references for those fossils (see footnotes). Specimens for which DNA data were available are coloured yellow. Simplified topography of the landscape is shown with approximate elevation represented by contour lines, rivers represented by fine lines, and biomes represented by shades of grey. **B** The distribution of eggshell thicknesses derived from the total number of eggshells collected across the range in the North, Southwest, and South of Madagascar.

### 6.4.3 RADIOCARBON DATING

Eggshell samples for radiocarbon dating were mechanically cleaned then reduced by 50% with the stoichiometric addition of 2N HCl *in vacuo*. Cleaned fragments were converted to graphite at the INSTAAR Laboratory for AMS Radiocarbon Preparation and Research (NSRL) before measurement by Accelerator Mass Spectrometry at the Keck Carbon Cycle AMS Laboratory at the UC Irvine (KCCAMS). Conventional radiocarbon ages have been calibrated using CALIB v.7.1 and SHcal13 (Stuiver and Raimer 1993; Stuiver et al. 2005; Hogg et al. 2013). Sample AD1739 was found in an archaeological deposit by KD; eggshell from the same context were radiocarbon dated as above.

### 6.4.4 AMINO ACID RACEMISATION

The ratio of two enantiomers (A/I), the protein amino acid L-isoleucine and the non-protein diastereomer D-alloisoleucine, was measured in eggshell by ion-exchange high-pressure liquid chromatography (S6.8.2); A/I reflects time and the integrated thermal history experienced by the sample. Quality control is monitored with a laboratory standard, ILC-G (Wehmiller 2013); 383 A/I analyses of the ILC-G standard in the lab average  $0.457 \pm 0.012$ . Levels of amino-acid racemisation in the eggshell samples were used to prioritise samples for aDNA extraction.

### 6.4.5 ADNA EXTRACTION *from* EGGSHELL

Ancient DNA was extracted from 33 eggshell samples across each thickness from each location (North, South, and Southwest; Figure 6.1). Samples were prioritised for DNA extraction based on their level of amino-acid racemisation (with low-levels of racemisation preferred) as well as on their exposure to the environment at the time of collection (with those found buried prioritised; Table S6.8.1). The same locality was not sampled twice for DNA in order to minimise the chance that two samples may have come from the same egg or same female. Ancient DNA was extracted from 200 mg of eggshell powder per sample in the Trace Advanced Ultra-Clean Environment (TrACE) at Curtin University, WA (Australia) following the

protocol described by Dabney et al. (2013) with minor changes (S6.8.3), and in keeping standard aDNA practice (Willerslev and Cooper 2005; Knapp et al. 2012).

#### **6.4.6 SHOTGUN LIBRARY PREPARATION**

Shotgun sequencing libraries were prepared following the protocol described by (Gansauge and Meyer 2013) with minor changes (S6.8.4; Table S6.8.2).

#### **6.4.7 HYBRIDISATION ENRICHMENT of MITOCHONDRIAL DNA**

3,083 80-mer mitochondrial baits with 4X (20 bp) tiling were designed based on a consensus sequence of two published *Aepyornis* (Mitchell et al. 2014a, NCBI accession #KJ749824; Grealy et al. 2017, NCBI accession #KY412176) reference genomes and one *Mullerornis* reference genome (Mitchell et al. 2014a, NCBI accession #KJ749825), and were manufactured through *MYcroarray*. Hybridisation enrichment of mitochondrial DNA was performed by following the MYbaits (*MYcroarray*) protocol (v.3, 2015) as per the manufacturer's instructions, with minor changes (S6.8.5; Table S6.8.3).

#### **6.4.8 HIGH-THROUGHPUT DNA SEQUENCING**

Enriched libraries were quantified using a LabChip GX Touch HT (*Perkin Elmer*) following the manufacturer's instructions (S6.8.6), and were pooled in equimolar concentrations in a total volume of 60 µl. In order to remove low-molecular weight primer dimer and library-build/capture artifacts, fragments between 140 bp and 300 bp were size-selected from the pooled library using two lanes of a Pippin Prep (*Sage Science*) eGel cassette following the manufacturer's instructions. The two lanes of size-selected library were recombined, and were purified and concentrated through a *QIAGEN* PCR Purification kit, following the manufacturer's instructions, with minor changes (S6.8.7). The final sequencing library was quantified again on the LabChip GX Touch HT. The library was diluted to 4 nM in ultrapure water and was sequenced using *Illumina*'s high-throughput platform NextSeq, following the manufacturer's instructions with minor changes (S6.8.8; Table S6.8.4).

#### **6.4.9 QUALITY CONTROL *and* FILTERING**

Sequences were trimmed using USEARCH v.8 (Edgar 2010) and sequences below 30 bp in length were discarded, as they could not be meaningfully mapped to reference genomes. USEARCH v.8 was used to quality filter sequences (by employing an expected error rate of 1% of the length of the sequence), find unique sequences, and remove chimeric sequences (S6.8.9; Table S6.8.5).

#### **6.4.10 RECONSTRUCTION *of* MITOCHONDRIAL GENOMES**

For each sample, sequences were iteratively mapped against a consensus elephant bird reference mitochondrial genome in Geneious v.8.1.6 (Kearse et al. 2012) using the default parameters under a ‘medium-low sensitivity’ option with 10 iterations. Mapped reads were then aligned to NCBI’s GenBank reference database (Benson et al. 2006) using BLAST 2.2.30+ (Altschul et al. 1990) implemented through the Pawsey Centre’s supercomputing facilities in order to obtain taxonomic assignments for the sequences. The BLASTn algorithm parameters evoked were as described by Grealy et al. (2017). Sequence taxonomy was assessed in MEGAN v.4.70.4. (Huson et al. 2007; S6.8.10). To remove potential contaminating sequences, reads aligning best to avian reference genomes were remapped onto the consensus genome generated from the last round of mapping, as before. A final strict consensus sequence with 50% majority-ruled based calling was generated, with positions having a coverage of less than two called as an ‘N’ and positions with no data represented by ‘?’. These final mitochondrial genomes can be found in GenBank (accession #: TBA) or downloaded from DataDryad (doi: TBA). The authenticity of mapped reads was assessed by charting the frequency of nucleotide substitutions across reads in mapDamage v.2.0.6 (Ginolhac et al. 2011; Jonsson et al. 2013; S6.8.10; Figure S6.8.1).

#### **6.4.11 PHYLOGENY RECONSTRUCTION *and* MOLECULAR DATING**

20 elephant bird mitochondrial genomes were aligned with two previously published elephant bird mitochondrial genomes (Mitchell et al. 2014a) and eight outgroup

ratites (Table S6.8.6) using MAFFT (Katoh et al. 2002) and MUSCLE (Edgar 2004) as implemented in Geneious v.8.1.6 (Kearse et al. 2012) using the default parameters. All protein-coding, rRNA, and tRNA genes as well as the control region were extracted from the alignment and partitioned (S6.8.11) by codon position (protein-coding genes), and loops and stems (RNA genes). RCV and stemminess tests (Phillips and Pratt 2008) were performed in PAUP v.4a150 (Swofford 2003) as described previously (Greal et al. 2017) with minor changes (S6.8.12; S6.8.13). These tests were used to assess base composition bias and the extent of phylogenetic signal erosion in order to determine which partitions may benefit from RY coding that will alleviate the biases (Table S6.8.7; Table S6.8.8; Table S6.8.9). The best-fitting substitution model for each partition was determined using jModelTest v.2.1.7 (Guindon et al. 2003; Darriba et al. 2012; S6.8.14; Table S6.8.10). Mitochondrial phylogenetic trees were constructed on both standard-coded and completely RY-coded data, using maximum likelihood and Bayesian approaches implemented in RAxML (Stamatakis 2014) and MrBayes (Huelsenbeck and Ronquist 2001, executed through the CIPRES online bioinformatic toolkit; Miller et al. 2010), respectively (S6.8.15; S6.8.16). Tracer v.1.6.1 was used to examine the convergence of Bayesian runs (Rambaut et al. 2003; S6.8.16).

Molecular dating of both standard and RY-coded datasets was performed using MCMCTree (Yang and Rannala 2006) implemented in PAML v.4.8 (Yang 2007) as previously described (Greal et al. 2017) with minor changes (S6.8.17). Two fossil-based age priors were used for calibration (Table S6.8.11).

#### **6.4.12 ASSESSMENT of GENETIC VARIATION in BARCODING GENE *COI***

In order to determine whether the clades identified through phylogenetic analysis might represent different species, genetic distance within and between elephant bird specimens from each region exhibiting less than 10% missing data across 596 bp of *cytochrome oxidase I (COI)* were calculated in MEGA v.6.06 (Tamura et al. 2013) using the Kimura 2-parameter model (Kimura 1980) with pairwise deletion of missing data alongside default parameters for the remaining options (S6.8.18). To gauge the limits of intra- and inter-specific variation in this barcoding region, the

distance within and between genera of moa, rhea, emu, cassowary, and kiwi were also estimated in the same way using published sequences (S6.8.18).

#### 6.4.13 CHARACTERISATION OF EGGSHELL MICROSTRUCTURE

20 eggshell samples of varying thicknesses across each location were imaged using micro computed tomography (Skyscan 1175 micro-CT, *Bruker-microCT*) at the Centre for Microscopy, Characterisation and Analysis, The University of Western Australia (S6.8.19, Table S6.8.12; n.b., note that not every eggshell imaged yielded aDNA and some eggshell that yielded DNA did not leave enough sample to be imaged). Analyses were performed on a central slice in order to minimise the effects of imaging artifacts and potential weathering of pores close to the surfaces of the eggshell. For each sample, 2D analyses of pore density and pore area were performed on a 20.07 mm<sup>2</sup> region of interest (ROI; Figure S6.8.2) from the central slice, alongside 3D analysis of pore volume and percentage porosity on a 1 mm-thick volume around the central ROI, using Bruker CTAn v.1.16.4.1+ software (S6.8.19; SkyScan 2003-2011, *Bruker microCT* 2012-2016). Outer, inner, and pore structure surface images were rendered in Bruker CTAn and visualized in FEI Avizo Fire v.8.1.1 (*Konrad-Zeuse-Zentrum* Berlin 1995-2014; FEI, SAS 1999-2014; S6.8.19). One-way ANOVA and pairwise Student's t-tests with a Bonferroni correction were performed in PAST v.3.1.1 (Hammer et al. 2001) to compare the pore density, average pore volume, and percent porosity between eggshell morphotypes. Outliers (typically samples exhibiting imaging artifacts even after denoising) were excluded from these analyses.

#### 6.4.14 STABLE ISOTOPE ANALYSIS

Samples were prepared for isotopic analysis following the procedures outlined in Miller et al. (2005; S6.8.20).  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  were determined using an elemental analyzer (NC 2500; *CE Elantech*, Lakewood, NJ) interfaced with *Thermo Finnigan* (San Jose, CA) Delta Plus XL or Delta V Plus mass spectrometers (Carnegie Institution of Washington, Washington, DC; for details, see S6.8.20).

## 6.5 RESULTS *and* DISCUSSION

---

The rarity of elephant bird skeletons has limited our ability to explore the evolutionary relationships among them. However, features of elephant bird fossil eggshell may provide clues about the evolution of these birds instead. Here we detail the first phylogeographic survey of elephant birds, which has been achieved through the application of various morphological and molecular assays applied to a relatively large set of eggshells from across Madagascar.

### 6.5.1 RADIOCARBON DATING

We provide nine new radiocarbon dates for elephant bird eggshells that date to between approximately 1300 BP and 2000 BP (Table S6.8.1). The youngest sample dates to  $1305 \pm 15$  years BP, suggesting that elephant birds were extant at this time but may have become extinct soon thereafter. These dates are consistent with dated skeletal specimens (Turvey 2009). The temporal difference between these late Holocene samples is close enough to be considered negligible in terms of the effect on genetic distance.

### 6.5.2 EGGSHELL MORPHOTYPES

In Madagascar's South and Southwest, a bimodal distribution of thicknesses was observed, with each mode corresponding to two morphotypes of eggshell, one that is, on average, less than 1.1 mm thick, and one that is over twice as thick exhibiting a mean thickness of more than 3 mm (Figure 6.1). The distribution of eggshell thicknesses across the South and Southwest regions of Madagascar do not appear to be different to one another, although they significantly deviate from normality such that this could not be statistically confirmed with typical parametric two-sample tests. It is generally accepted that thickness of the eggshell is unique to the genus of elephant bird that laid it, with a 'thin' (i.e., less than 1.5 mm thick) morphotype corresponding to birds of the more gracile genus *Mullerornis*, and a 'thick' (greater than 2 mm thick) morphotype corresponding to birds of the more robust genus *Aepyornis* (Clarke 2006; Tovondrafale et al. 2014). However, moa eggshell displays a continuum of thicknesses and is "a poor discriminator for species assignment of

shell fragments in moa” as it varies widely within and between the 10 currently accepted species (Oskam et al. 2011; Huynen et al. 2010). To date, eggshells have not been found in association with identifiable elephant bird skeletal fossils (except one embryonic skeleton inside an intact egg<sup>1</sup>; Balanoff and Rowe 2007). Therefore, whether there is a significant correlation between elephant bird taxonomy and eggshell thickness “awaits verification from techniques such as those of contemporary molecular biology” (Clarke 2006).

The distribution of thicknesses of eggshell from the North of Madagascar is unimodal, with the average thickness being almost entirely intermediate between the ‘thin’ and ‘thick’ morphotypes of the South / Southwest at 1.95 mm (Figure 6.1). This pattern could indicate that a kind of disruptive selection (where extreme thickness phenotypes are favoured) may have been operating in South and Southwest of Madagascar, while the Northern phenotype has arisen via stabilising selection (where an intermediate phenotype is favoured). However, selection may be operating on another phenotypic trait, such as body size, that would have an indirect effect on eggshell thickness.

### 6.5.3 MITOCHONDRIAL GENOMES

An average of 3,045,033 filtered, unique sequences were retrieved per sample across 32 eggshell samples of varying thicknesses from across the North, Southwest and South of Madagascar (Table S6.8.5). Of these, an average of 1,298 shotgun avian reads mapped onto the elephant bird consensus reference mitochondrial genome (0.02% endogenous mitochondrial aDNA), while an average of 41,115 captured avian reads mapped onto the same reference (7.03% endogenous mitochondrial aDNA), resulting in an average fold-enrichment of over 300X (Table S6.8.5). On average, mapped reads had a mean length of approximately 44 bp and a maximum length of 101 bp (Table S6.8.5). 1,756 reads were sequenced from a shotgun negative control library; however, none of these reads mapped to the elephant bird

---

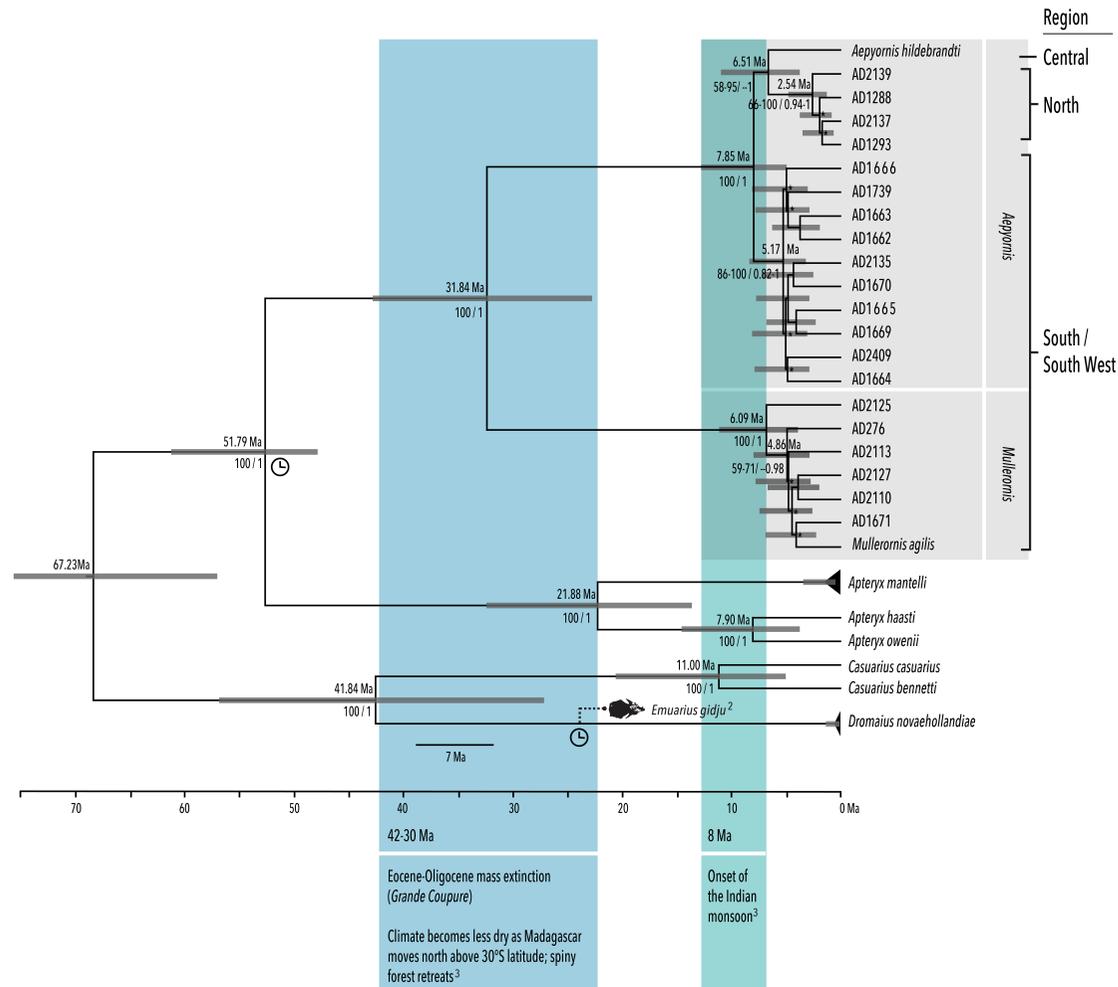
<sup>1</sup> This specimen was identified as *Aepyornis* based on the size of the eggshell; however, the thickness of the eggshell is not reported and as the skeleton is embryonic it cannot definitively be compared to type specimens.

mitochondrial reference genome, suggesting that avian contamination arising from the laboratory environment was negligible.

We retrieved 16 near-complete (more than 15,000 bp, average coverage 236X) and four partial (more than 10,000 bp, average coverage 19X) mitochondrial genomes (Table S6.8.5). The damage profile of mapped reads, generated through MapDamage v.2.0 (Ginolhac et al. 2011; Jonsson et al. 2013), showed a higher proportion of cytosine-to-thymine nucleotide misincorporations at both the 5' and 3' ends, which is characteristic of *bona fide* aDNA (S6.8.10; Figure S6.8.1). Less than 5,000 bp were retrieved from an additional 13 samples (average coverage 0.7X), with seven of these yielding less than 200 bp of sequence with an average coverage of 0.2X (Table S6.8.5). These 13 samples were not included in subsequent analyses due to their lack of informative sequence data, propensity to reduce the number of informative sites able to be included, and an inability to verify the authenticity of the aDNA through an examination of their damage profiles. Samples that yielded the most aDNA were typically younger than 2000 years BP, were found in buried deposits (as opposed to surface scatter), and tended to exhibit lower levels of amino-acid racemisation (Table S6.8.1).

#### 6.5.4 ELEPHANT BIRD PHYLOGEOGRAPHY *and* TAXONOMY

**GENETIC RELATIONSHIPS.** The elephant birds separate into two distinct monophyletic clades corresponding to the two genera that have been described from skeletal morphology (i.e., *Aepyornis* and *Mullerornis*): all eggshell samples thinner than 1.5 mm in thickness fall into a cluster with the *Mullerornis* genome previously derived (Mitchell et al. 2014a) from a vouchered bone specimen, and similarly, all eggshell samples greater than 1.5 mm in thickness cluster with the *Aepyornis* genome previously reconstructed (Mitchell et al. 2014a) from a vouchered bone specimen (Figure 6.2). Regardless of RY-coding scheme, these groupings receive the highest statistical support from both maximum likelihood and Bayesian analyses, with bootstrap support for these nodes being 100% and Bayesian posterior probabilities being 1 (Figure 6.2). Furthermore, the average genetic distance in an approximately 600 bp barcoding region of *cytochrome oxidase I* (*COI*) between these two clades is approximately ten times greater (11.6%) than the average genetic



**FIGURE 6.2 | MOLECULAR DATED PHYLOGENETIC TREE** showing the relationships between eggshell specimens from the North, South, and Southwest of Madagascar. Fossil calibrated notes are indicated by a clock symbol. Grey bars represent 95% HPDs of the date estimate for the node. Numbers below nodes give the ML bootstrap support and Bayesian posterior probability for both standard-coded and fully RY-coded mitochondrial datasets. Notes marked by an asterisk had less than 70% ML bootstrap support. The timing of major geological and climatic events are indicated on the timeline below the tree.

**TABLE 6.2** | ESTIMATES OF AVERAGE EVOLUTIONARY DIVERGENCE OVER SEQUENCE PAIRS WITHIN AND BETWEEN GROUPS ACROSS A 596 BP BARCODING REGION OF COI FOR ELEPHANT BIRDS AND OTHER RATITES. The number of base substitutions per site from averaging over all sequence pairs within each group are shown. Analyses were conducted using the Kimura-2-parameter model. Codon positions included were 1<sup>st</sup>+2<sup>nd</sup>+3<sup>rd</sup>+Noncoding. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA6. Minimum and maximum genetic distance is shown in brackets where possible). *n* refers to the number of sequences compared.

Comparison	Average Genetic distance across 596 bp of COI (K2P)
Within 8 moa species ( <i>n</i> =33)	0.005 (0-0.014)
Within 2 rhea species ( <i>n</i> =10)	0.001
Within 1 emu species ( <i>n</i> =2)	0.000
Within 2 cassowary species ( <i>n</i> =4)	0.000
Within 4 kiwi species ( <i>n</i> =26)	0.001 (0-0.003)
Between moa 7 species within 3 genera ( <i>n</i> =26)	0.019 (0.012-0.029)
Between 2 rhea species within 1 genus ( <i>n</i> =10)	0.072
Between 2 cassowary species within 1 genus ( <i>n</i> =4)	0.033
Between kiwi 4 species within 1 genus ( <i>n</i> =26)	0.051 (0.017-0.075)
Between 6 moa genera ( <i>n</i> =26)	0.048 (0.023-0.064)
Within <i>Aepyornis</i> ( <i>n</i> =9)	0.013
Within <i>Mullerornis</i> ( <i>n</i> =4)	0.004
Between <i>Aepyornis</i> and <i>Mullerornis</i>	0.116
Within <i>Mullerornis</i> South/Southwest ( <i>n</i> =4)	0.004
Between <i>Mullerornis</i> South/Southwest and <i>M. agilis</i>	0.004
Within <i>Aepyornis</i> South/Southwest ( <i>n</i> =6)	0.005
Within <i>Aepyornis</i> North ( <i>n</i> =3)	0.002
Between <i>Aepyornis</i> South/Southwest and North	0.013
Between <i>Aepyornis</i> South/Southwest and Central <i>A. hildebrandti</i>	0.014
Between <i>Aepyornis</i> North and Central <i>A. hildebrandti</i>	0.003

distance within each clade (1.3% for *Aepyornis* and 0.4% for *Mullerornis*; Table 6.2); this is comparable to other ratites such as moa, where the average within-genus genetic distance is 1.9% (between 1.2-2.7%) while the average distance between genera is 4.8% (2.3-6.4%) at the same locus. Thus, these data confirm that the two modes of eggshell thickness observed in South/Southwest Madagascar (Figure 6.1) correspond to the two genera of elephant birds. While the genetic evidence supports the hypothesis that there are two different genera of elephant birds, the high level of divergence between them in *COI* (almost double that of moa) may suggest that they belong to more than one family; currently, both genera are classified within the family Aepyornithidae, but reclassification of *Mullerornis* into a new family (perhaps “Mullerornithidae”) may be appropriate, pending further molecular or morphological research.

Within *Mullerornis*, there appears to be no correlation between geographic region and the genetic relationships among specimens, with some samples from the Southwest being more closely related to samples in the South than others from the Southwest, and vice versa, albeit with low support (Figure 6.2). For example, AD#2127 from the Southwest is more closely related to AD#2110 from the South than it is to other samples from the Southwest such as AD#1671 (Figure 6.2). Furthermore, the voucher *Mullerornis agilis* specimen identified from Southern Madagascar is more closely related to a specimen from the Southwest than other Southern specimens (this specimen is consistently nested in the clade with all other putative *Mullerornis* specimens to the exclusion of AD#2125 with high support (71% bootstrap support, 0.98 posterior probability), except with extreme RY-coding, with 59% ML-bootstrap support and no support from Bayesian analyses. Support for the alternative grouping of all South/Southwestern specimens to the exclusion of *M. agilis* by Bayesian posterior probability is 0.93). Low and inconsistent statistical support for the groupings within the genus *Mullerornis* suggest that all samples represented belong to the same species, and that this species is *M. agilis*. This interpretation is supported by comparisons of the genetic distance within and between *Mullerornis* specimens: the average genetic distance in *COI* within *Mullerornis* from the South is 0.2%, while the average genetic distance between South and Southwest *Mullerornis* specimens is 0.5%, between South *Mullerornis* specimens and *M. agilis* is 0.3%, and between Southwest *Mullerornis* specimens and

*M. agilis* is 0.7% (Table 6.2). Grouping the South and Southwest specimens, the overall genetic distance within them is 0.4%, with the South/Southwest clade being 0.4% different from *M. agilis* (Table 6.2). As the genetic distance in *COI* within ratite species is typically less than 1% (Table 6.2), the distances observed between South, Southwest specimens and *M. agilis* suggest that, relative to other ratites, they are not genetically different enough to be considered different species. Thus, *M. agilis* appears to be the only species within *Mullerornis* from the coastal areas of South and Southwest Madagascar. Although we did not sample around the other locations where *Mullerornis* species have been found (e.g., West Madagascar), these results hint that there could be far fewer species than have been morphologically described (four; Table 6.1).

Likewise, within *Aepyornis* there is no division between samples from the South and Southwest (Figure 6.2); in all analyses, several samples from the Southwest are more closely related to samples from the South than they are to their Southwestern neighbors, and vice versa, albeit with low support. Although this does not rule out the possibility that two species of *Aepyornis* (i.e., *A. medius* and *A. maximus*) coexisted across the South/Southwest, the genetic distance in *COI* within *Aepyornis* specimens from the South is 0.5%, while the distance between *Aepyornis* specimens from the South and Southwest is 0.4%, comparable to the within-species *COI* variation in other ratites at < 1% (Table 6.2). There is also no genetic distinction between thinner (< 3 mm) and thicker (> 3 mm) *Aepyornis* eggshell samples from the South/Southwest, with some thick samples more closely related to thin samples than other thick samples (e.g., AD#2409 and AD#1664). This result suggests that the thinner eggshells from the South/Southwest simply fall within the left tail of the distribution of thicknesses for *Aepyornis* in the South (Figure 6.1), rather than being a genetically or taxonomically distinct unit. Relationships among samples from the South/Southwest are weakly supported, and branch lengths are extremely short, suggesting that the samples represented here belong to the same species of *Aepyornis*, potentially *Aepyornis maximus* that has been described from skeletal fossils (although aDNA from a vouchered skeletal specimen of *A. maximus* would be

needed to confirm this hypothesis)<sup>2</sup>. However, there may be significant genetic structure between populations of this species throughout the South and Southwest that we were unable to resolve with mitochondrial DNA; highly polymorphic regions of the nuclear genome such as microsatellites, or even the mitochondrial hypervariable D-loop, may reveal additional genetic diversity within *Aepyornis* from the South/Southwest. Nevertheless, these results advocate for fewer species of *Aepyornis* in Southern/Southwestern Madagascar than have been previously described from morphological analysis of the skeletal fossils: *A. medius* (synonym *A. grandideri*) and *A. maximus* have both been found in Madagascar's South and Southwest, but it is possible that one represents sexually dimorphic members of the opposite sex. Similar to moa (Bunce et al. 2003; Huynen et al. 2003) and other ratites including the kiwi where females can be between 20-80% bigger than males, as suggested by Hume and Walters (2012), the smaller *A. medius* may represent males of the species, while the larger specimens identified as *A. maximus* represent females of the same species. Sex typing of vouchered bone specimens from these species would be required to elucidate this, and in this endeavor aDNA derived from eggshell is of little use as the DNA, if originating from the mother, is expected to be female. Finally, not all areas in Southern Madagascar were sampled; indeed, more species may have existed through Southeast Madagascar or South Central Madagascar where the environment and climate significantly differs from the South/Southwest (Figure 6.1).

One of the most significant findings of this study is that there is a clear and consistent division between *Aepyornis* samples from the North and samples from the South/Southwest of Madagascar (Figure 6.2). Samples from the South/Southwest always form a monophyletic clade with high statistical support (bootstrap support = 88-100%, posterior probability = 0.82-1). Northern samples also form a monophyletic clade with high statistical support (bootstrap support = 66-100%, posterior probability = 0.94-1). This topology was retrieved regardless of whether some or all partitions were RY-coded, and regardless of whether one or 19 outgroups were included, indicating that the division observed between Northern and

---

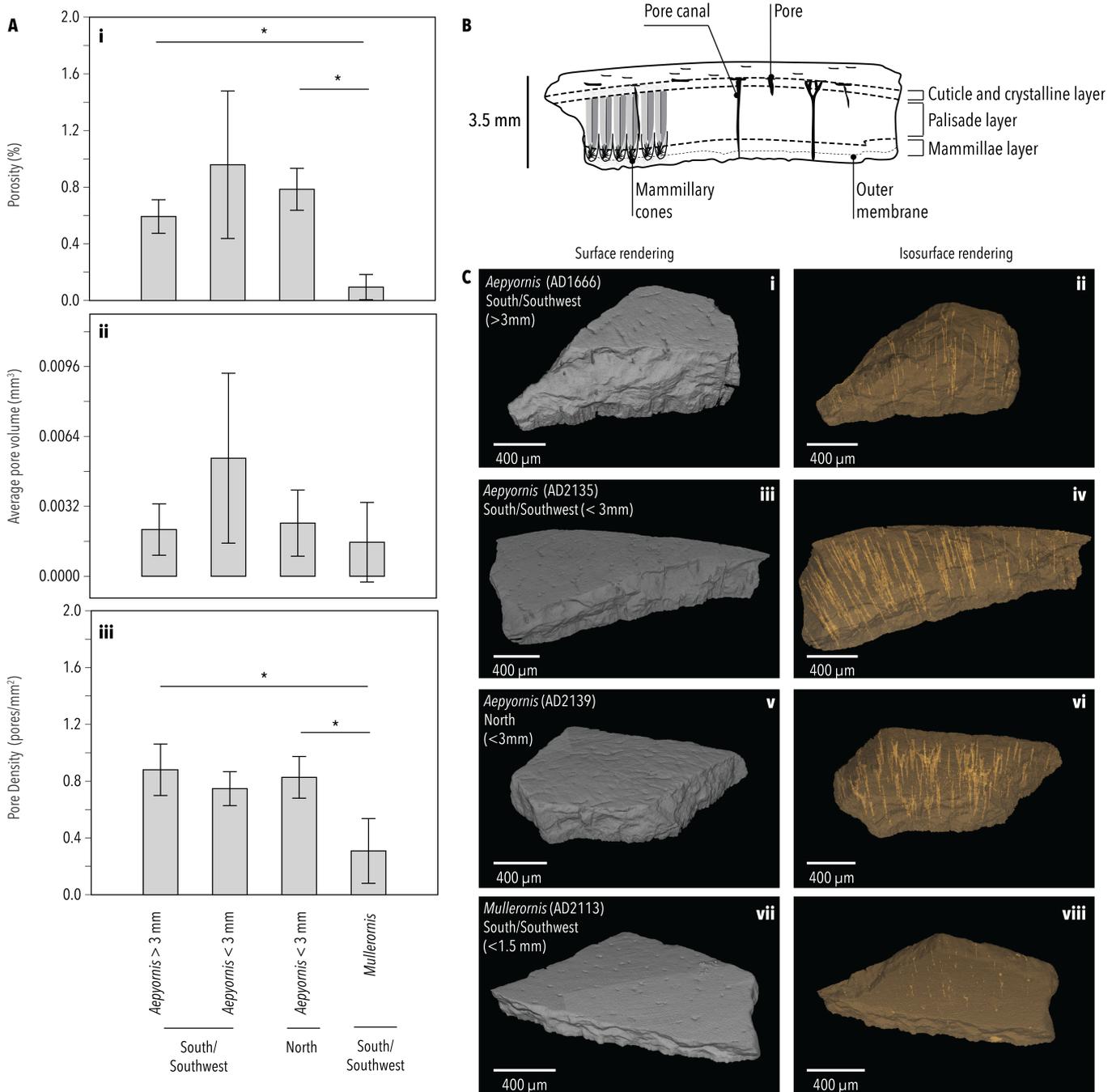
<sup>2</sup> Yonezawa et al. (2016) sequenced a mitochondrial genome from an *Aepyornis maximus* skeletal specimen that is currently being added to these phylogenetic analyses.

Southern *Aepyornis* is robust in the face of alternative analysis. Analysis of the genetic variation in *COI* supports this result, as the average genetic distance within the Northern clade is 0.2%, while the average genetic distance between the Northern and South/Southwestern clade is 1.3%; above the level of intra-specific variation observed in ratites and other birds (<1%) and within the limits of inter-specific variation which is estimated to be between 1.2-7.5% (Table 6.2). It is also above the accepted cut-off for distinguishing between species in moa (the closest ecological analog of elephant birds) using *COI*, which is 1.25% (Lambert 2005; Huynen and Lambert 2014), as well as 260 species of North American birds (Hebert et al. 2004). The Northern eggshell clade is sister to the voucher *A. hildebrandti* specimen from Central Madagascar with 98% bootstrap support and a Bayesian posterior probability of 1 with standard coding (Figure 6.2). However, the placement of *A. hildebrandti* is somewhat unstable, with only 58% ML bootstrap support and no support from Bayesian analysis when fully RY-coded (Figure 6.2). Nevertheless, support for the alternate grouping of *A. hildebrandti* as sister to South/Southwest *Aepyornis* is equally ill-supported, only garnering a posterior probability of 0.42 and no support from ML analysis. The consistent exclusion of *A. hildebrandti* from the Northern clade with high confidence could potentially indicate that this clade belongs to a new, morphologically unidentified species of *Aepyornis*, or represent an evolutionarily significant unit, such a subspecies (e.g., *A. hildebrandti* subsp. *avaratraoolithus*). To the best of our knowledge, no skeletal specimens of *Aepyornis* have thus far been found in the far North (James Hansford, *pers. comm.*). Conversely, this relationship may suggest that the Northern eggshell belongs to the species *A. hildebrandti*, which could explain why the eggshell is, on average, statistically significantly thinner than their Southern counterparts, as the bones of *A. hildebrandti* are smaller than *A. maximus* (Hume and Walters 2012). Moreover, the genetic distance in *COI* between Northern *Aepyornis* and *A. hildebrandti* is 0.3% whereas the genetic distance between the South/Southwest *Aepyornis* and *A. hildebrandti* is 1.4% (Table 6.2), supporting the idea that the Northern *Aepyornis* could be classified as *A. hildebrandti* (or *A. hildebrandti* subsp.?), while the South/Southwestern *Aepyornis* are a different species, probably *A. maximus*. Again, *A. gracilis*, also described from Central Madagascar but less robust than *A. hildebrandti* (Hume and Walters 2012), may be the male member of *A. hildebrandti* and has been described as a “dubious species” (Hume and Walters 2012). These

results also could indicate that the distribution range of the species may be vast, spanning multiple climatic zones across more than 900 km (Figure 6.1), which would suggest that diversity within elephant birds is not only much lower than described from the fossil record, but also much lower than one might expect from a country with multiple discrete climatic zones (Figure 6.1), numerous geographic barriers to gene flow, and one of the highest faunal and floral biodiversities (particularly endemic biotas) in the world (Wilme' et al. 2006).

Alternatively, the short branch lengths among all *Aepyornis* specimens represented here are indicative of little divergence between them, and as such a genetic cline may exist within *Aepyornis*, in which there is gene flow between consecutive populations along the length of Madagascar (potentially around the coast), resulting in what appears to be different species at the extremes of the geographic range but what is, in reality, an effect of inadequate sampling along the gradient. Another explanation could be that, like moa, females hold home ranges while males travel long distances to breed; under this scenario, mitochondrial DNA would not fully reflect the relationships between elephant birds as it is inherited maternally. Analysis of nuclear aDNA would allow us to ascertain whether this is the case; however, again, depending on whether the DNA within eggshell originates from the mother or the fetus, it may not be possible to retrieve male DNA from eggshell.

**EGGSHELL MICROSTRUCTURE.** Avian eggshell porosity is a proxy the amount of gas exchange (oxygen, carbon dioxide, and water vapor) between the developing fetus and the environment (Burton and Tullett 1983). As such, the porosity of an eggshell, in addition to its thickness, can be telling about the “size of the egg, the time required for incubation, and the metabolic rate of the embryo” (Tullett 1984), as well as the environment in which it was incubated. Although some eggshell microstructures can be altered through diagenic modifications and embryonic development (for example, mammillary cones become degraded as the embryo draws calcium from the shell as it develops)—and indeed, even thickness can be influenced by diet and post-depositional weathering—pore density and volume through the centre of the eggshell are expected to remain unadulterated by weathering and may offer another means by which to distinguish taxa. However, we must take care not to over-interpret the results in light of possible post-mortem damage, scanning artifacts and limitations,

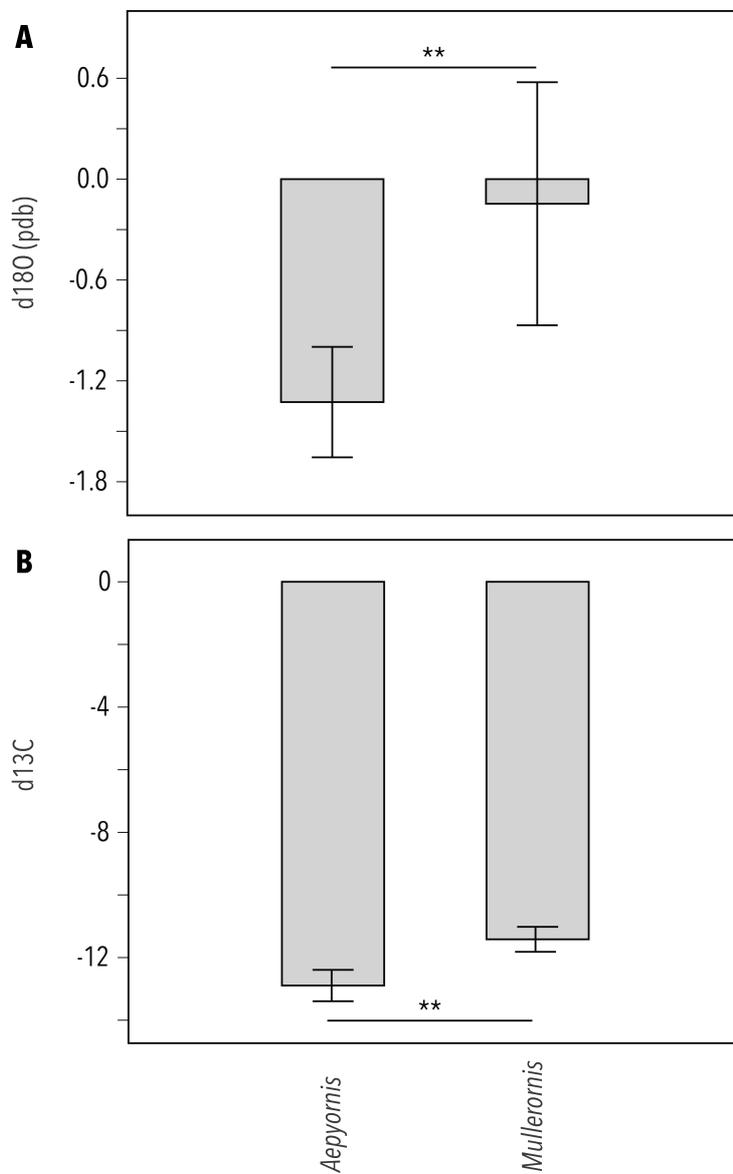


**FIGURE 6.3 | EGGSHELL MICROSTRUCTURE.** **A** Bar chart comparing the mean porosity, mean ‘average volume per pore’, and mean pore density within the ROI examined of each eggshell morphotype. Significant differences ( $p$ -value < 0.013) are indicated by an asterisk. **B** A cross-sectional schematic diagram of elephant bird eggshell macromolecular structure depicting the main morphological features, including both simple (unbranched) and multifurcate (branching) longitudinal pore structures. **C** Representative micro CT scans of each eggshell morphotype showing the outer surface and internal pore structures.

unknown within-egg variation, and small sample size ( $n=20$ , five per morphotype). Furthermore, other aspects of eggshell microstructure were not examined given the limits of resolution achievable with the technology used.

The pores in elephant bird eggshell, similar to some ratites (i.e., rheas) but very different from others (i.e., kiwi, ostrich), are relatively simple in structure (Tullett 1984), exhibiting some unbranched pores, some bifurcate branching pores, and some pores that ‘fan-out’ toward the outer surface of the eggshell (Figure 6.3b, c). Although some eggshells appear to have a greater proportion of branching pores to unbranched pores than others, this seems the presence of this feature appears to be random with respect to genus, thickness, or geographic origin. However, the porosity was statistically higher in *Aepyornis* eggshells from the South/Southwest ( $0.802\% \pm 0.395$  95% CI) than *Mullerornis* eggshells ( $0.095\% \pm 0.25395\%$  CI;  $p$ -value = 0.011,  $df = 10$ ,  $\alpha = 0.013$ ; Figure 6.3ai) from the same region. As porosity is a function of pore volume and pore density, this difference can be attributed to differences in pore density (Figure 6.3aiii) between South/Southwest *Aepyornis* ( $0.797$  pores/mm<sup>2</sup>  $\pm 0.126$  95% CI) and *Mullerornis* ( $0.309$  pores/mm<sup>2</sup>  $\pm 0.322$  95% CI;  $p$ -value = 0.001,  $df = 12$ ,  $\alpha = 0.013$ ) and not differences in the volume of pores ( $p$ -value = 0.283,  $df = 17$ ,  $\alpha = 0.013$ ; Figure 6.3aii). With a thinner and less porous eggshell than *Aepyornis* that inhabited the same environment, *Mullerornis* embryos were likely to have been smaller than *Aepyornis* embryos, which is no surprise given that skeletal morphology indicates that *Mullerornis* adults were smaller than *Aepyornis* adults (Hume and Walters 2012).

Within South/Southwest *Aepyornis*, no statistical difference was detected in porosity ( $p$ -value = 0.303,  $df = 6$ ,  $\alpha = 0.013$ ), average pore volume ( $p$ -value = 0.154,  $df = 9$ ,  $\alpha = 0.013$ ), or pore density ( $p$ -value = 0.256,  $df = 7$ ,  $\alpha = 0.013$ ; Figure 6.3a) between eggshell at the thicker-end of the distribution ( $> 3$  mm thick) versus eggshell at the thinner-end of the distribution ( $> 2$  mm but  $< 3$  mm thick). This supports the genetic evidence that suggests thin and thick *Aepyornis* eggshell from the South/Southwest belong to the same species. However, this finding does not take into account differences that may exist in the pore volume as it traverses the eggshell (e.g., tapering, branching), or that pore density can differ over the surface of the eggshell, pores often being more concentrated at the blunt pole of the eggshell.



**FIGURE 6.4** | A COMPARISON OF MEAN  $\delta^{13}\text{C}$  AND  $\delta^{18}\text{O}$  ISOTOPE CONTENT OF *AEPYORNIS* AND *MULLERORNIS* EGGSHELLS FROM SOUTH/SOUTHWEST MADAGASCAR. Asterisks indicate significant differences ( $p$ -value < 0.005).

There is a difference in porosity ( $p$ -value  $< 0.001$ ,  $df = 6$ ,  $\alpha = 0.013$ ; Figure 6.3ai) and pore density ( $p$ -value =  $0.006$ ,  $df = 9$ ,  $\alpha = 0.013$ ; Figure 6.3aiii) between Northern eggshell and the *Mullerornis* eggshell, supporting the classification of the Northern eggshell within the genus *Aepyornis*. However, there is no statistical difference in porosity ( $p$ -value  $> 0.118$ ,  $df = 5$ ,  $\alpha = 0.013$ ; Figure 6.3ai), average pore volume ( $p$ -value  $> 0.247$ ,  $df = 8$ ,  $\alpha = 0.013$ ; Figure 6.3aii), or pore density ( $p$ -value  $> 0.433$ ,  $df = 9$ ,  $\alpha = 0.013$ ; Figure 6.3aiii) between Northern eggshell and South/Southwest *Aepyornis* eggshell of either thickness. Because “species nesting in hot and humid environments [tend to] have a greater pore area than similarly sized eggs of species nesting in colder and drier environments” (Stein and Badyaev 2011), we would expect to observe that eggshells from the North, where average annual rainfall is higher than the South/Southwest, to exhibit greater porosity than eggshells of the same size from the Southern environment. However, “there is a wide variation in shell porosity between eggs within a species” (Burton and Tullett 1983), so it may not be surprising that we did not observe our expectations, especially given the limited sample size. Perhaps an equally porous but overall smaller egg would be selected in a humid environment. Although these data do not offer any insight into the phylogenetic affinities of the Northern eggshell, they also do not contradict the genetic evidence.

### 6.5.5 ELEPHANT BIRD DIVERSIFICATION

Molecular dating estimates that the divergence between *Aepyornis* and *Mullerornis* occurred approximately 31.84 Ma (95% HPD 22.38-42.12 Ma; Figure 6.2), which is consistent with the date estimated by Grealy et al. (2017) that was based on both mitochondrial and nuclear aDNA data. This timing coincides with the Eocene-Oligocene mass extinction (*Grande Coupure* faunal turnover), where global cooling took place. It also coincides with the time when Madagascar drifted north above 30°S in latitude: before this time, the climate was largely dry and the island was dominated by spiny forest, but as precipitation increased as Madagascar moved north during the Eocene, the range of this biome contracted towards the South/Southwest (Yoder and Nowak 2006) and the humid and dry forests of the North and East originated and expanded (Buerki et al. 2013). The changes in palaeoclimate and dominant vegetation (including origin of Madagascar’s endemic flora; Buerki et al.

2013) during this time may have driven divergence in elephant birds as it did in Madagascar's lemurs (Yoder and Yang 2004), giving rise to the two genera. In fact, there is a statistically significant ( $p$ -value = 0.002,  $df$  = 65,  $\alpha$  = 0.05; Figure 6.4b) difference in the mean organic carbon 13 isotope ( $\delta^{13}\text{C}$ ) content of South/Southwest *Aepyornis* and *Mullerornis* eggshells, where the average  $\delta^{13}\text{C}$  of *Aepyornis* is -22.9 (95% CI  $\pm$  0.52) while the average  $\delta^{13}\text{C}$  of *Mullerornis* is -21.42 (95% CI  $\pm$  0.432): for *Aepyornis*, this value falls within the distribution of  $\delta^{13}\text{C}$  for  $\text{C}_3$ -type vegetation (trees and shrubs), whereas for *Mullerornis*, values fall within the distribution of  $\delta^{13}\text{C}$  for CAM-type vegetation (succulents; Clarke et al. 2006). These results may answer the question posed by Clarke et al. (2006) as to why *Aepyornis* browsed on  $\text{C}_3$ -type foliage in a predominantly succulent landscape. A significant difference in eggshell oxygen 18 ( $\delta^{18}\text{O}$ ) isotopes ( $p$ -value < 0.001,  $df$  = 63,  $\alpha$  = 0.05; Figure 6.4a) is also observed between South/Southwest *Aepyornis* and *Mullerornis*, with the average  $\delta^{18}\text{O}$  of *Aepyornis* eggshells ( $-1.33 \pm 0.33$  95% CI) being significantly more negative than the average  $\delta^{18}\text{O}$  of *Mullerornis* eggshells ( $-0.15 \pm 0.78$  95% CI). According to Clarke et al. (2006), "trends towards more positive  $\delta^{18}\text{O}$  values in biominerals can be expected to reflect hotter, drier environments" and  $\delta^{18}\text{O}$  in *Aepyornis* eggshell indicates "a dependence on groundwater-fed perennial ponds during the breeding season": if *Mullerornis* primarily fed on succulent vegetation, they may not have relied as heavily on additional reservoirs to obtain enough water, which could account for the higher  $\delta^{18}\text{O}$  values observed in *Mullerornis* eggshell compared with *Aepyornis* eggshell. Alternatively, as Clarke et al. (2006) suggest, "it is plausible that *Aepyornis* subsisted elsewhere during the majority of the year, and migrated to the region to breed" while *Mullerornis* stayed put (however, this behavior has not been documented in other ratites). The range overlap between *Aepyornis* and *Mullerornis* along with the differences observed in  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$ , suggests that niche partitioning (at least, during the breeding season) occurred sometime during the evolution of elephant birds, perhaps around the time *Aepyornis* and *Mullerornis* lineages split—particularly if such a time was one of palaeoenvironmental upheaval, as occurred around the Eocene-Oligocene boundary. Alternatively, could the lineages have split when one rafted or swam to Madagascar from Africa across the Mozambique Channel during a time of low sea level, followed later on by the other? According to Goodman and Patterson (1997), favourable winds and currents before the end of the Eocene and low sea level during

the Eocene/Oligocene may have made crossings “somewhat easier”, although this scenario is unlikely.

There also appears to be a later diversification within both elephant bird genera. However, using older fossil calibrates to date the divergence of these younger nodes is inherently problematic: currently no established crown Aepyornithiform or Apterygiform fossils have been found that would offer a more accurate calibration for the molecular clock, and as such, the dates estimated typically have wide HPDs, and may appear older than they are. Nevertheless, diversification within both *Aepyornis* and *Mullerornis* is estimated to have occurred approximately 6-8 Ma (95% HPDs 3.72-12.55 Ma; Figure 6.2); this timing corresponds to a second diversification within lemurs between 8-12 Ma hypothesised to have been sparked by the onset of the Indian monsoon systems in Madagascar (Goodman and Patterson 1997; Yoder and Nowak 2006). Again, the increase in rainfall and humidity would have driven the evolution of rainforest vegetation and further established the distinct biomes seen in Madagascar’s East, North, and Central Highlands today (Figure 6.1). This supports the idea of separate *Aepyornis* species existing in Madagascar’s North and South/Southwest, as elephant bird populations would have been subjected to different selection pressures in these regions: in the South/Southwest, there would have been continued selection for drought tolerance, which may explain features of the eggshell that may provide resistance to desiccation (e.g., reduced pore density), while in the North, there may have been selection for smaller body-size that would be advantageous navigating through more humid forest, which could result in the thinner eggshell seen in the North.

Molecular dating also suggests that there is several million-year difference separating *A. hildebrandti* and the Northern specimens, whereas the divergence within *Aepyornis* South/Southwest and *Mullerornis* spans less than 1 Ma. The little genetic divergence observed in *COI* between *A. hildebrandti* and *Aepyornis* North may be because evolutionary rates of barcoding genes are typically slower in large-bodied birds with long generation times, causing them to appear more closely related than perhaps they truly are (Thomson et al. 2014). Several million years is arguably enough time for speciation to have occurred, which further speaks to the hypothesis that Northern and Central populations of *Aepyornis* are at least evolutionarily distinct

units, if not different species. Furthermore, it has been suggested that the valley known as the Mandritsara window “may have formed the isolation obstacle for certain organisms between the Northern Highlands and the Central Highlands” leading to the evolution of “species pairs” (Goodman 2005). In contrast, the rapid divergence within *Aepyornis* and *Mullerornis* genera in the South/Southwest, as well as the fact that mean  $\delta^{13}\text{C}$  and  $\delta^{18}\text{O}$  values of either genus do not significantly differ between eggshell from the South and the Southwest (*Aepyornis*:  $p$ -value  $> 0.88$ ,  $df = 47$ ,  $\alpha = 0.05$ ; *Mullerornis*:  $p$ -value  $> 0.19$ ,  $df = 16$ ,  $\alpha = 0.05$ ; Figure 6.4), adds support to the hypothesis that there was one species from each genus co-habiting the South/Southwest. Unfortunately, no isotope data is currently available for the eggshell from Madagascar’s North; however, this would be an ideal avenue to pursue in the future<sup>3</sup>.

## 6.6 CONCLUSION

---

The systematics of elephant birds has been confused since their discovery due to the scarcity of skeletal fossils, with little additional evidence found in the last several hundred years to refute or support initial classifications. Here, several lines of molecular evidence were employed to revisit the taxonomy of elephant birds and elucidate phylogeographic patterns. We used whole mitochondrial genomes extracted from Northern, Southern, and Southwestern elephant bird eggshell, eggshell macro- and microstructure, and isotope data to help resolve the evolutionary relationships within elephant birds. We find that the current classification of two elephant bird genera is supported by phylogenetic analyses, with thin *Mullerornis* eggshell and thicker *Aepyornis* eggshell forming monophyletic clades. Differences between the two genera are also observed in the microstructure of their eggshells, with *Mullerornis* exhibiting less dense pores than *Aepyornis*. Carbon and oxygen isotope data also indicate differences in the diet of South/Southwest *Mullerornis* and *Aepyornis* that may suggest niche partitioning, potentially initiating their split from a common ancestor around 30 Ma as Madagascar became less arid and spiny forest retreated during the Eocene-Oligocene. However, little genetic variation in barcoding

---

<sup>3</sup> Isotopic data from Northern eggshell is currently being collected and will be available in the published version of the article.

genes and low support for the phylogenetic topology within these clades suggests that species diversity within genera is lower than previously described, with one species of *Mullerornis* and one species of *Aepyornis* cohabiting the South/Southwest during the late Holocene; potentially more species existed at an earlier time that went extinct. Differences in skeletal morphology that were ascribed to different species may represent sexual dimorphism within a species, as in other ratites. The unique morphotype of eggshell found in Madagascar's far North places them unambiguously among *Aepyornis*, within a monophyletic clade of their own. The onset of the Indian monsoon season in Madagascar around 8 Ma may have driven divergence between Northern and Southern *Aepyornis* clades. Although little genetic divergence from their sister *A. hildebrandti* is observed in the Northern specimens, molecular dating suggests that there is ca. 4 Ma separating the two, suggesting that these eggshell may represent a novel oospecies or subspecies, particularly if they were geographically isolated. No differences in eggshell porosity were found between Northern and Southern *Aepyornis*, and the possibility of a genetic connectivity existing between populations of *Aepyornis* along the length of Madagascar cannot be ruled out without additional sampling. Nevertheless, we propose that a revision in elephant bird taxonomy and systematics is needed, incorporating a palaeogenomic perspective. A greater understanding of elephant bird genetic diversity and evolution that could be achieved with the addition of nuclear DNA will contribute to understanding how they lived and functioned within Madagascar's unique ecosystems, and why they eventually became extinct.

## 6.7 REFERENCES

---

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Balanoff AM and Rowe T. 2007. Osteological description of an embryonic skeleton of the extinct elephant bird, *Aepyornis* (Palaeognathae: Ratitae). *Society of Vertebrate Palaeontology Memoir* **9**: 1-53.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2006. GenBank. *Nucleic Acids Research* **34**: D16-D20.

- Brodkorb P. 1963. Catalogue of fossil birds. *Bulletin of the Florida State Museum Biological Sciences* **7**: 206-207.
- Buerki S, Devey DS, Callmander MW, Phillipson PB, Forest F. 2013. Spatio-temporal history of the endemic genera of Madagascar. *Botanical Journal of the Linnean Society* **171**: 304-329.
- Bunce M, Worthy TH, Ford T, Hoppitt W, Willerslev E, et al. 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**: 172-175.
- Burckhardt R. 1893. Ueber *Aepyornis* (*Ae. hildebrandti*). *Geological Magazine*.
- Burton FG and Tullett SG. 1983. A comparison of the effects of eggshell porosity on the respiration and growth of domestic-fowl, duck and turkey embryos. *Comparative Biochemistry and Physiology A-Physiology* **75**: 167-174.
- Clarke SJ, Miller GH, Fogel ML, Chivas AR, Murray-Wallace CV. 2006. The amino acid and stable isotope biogeochemistry of elephant bird (*Aepyornis*) eggshells from southern Madagascar. *Quaternary Science Reviews* **25**: 2343-2356.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**.
- Dewar R. 1984. Extinctions in Madagascar: the loss of the subfossil fauna. In. *Quaternary extinctions: A prehistoric revolution*. Tuscon: University of Arizona Press.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
- Gansauge MT and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737-748.
- Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**: 2153-2155.
- Goodman SM and Patterson BD. 1997. *Natural change and human impact in Madagascar*. Washington: Smithsonian Institution Press.
- Goodman SM, Rakotondravony D, Randriamanantsoa HN, Rakotomalala-Razanahoera M. 2005. A new species of rodent from the montane forest of central eastern Madagascar (Muridae: Nesomyinae: *Voalavo*). *Proceedings of the Biological Society of Washington* **118**: 863-873.
- Grealy A, Douglass K, Haile J, Bruwer C, Gough C, et al. 2016. Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar. *Journal of Archaeological Science* **75**: 82-88
- Grealy A, Phillips M, Miller G, Gilbert MTP, Rouillard J-M, et al. 2017. Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Molecular Phylogenetics and Evolution* **109**: 151-163.
- Guindon S and Gascuel, O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systems Biology* **52**: 696-704.
- Hammer Ø, Harper, DAT, Ryan PD. 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 9.

- Hebert PDN, Stoeckle MY, Zemplak TS, Francis C. 2004. Identification of birds through DNA barcodes. *PLOS Biology* **10**: 1657-1663.
- Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, et al. 2013. SHCAL13 Southern Hemisphere calibration, 0-50,000 years cal BP. *Radiocarbon* **55**: 1-15.
- Huelsenbeck JP and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Hume JP and Walters M. 2012. *Extinct Birds*. Bloomsbury Publishing.
- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Huynen L, Gill BJ, Millar CD, Lambert DM. 2010. Ancient DNA reveals extreme egg morphology and nesting behavior in New Zealand's extinct moa. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 16201-16206.
- Huynen L, Millar CD, Scofield RP, Lambert DM. 2003. Nuclear DNA sequences detect species limits in ancient moa. *Nature* **425**: 175-178.
- Huynen L and Lambert DM. 2014. Complex species status for extinct moa (Aves: Dinornithiformes) from the genus *Euryapteryx*. *PLOS One* **9**.
- Jonsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682-1684.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059-3066.

- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**:111-120.
- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage—Building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN, et al. 2005. Is a large-scale DNA-based inventory of ancient life possible? *Journal of Heredity* **96**: 279-284.
- Lambrecht K. 1933. *Handbuch der Palaeornithologie*. Berlin: Borntraeger.
- MacPhee RDE, Burney DA, Wells NA. 1985. Early Holocene chronology and environment of Ampasambazimba, a Malagasy subfossil lemur site. *International Journal of Primatology* **6**: 463-489.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*: 1-8.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014a. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Mitchell KJ, Wood JR, Scofield RP, Llamas B, Cooper A. 2014b. Ancient mitochondrial genome reveals unsuspected taxonomic affinity of the extinct Chatham duck (*Pachyanas chathamica*) and resolves divergence times for

- New Zealand and sub-Antarctic brown teals. *Molecular Phylogenetics and Evolution* **70**: 420-428.
- Oskam CL, Haile J, McLay E, Rigby P, Allentoft ME, et al. 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **277**: 1991-2000.
- Oskam CL, Jacomb C, Allentoft ME, Walter R, Scofield RP. 2011. Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *Journal of Archaeological Science* **38**: 2589-2595.
- Phillips MJ and Pratt RC. 2008. Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums). *Molecular Phylogenetics and Evolution* **46**: 594-605.
- Rambaut A, Suchard MA, Xie W, Drummond AJ. 2003. Tracer: MCMC Trace Analysis Tool. Version 1.6.1.
- Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.
- Stein LR and Badyaev AV. 2011. Evolution of eggshell structure during rapid range expansion in a passerine bird. *Functional Ecology* **25**.
- Stuiver M and Reimer PJ. 1993. Extended 14C data base and revised CALIB 3.0 14C age calibration program. *Radiocarbon* **35**: 215-230.
- Stuivier M, Reimer PJ, Reimer RW. 2005. CALIB 5.0.
- Swofford DL. 2003. PAUP\* Version 4.0b. Sunderland, Massachusetts: Sinauer Associates.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

- Thomson CE, Gilbert JDJ, Brooke M de L. 2014. Cytochrome b divergence between avian sister species is linked to generation length and body mass. *PLOS One* **9**.
- Tovondrafale T, Razakamanana T, Hiroko K, Rasoamiaramanana A. 2014. Palaeoecological analysis of elephant bird (Aepyornithidae) remains from the Late Pleistocene and Holocene formation of southern Madagascar. *Malagasy Nature* **8**.
- Tullett SG. 1984. The porosity of avian eggshells. *Comparative Biochemistry and Physiology A-Physiology* **78**: 5-13.
- Turvey S. 2009. *Holocene Extinctions*. Oxford Biology: Oxford.
- Wehmiller JF. 2013. Interlaboratory comparison of amino acid enantiomeric ratios in Pleistocene fossils. *Quaternary Geochronology* **16**: 173-182.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Wilme' L, Goodman SM, Ganzhorn JU. 2006. Biogeographic evolution of Madagascar's Microendemic biota. *Science* **312**. 1063-1065.
- Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586-1591.
- Yang Z and Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212-226.
- Yoder AD and Nowak MD. 2006. Has vicariance or dispersal been the prominent biogeographic force in Madagascar? Only time will tell. *Annual Review of Ecology, Evolution, and Systematics* **37**: 405-431.

Yoder AD and Yang Z. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Molecular Ecology* **13**: 757-773.

Yonezawa T, Segawa T, Mori H, Campos PF, Hongoh Y, et al. 2016. Phylogenomics and morphology of extinct palaeognaths reveal the origin and evolution of ratites. *Current Biology* **27**: 1-10.

## 6.8 SUPPLEMENTARY INFORMATION

---

### S6.8.1 EGGSHELL SPECIMEN COLLECTION

Eggshell of *Aepyornis* and *Mullerornis* were collected during field seasons conducted in 2006 and 2007 with the assistance of Ramilisonina of the Institut de Civilizations Musée d'Art et d'Archéologie, Antananarivo, Madagascar, and Retsihisatse Analamahery, Ambovombe, Madagascar, as well as Dr. Jean-Luc Schwenninger, Oxford University, Dr. John Magee, Australian National University, Canberra, and Steve DeVogel, University of Colorado. Eggshell collecting sites were chosen based on satellite imagery. Most eggshell were found as surface exposures in recently wind-deflated sand dunes, but some were collected *in situ* in vertical exposures. AD1739 was collected *in situ* by Kristina Douglass and MAP team during an archaeological excavation. Samples without a field ID (Table S6.8.1) were collected by Alicia Grealy, James Haile, Kristina Douglass and the MAP team in 2014, and were mostly found on the surface of rocky outcrops and beaches.

### S6.8.2 AMINO ACID RACEMISATION

The utility of avian eggshell for AAR has been recognized since the work of (Brooks 1990). Unlike molluscan biominerals, in which the majority of the protein residues are preserved in intercrystalline locations where they are subject to slow diffusional loss, eggshell protein is encapsulated within the calcite crystals and effectively immune from diffusional processes (Miller 2000). We rely on the extent of isoleucine epimerization, in which the protein amino acid L-isoleucine racemises about one of its two central carbon atoms to form the non-protein diastereomer D-alloisoleucine at a rate dependent on ambient temperature. The ratio of the two enantiomers (A/I) is measured by ion-exchange high-pressure liquid chromatography; A/I reflects time and the integrated thermal history experienced by the sample. Quality control is monitored with a laboratory standard, ILC-G (Wehmiller 2013) 383 A/I analyses of the ILC-G standard in our lab average  $0.457 \pm 0.012$ .

### S6.8.3 ADNA EXTRACTION

Ancient DNA was extracted from eggshell samples in the Trace Advanced Ultra-Clean Environment (TrACE) at Curtin University, WA (Australia). Personal protective equipment including coveralls, hairnet, double gloves, facemask, eyewear, and boots were worn to minimise the introduction of further exogenous DNA contamination. All surfaces were decontaminated with a solution of 10% bleach, followed by 70% ethanol as is the standard protocol for contamination avoidance (Willerslev and Cooper 2005; Knapp et al. 2012).

The surfaces of eggshell samples were cleaned with a solution of 10% bleach, followed by 70% ethanol. The surface layer was then removed and discarded through gentle grinding at low speed with a *Dremel* drill sporting a sterile drill bit. 400 mg of eggshell was then ground into a fine powder inside a sealed, sterile, stainless steel pot using a *Retsch* PM200 planetary ball mill at 400 rpm for five min or until completely powdered. Powder was transferred to clean 2.0-ml *Eppendorf* tubes in batches of 200 mg and stored in a -20°C freezer until extraction. Stainless steel pots were scrubbed in a solution of 10% bleach, followed by rinsing with Ultrapure water, DNAErase (*Sigma-Aldrich*) treatment for five min, wiping dry with 70% ethanol and UV irradiation for a minimum of 1 hr. Prior experiments have shown that this treatment is sufficient to avoid any cross contamination between samples powdered in the same pot at different times (data not shown).

Filter pipette tips, and certified DNase/RNase-free tubes and solutions were used throughout all DNA procedures. One DNA-free control was included for extraction alongside every 10 samples. DNA was extracted as per Dabney et al. (2013) with minor changes. Digest buffer contained final concentrations of approximately 0.25 mg/ml Proteinase K (*Astral*) and 0.45 M EDTA (*Sigma-Aldrich*) (i.e., 0.84 mg Proteinase K in 3.32 ml of 0.5 M EDTA is required per reaction). Digest buffer was inverted to mix and 1.66 ml was added to 200 mg of eggshell powder per sample in a 2-ml Safe-Lock *Eppendorf* tube. The tube was sealed tightly with Parafilm M (*Bemis NA*) and incubated for 30 min with rotation (10 rpm) at 55°C in a hybridisation oven. After 30 min, samples were centrifuged for 10 min at maximum speed in a bench-top centrifuge (*Eppendorf*) to pelletise debris. The supernatant was removed and

**TABLE S6.8.1 | EGGSHELL SPECIMENS PRIORITISED FOR ADNA EXTRACTION.**  
Asterisked samples did not yield enough useable aDNA and were excluded from subsequent analyses aDNA analyses.

Field ID	AD#	MB#	Location	Co-ordinates	Thickness (mm)	Amino acid racemisation	Estimated age (ya)
M06-M028	1666	2173 2616 2771	Talaky	-25.4, 45.7	4.12	0.0450	1330+/-20
M06-M025	1665	2618 2772 2778	Talaky	-25.2, 45.7	3.7	0.0455	1305+/-15
M06-M164	2121	2968*	Itampolo	-24.7, 43.9	1.9	0.0955	NA
M07-M041	2110	2973	Coast S. Ambovombe	-25.2, 46.3	0.9	0.0530	NA
M06-M112	2135	2986	Faux Cap	-25.6, 45.5	2.97	0.0615	NA
M07-M167	2137	2987	Isle E. Ambolobozokely	-12.5, 49.6	2.2	0.0700	NA
M06-M089	2125	2988	Cap St. Marie	-25.6, 45.1	1.14	0.0600	NA
M07-M199	1293	2997	Isle E. Ambolobozokely	-12.4, 49.6	1.94	0.0605	NA
MAD97-26	276	3004	Talaky	-25.4, 45.7	1<x<3	NA	NA
M07-M218	1288	3011	E. Ramena Beach	-12.3, 49.4	1.75	0.0610	NA
M06-M021	1664	3026	Talaky	-25.4, 45.7	1.38	0.0450	1495+/-15
M07-M161	2107	2962*	Isle E. Ambolobozokely	-12.4, 49.6	0.6	0.0540	NA
M06-M164	2112	2975*	Itampolo	-24.7, 43.9	1.51	0.0955	NA
M06-M229	2113	2980	Anakao	-23.7, 43.6	1.4	0.0690	NA
M06-M164	2118	2982*	Itampolo	-24.7, 43.9	3.15	0.0645	NA
M06-M292	2120	2984*	Ifaty	-23.1, 43.6	3.5	0.1985	NA
M06-M230	2127	2990	Anakao	-23.7, 43.6	1.2	0.0630	NA
M07-M217	2139	2998	E. Ramena Beach	-12.3, 49.4	2.69	0.0905	NA
CHANTAL2	1662	3024	Segeny	-23.5, 43.8	2.16	0.0350	1855+/-15
CHANTAL3	1663	3025	Segeny	-23.5, 43.8	2.7	0.0360	1785+/-15
M06-M172	1669	3219	Itampolo	-24.7, 43.9	4.04	0.0710	2000+/-500
M06-M194	1670	3220	Itampolo	-24.7, 43.9	3.74	0.0590	1500+/-500
M06-M296	1671	3221	Ifaty	-23.0, 43.5	1.0	0.0760	2000+/-500
TONY2 A1-1-3	1739	3222	Lakatom-Bato	-22.1, 43.2	4.01	NA	2004+/-37
NA	2384	3224*	Ampasilava	NA	3.46	NA	NA
NA	2394	3243*	Ankaranduka	-22.1, 43.2	2.92	NA	NA
NA	2402	3251*	Ankaranduka	-22.1, 43.2	1.06	NA	NA
NA	2406	3255*	Ampasilava	NA	2.95	NA	NA
NA	2409	3258	S. Salary	-22.6, 43.3	3.5	NA	NA
NA	2416	3260*	Ampasilava	NA	4.11	NA	NA
NA	2418	3262*	Andavadoaka	-22.0, 43.2	2.47	NA	NA
NA	2419	3263*	Andavadoaka	-22.0, 43.2	1.75	NA	NA
NA	2420	3264*	Andavadoaka	-22.0, 43.2	0.95	NA	NA

discarded. During optimisation of the extraction protocol, this pre-digestion step was found to reduce proportion of bacterial DNA relative to endogenous DNA without significantly decreasing endogenous DNA yield (data available upon request). A further 1.66 ml of digest buffer was added to the debris pellet, which was resuspended by vortexing and sealed with Parafilm. Samples were left to digest for a further 20 hr as above.

Samples were removed from the hybridisation oven and were centrifuged for 10 min at maximum speed to pelletise debris. Binding buffer was prepared with final concentrations of: 40% isopropanol, 0.05% Tween-20, 90 mM sodium acetate pH 5.2, and 5 M guanidine hydrochloride, in ultrapure water (i.e., 8.67 ml of 100% isopropanol, 10.83 µl of 100% Tween-20, 650 µl of 3 M sodium acetate pH 5.2, and 10.34 g of guanidine hydrochloride (MW 95.53 g/mol), topped up to 21.66 ml with ultra-pure water is required per sample). 1 µl of *QIAGEN* pH indicator (cat: 28004) per ml was added to the buffer to ensure that the pH remained within the DNA-binding range of the silica column. 1.66 ml of the digest supernatant was added to 21.66 ml of binding buffer in a 50 ml Falcon tube, and the solution was inverted to mix. A bleached and UV irradiated Zymo-Spin-V 15 ml extension reservoir (*Zymo Research*, cat: C1016-25) was gently but firmly fitted onto a MinElute silica spin column (*QIAGEN*, cat: 28004). Parafilm was wound around the junction between the extension reservoir and spin column to create a filter assembly that was placed inside a 50 ml Falcon tube. 10 ml of the digest-binding buffer solution was poured into the extension reservoir, and the cap was sealed with Parafilm. This assembly was centrifuged at 1,500 x g in a balanced bench-top centrifuge with a soft ramp for four minutes, after which the flow-through was discarded in a sealed, disposable waste container, and the remainder of the digest-binding buffer solution was poured into the extension reservoir and sealed. The assembly was centrifuged as before. The flow-through was discarded and the extension reservoir and spin column were gently disassembled. The spin column was placed in a provided 2 ml collection tube (*QIAGEN*, cat: 28004), and centrifuged for one min at 13,000 rpm. Flow-through was discarded, and 700 µl of Buffer PE (*QIAGEN*, cat: 28004) was added to the spin column, and the column was centrifuged for one min at 13,000 rpm. This step was repeated for a second time using a clean 2 ml collection tube. After discarding the

flow-through, the spin column was placed in a clean 1.5 ml *Eppendorf* tube and dry spun for one min at 13,000 rpm. To elute the DNA, 15 µl of EB Buffer (*QIAGEN*, cat: 28004) warmed to 37°C was added directly to the membrane, the column was incubated for five min at 37°C, and then centrifuged at 10,000 rpm for one min. This elution step was repeated such that 30 µl of eluate was obtained. The 30 µl of eluate was passed back through the column one more time after a final five min incubation at 37°C, and transferred to a clean 0.5 ml Lo-Bind Safe-lock *Eppendorf* tube. 1 µl of 1% TE-Tween-20 was added per 20 µl of eluate (i.e., 1.5 µl 1% TE-Tween-20) to ensure DNA did not adhere to the tube. DNA was stored at -20°C until further use.

#### **S6.8.4 PREPARATION of SHOTGUN SEQUENCING LIBRARY**

Shotgun sequencing libraries were prepared by following (Gansauge and Meyer 2013) with minor changes in an ultra-clean environment. Modifications to the adapters used are listed in Table S6.8.2.

An extraction control, no-template (water) control, and CL104 positive control were also included in the library building process. At step 1, reactions were performed in 0.2 ml 8-well PCR strip tubes. 12 µl of DNA extract was used, and Afu UDG was replaced by ultrapure water. We elected to not use uracil-DNA glycosylase because “many aDNA researchers are reassured of the authenticity of the resulting ancient sequences when random C-T transitions are observed in cloned products of a PCR, as this form of damage is common in ancient samples” (Shapiro and Hofreiter 2012). At step 5, a final concentration of 2.5 U/µl CircLigase II was used (i.e., 2 µl of 100 U/µl CircLigase II). At step 13, tubes were incubated in a thermal cycler for two min at 65°C as opposed to a thermal shaker. At step 13, tubes were transferred to a thermal shaker (*Eppendorf*) pre-cooled to 15°C as opposed to a thermocycler. Steps 14, 15, 18, 19, 23, and 25 were performed in a thermal shaker. At step 25, the supernatant was stored in a 1.5 ml Lo-Bind *Eppendorf* tube at -20°C. After step 28 (in a physically separated post-PCR laboratory), 10 µl of PCR product was combined with 0.5 µl of 6X loading dye (*QIAGEN*) and run alongside 3 µl of 50 bp DNA ladder (GeneRuler, *Fermentas*) for 40 min at 96 V on a 2% agarose gel electrophoresis (2.2 g agarose, 110 ml 1X TAE buffer, 8 µl GelRed, 1x20-well comb) that was visualised and photographed using a *BioRad* transilluminator,

**TABLE S6.8.2** | SINGLE-STRANDED LIBRARY BUILD PRIMERS used that have been modified from Gansauge and Meyer (2013).

Name	Sequence (5'-3')
CL53	ACACGACGCTCTC/3ddC/
CL78	/5Phos/AGATCGGAAG/iSp9//iSp9//iSp9//3BioTEG/

in order to confirm the library building process.

At step 30 (Gansauge and Meyer 2013), the libraries were amplified in eight replicate qPCR reactions with unique fusion-tag indexing primers suitable for the *Illumina* sequencing platforms. The PCR reaction contained reagents in final concentrations of: 1X *ABI* Power SYBR Master Mix (12.5  $\mu$ l of a 2X stock), 0.4  $\mu$ M *IDT* forward primer (1  $\mu$ l of a 10  $\mu$ M stock); 0.4  $\mu$ M *IDT* reverse primer (1  $\mu$ l of a 10  $\mu$ M stock); 5  $\mu$ l of neat library, made up to a total of 25  $\mu$ l final volume with HPLC-grade water (9.5  $\mu$ l). Thermal cycling conditions were: 95°C for two min, 20 cycles of 95°C for 15 sec, 60°C for 30 sec, 68°C for one min. Replicate reactions were combined and vortexed to mix.

### **S6.8.5 HYBRIDISATION ENRICHMENT of MITOCHONDRIAL DNA**

In a 0.2-ml Lo-bind PCR tube (round capped) in an ultra-clean environment, a solution was prepared (per reaction) containing final concentrations of 0.45  $\mu$ g/ $\mu$ l Chicken Cot-1 (2.5  $\mu$ l of a 1  $\mu$ g/ $\mu$ l stock), 9.1  $\mu$ M *IDT* forward blocking primer (Table S6.8.3; 0.25  $\mu$ l of a 200  $\mu$ M stock), 9.1  $\mu$ M *IDT* reverse blocking primer (Table S6.8.3; 0.25  $\mu$ l of a 200  $\mu$ M stock). In a separate 0.2-ml tube, a solution was prepared (per reaction) containing final concentrations of 9X Hyb#1 (i.e., 9  $\mu$ l of a 20X stock of Hyb#1 which is 20X SSPE), 0.0125 M Hyb#2 (i.e., 0.5  $\mu$ l of a 0.5 M stock of Hyb#2 which is 0.5M EDTA, pH 8.0), 8.75X Hyb#3 (i.e., 3.5  $\mu$ l of a 50X stock of Hyb#3 which is Denhardt's solution), 0.25% Hyb#4 (i.e., 0.5  $\mu$ l of a 1% stock of Hyb#4, which is 10% SDS), 1 U/ $\mu$ l Rnase Block (*SUPERase*; i.e., 1  $\mu$ l of a 20 U/ $\mu$ l solution of Rnase Block), and 5.5  $\mu$ l of baits for a final volume of 20  $\mu$ l. Solutions were gently vortexed to mix and briefly spun in a bench-top microcentrifuge to collect the solution. In a post-PCR environment, 5  $\mu$ l of the first solution was then added to 7  $\mu$ l of library (in a post-PCR environment) for a total of 12  $\mu$ l.

The first solution was placed in a thermalcycler and incubated for five min at 95°C. The second solution was then placed in the thermalcycler and both solutions were incubated for five min at 55°C, after which 18  $\mu$ l of the second (baits) solution was

**TABLE S6.8.3** | CUSTOM BLOCKING PRIMERS (for sequencing the NextSeq platform) used in the enrichment protocol and primers used to amplify the captured libraries.

Name	Sequence (5'-3')	Function
NextSeq_Blocking_F	AATGATACGGCGACCACCGAGATCTACACIIIIIII ACACTCTTCCCTACACGACGCTCTT /3InvdT/	Forward capture blocking primer
NextSeq_Blocking_R	CAAGCAGAAGACGGCATAACGAGATIIIIIII GTGACTGGAGTTCAGACGTGTGCTCT /3InvdT/	Reverse capture blocking primer
P5	AATGATACGGCGACCACCGAGATCTACAC	Forward library amplification primer ( <i>Illumina</i> )
P7	CAAGCAGAAGACGGCATAACGAGAT	Reverse library amplification primer ( <i>Illumina</i> )

added to the first (library) solution. The solution was gently vortexed to mix. 10  $\mu\text{l}$  of mineral oil was added on top of the reaction to prevent evaporation, and briefly spun in a microcentrifuge to collect liquid. The solution was incubated in the thermalcycler with a heated lid for 40 ht at 55°C. The remainder of the MYbaits protocol v.3 (*MYcroarray*) was performed as per the manufacturer's instructions between steps 2A.1 and 3.1.

Captured libraries were amplified in eight replicate qPCR reactions containing final concentrations of: 1X *ABI* Power SYBR Master Mix (25  $\mu\text{l}$  of a 2X stock), 0.2  $\mu\text{M}$  *IDT* forward primer P5 (1  $\mu\text{l}$  of a 10  $\mu\text{M}$  stock; Table S6.8.3); 0.2  $\mu\text{M}$  *IDT* reverse primer (1  $\mu\text{l}$  of a 10  $\mu\text{M}$  stock; Table S6.8.3); 5  $\mu\text{l}$  of neat library, made up to a total of 50  $\mu\text{l}$  final volume with HPLC-grade water (18.5  $\mu\text{l}$ ). Thermal cycling conditions were: 95°C for 30 seconds, 20 cycles of 95°C for 20 sec, 60°C for 30 sec, 72°C for 30 sec. Replicate reactions were combined and vortexed to mix.

#### **S6.8.6 QUANTITATION of THE CAPTURED LIBRARIES and POOLING**

The total DNA concentration of an aliquot of 1  $\mu\text{l}$  of each captured library was quantified using a Nanodrop 2000 spectrophotometer following the manufacturer's instructions. Based on this rough quantitation, each library was diluted to between 2 and 5 ng/ $\mu\text{l}$  in ultrapure water for a final volume of 10  $\mu\text{l}$ , in order for their concentrations to be within the dynamic range of the LabChip GX Touch HT (*Perkin Elmer*); this usually required a 1 in 10 dilution (i.e., 1  $\mu\text{l}$  of sample in 9  $\mu\text{l}$  of ultrapure water). 10  $\mu\text{l}$  of this dilution was loaded onto the LabChip GX Touch HT for quantitation of fragments between 140 and 300 bp, following the manufacturer's instructions. Libraries were then pooled in equimolar concentrations in a total volume of 60  $\mu\text{l}$  and vortexed to mix.

#### **S6.8.7 SIZE SELECTION and PURIFICATION of POOLED SEQUENCING LIBRARY**

To eliminate low-molecular weight artifacts, the 60  $\mu\text{l}$ -pooled library was divided into two lots of 30  $\mu\text{l}$  and run on two lanes of a Pippin Prep (*Sage Science*) ethidium bromide eGel cassette to select fragments between 140 and 300 bp, following the manufacturer's instructions. This size range was selected to capture the largest

fragments representing elephant bird sequences, with the remainder of fragments below 140 bp representing sequencing adapters or dimer. To buffer exchange and concentrate the size-selected library, it was purified using a PCR purification kit (*QIAGEN*). After the addition of PB buffer, 3 µl of 3 M sodium acetate was added to adjust the pH. DNA was eluted in 30 µl of ultrapure water.

#### **S6.8.8 FINAL QUANTITATION *and* SEQUENCING**

The final purified sequencing library was quantitated using a Nanodrop 2000 spectrophotometer and LabChip GX Touch HT as above.

Captured libraries were sequenced on the NextSeq next-generation sequencing platform (*Illumina*; Curtin University). 10.0 µl of the 2 nM dilution of the library was combined with 10 µl of 0.2 M molecular biology-grade NaOH and incubated for five min at 25°C, then placed on ice. 10 µl of 200 mM Tris-HCl was added to this mixture, and the solution was vortexed to mix and placed on ice. 970 µl of HT1 buffer was added to this mixture, vortexed, and placed on ice to make 1 ml of a 20 pM library. 97.5 µl of the 20 pM library was added to 1202.5 µl of HT1 buffer to make 1.3 ml of a 1.5 pM library. 1.2 µl of 20 pM PhiX was added to 1299 µl of the 1.5 pM library to make the final loading solution that was added to well 17 of the reagent cartridge. 6 µl of a 100 µM stock of custom sequencing primer CL72 (Table S6.8.4) was spiked into well 20 of the reagent cartridge, and a 6 µl of a 100 µM stock of custom i5 indexing primer (Table S6.8.4) was spiked into well 22 of the reagent cartridge.

The options selected when creating the sample sheet were: Other/ FASTQ only/ Sample preparation kit = TruSeq LT/ 2 Index reads/ Single end/ 150 cycles/ No custom primer/ No trimming. The library was run using a NextSeq 500 Mid-output 150 v.2 kit.

#### **S6.8.9 QUALITY CONTROL *and* FILTERING**

Sequencing data for each file was quality filtered in USEARCH v.8 (Edgar 2010) using the following command:

**TABLE S6.8.4** | CUSTOM I5 INDEXING PRIMER AND READ 1 SEQUENCING PRIMERS USED TO SEQUENCE LIBRARIES ON THE NEXTSEQ PLATFORM.

Name	Sequence (5'-3')	Function
Custom i5 indexing primer	GGAAGAGCGTCGTGTAGGGAAAGAGTGT	Custom forward index sequencing primer
CL72	ACACTCTTCCCTACACGACGCTCTTCC	Custom Read 1 sequencing primer

```
usearch8 -fastq_filter [FILENAME.fastq] -fastqout  
[FILENAME_QF.fastq] -fastq_truncqual 3 -fastq_maxee_rate  
0.01 -fastq_minlen 30
```

Quality filtered sequence data from each sample across the four NextSeq lanes were combined using the following command:

```
cat [SAMPLE_NAME]* > [SAMPLE_NAME_QF_.CAT].fastq
```

Sequences were dereplicated using the following command:

```
usearch8 -derep_fulllength [SAMPLE_NAME_QF_CAT].fastq -  
fastqout [SAMPLE_NAME_QF_CAT_DEREP].fastq -sizeout
```

Chimeric sequences were removed using the following command:

```
usearch8 -uchime_denovo [SAMPLE_NAME_QF_CAT_DEREP].fastq  
-nonchimerasq [SAMPLE_NAME_QF_CAT_DEREP_CF].fastq
```

Filtered sequences were imported into Geneious v.8.1.6 (Kearse et al. 2012) and any remaining adapter sequences (that were not automatically removed by *Illumina* software due to sequencing errors) were trimmed from the 3' end; these were found using the 'Annotate and Predict' / Trim Ends / Remove new trimmed regions / Allow mismatches 2 / Minimum match length 9 / Trim 3' end. Sequences below 30 bp were discarded and unique sequences were found through Edit / Find Duplicates / Extract unique sequences.

**TABLE S6.8.5** | SUMMARY TABLE OF THE READS LOST DURING QUALITY CONTROL, THE NUMBER OF READS MAPPED AND THEIR STATISTICS, AND THE AVERAGE COVERAGE OF THE MITOCHONDRIAL GENOME ACROSS THE NUMBER OF BASE-PAIRS RETRIEVED FOR EACH SAMPLE.

### **S6.8.10 RECONSTRUCTION *of* MITOCHONDRIAL GENOMES**

**TAXONOMIC ASSIGNMENT OF SEQUENCES.** LCA parameters in MEGAN v.4.70.4 (Huson et al. 2007) were: Min Support 1 / Min Score 35.0 / Top Percent 10.0 / Win Score 0.0 / Min Complexity 0.44. The tree was collapsed at the order level, and sequences within the node Aves were extracted in .fasta format.

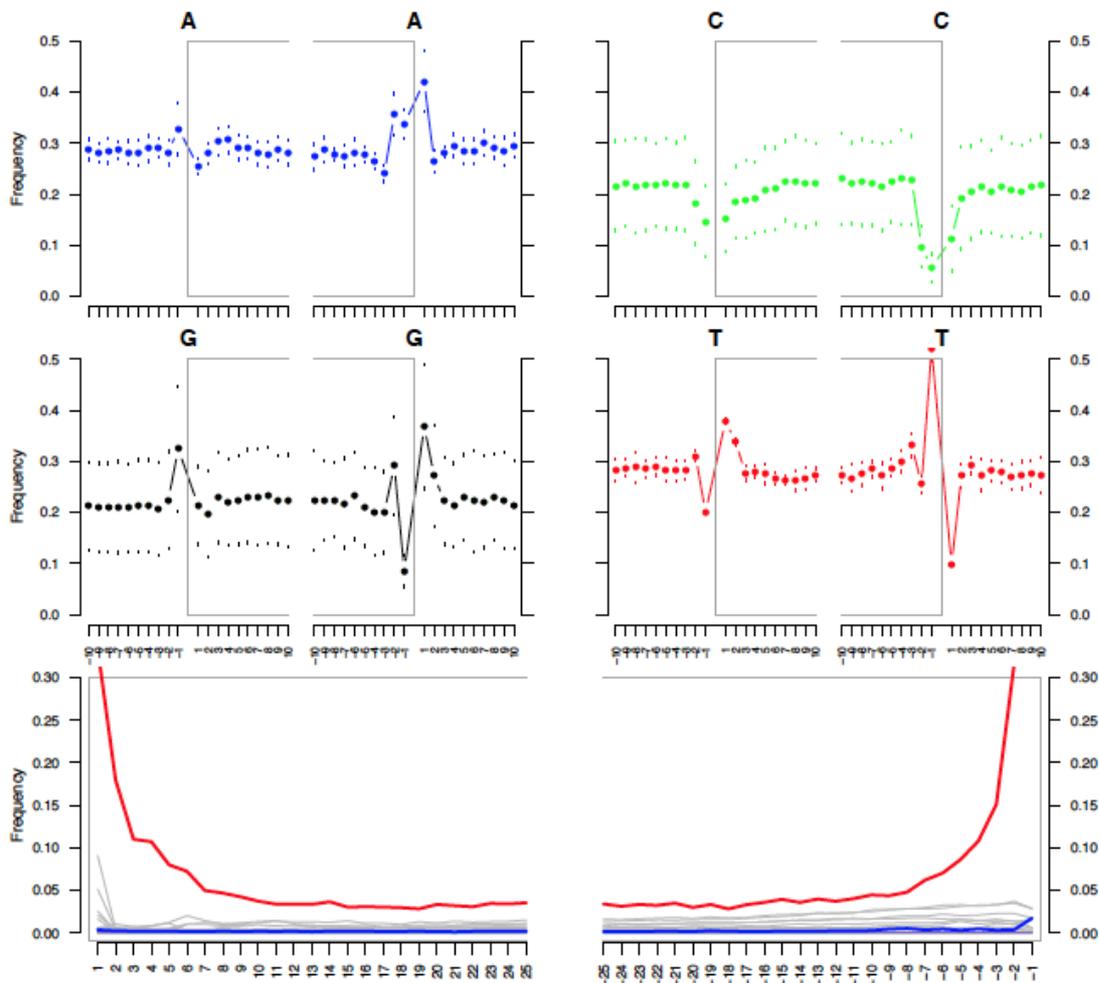
**AUTHENTICITY.** Final mapping files were exported from Geneious as a .sam file into /mapDamage/bin, also exporting the reference sequence in .fasta format. MapDamage v.2.0 was run by navigating to the /mapDamage/bin folder in a terminal window, and typing the command:

```
mapdamage -i [INPUT].sam -r [REFERENCE].fasta
```

Nucleotide misincorporation plots were examined for a higher proportion of C to T misincorporations at the 5' and 3' terminus of reads that is indicative of authentic ancient DNA. In the first position from the 5' end, the frequency of C to T mutations exceeded 0.3 in all samples (Figure S6.8.1 shows a typical example), suggesting that the mapped reads were truly of ancient origin.

### **S6.8.11 GENE ANNOTATION *and* PARTITIONING**

20 elephant bird mitochondrial genomes were annotated by aligning them with a kiwi mitochondrial genome, and transferring the annotations from the kiwi to the elephant bird genomes in Geneious v.8.1.6 (Kearse et al. 2012) using the 'Annotations→Copy all in selected region to' function. Annotated genomes were then aligned with eight outgroup ratite taxa (Table S6.8.6) and two previously published elephant bird genomes (*Aepyornis hildebrandti* KJ749824 and *Mullerornis agilis* KJ749825; Mitchell et al. 2014) using a MAFFT alignment (Katoh et al. 2002) as implemented in Geneious v.8.1.6 ('Align/Assemble→Multiple align→MAFFT alignment') with the default parameters. The alignment was refined by using the MUSCLE algorithm (Edgar 2004) as implemented in Geneious v.8.1.6 with the default parameters ('Align/Assemble→Multiple align→MUSCLE alignment'). As



**FIGURE S6.8.1** | DAMAGE PROFILE from MB2973 exemplifying a higher proportion of C to T nucleotide misincorporations (red) at both the 5' and 3' ends mapped reads (bottom panel). This damage pattern was observed for all that were used for subsequent analysis. Figure generated by MapDamage v.2.0 (Ginolhac et al. 2011; Jonsson et al. 2013).

aDNA damage can inflate genetic diversity, potentially damaged bases were identified as any G->A or C->T transition that is present in only one sample but where the 'wild-type' base is otherwise fixed in all ratites. These positions were RY-coded. The alignment contained a total of 30 complete mitochondrial genomes. All protein-coding genes (*ATP6*, *ATP8*, *CO1*, *CO2*, *CO3*, *CYTB*, *ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5*, and *ND6*), rRNA genes (*16SrRNA* and *12SrRNA*), tRNA genes (22), and the control region were individually extracted from the alignment using the 'Extract' function in Geneious v.8.1.6.

Protein coding genes were translated to ensure the ORFs were in-frame. Any termination mutations (i.e., base changes that convert an amino acid to a stop codon) were converted to RY-coding. Insertion or deletion mutations were dealt with by adding additional gaps such that the codons down stream of the mutation were still in-frame. Each protein-coding gene was partitioned by codon position in Geneious v.8.1.6 using the 'Tools->Strip alignment columns->Strip two adjacent columns per codon->Start with column 2' function (n.b.: starting with column 2 will extract 1<sup>st</sup> codon positions, starting with column 3 will extract 2<sup>nd</sup> codon positions, and starting with column 1 will extract 3<sup>rd</sup> codon positions, provided that every position in the alignment is in-frame). Codon positions for each gene were concatenated in Geneious v.8.1.6 using the 'Tools->Concatenate sequences or alignments' function, such that three files were generated, one containing the 1<sup>st</sup> codon positions for all protein-coding genes, one containing the 2<sup>nd</sup> codon positions for all protein-coding genes, and one containing the 3<sup>rd</sup> codon positions for all protein-coding genes.

The stems and loops of the *16srRNA* gene were annotated as defined by (do Amaral et al. 2010), and were partitioned in Geneious v.8.1.6 (using the 'Extract' function, as above). The stems and loops of the *12srRNA* gene were annotated as defined by (De Los Monteros 2003), and were partitioned in Geneious v.8.1.6 as above. The secondary structure of tRNAs were predicted using the program ARWEN (Laslett 2008), and the stems (acceptor stem, D stem, anticodon stem, and T stem) and loops (D loop, anticodon loop, variable loop, and T loop) were partitioned as above. Loop alignments of the RNA genes were concatenated as above, as were the stem alignments of the RNA genes.

**TABLE S6.8.6** | GENBANK ACCESSIONS OF THE EIGHT OUTGROUP RATITE TAXA USED FOR WHOLE MITOCHONDRIAL GENOME PHYLOGENETIC ANALYSES.

Species name	Common name	Accession Number	Reference
<i>Apteryx australis mantelli</i>	North island brown kiwi	AY016010	(Cooper et al. 2001)
<i>Apteryx australis mantelli</i>	North island brown kiwi	KU695537	(Liu J 2016)
<i>Apteryx haastii</i>	Great spotted kiwi	NC_002782	(Haddrath and Baker 2001)
<i>Apteryx owenii</i>	Little spotted kiwi	NC013806	(Phillips et al. 2010)
<i>Casuarius bennettii</i>	Dwarf cassowary	AY016011	(Cooper et al. 2001)
<i>Casuarius casuarius</i>	Southern cassowary	NC_002778	(Haddrath and Baker 2001)
<i>Dromaius novaehollandiae</i>	Emu	AY016014	(Cooper et al. 2001)
<i>Dromaius novaehollandiae</i>	Emu	NC_002784	(Haddrath and Baker 2001)

The control region for all taxa was annotated and partitioned into the three domains as defined in (Ruokonen et al. 2002) in Geneious v.8.1.6 using the ‘Add Annotation’ and ‘Extract Annotation’ functions. However, the control region was not used for downstream analyses.

For downstream analyses, all unknown characters marked by a ‘?’ were designated as gaps by ‘-’.

#### **S6.8.12 RCV ANALYSIS**

Relative composition variability (RCV; Phillips et al. 2010) for each mitochondrial partition with standard nucleotide coding as well as RY coding was calculated in PAUP v.4a150 (Swofford 2003), in order to determine which partitions may benefit from RY-coding. Partition MSAs were combined in NEXUS format, with each partition defined as a character set (‘charset’):

```
charset m1 = 1-3806;  
charset m2 = 3807-7609;  
charset m3 = 7610-11412;  
charset stems = 11413-13388;  
charset loops = 13389-15554;
```

An RY-coded MSA (R = A or G; Y = C or T) in NEXUS format was also generated with partitions defined by the character sets:

```
charset m1ry = 1-3806;  
charset m2ry = 3807-7609;  
charset m3ry = 7610-11412;  
charset stemsry = 11413-13388;  
charset loopsry = 13389-15554;
```

The MSA was imported into PAUP v.4a150 (Swofford 2003). Three estimates of base frequency for each partition were generated in which: (1) constant sites and gaps were excluded; (2) only gaps were excluded; and (3) no sites were excluded.

```
> execute [filename].nex
> exclude all; include [partition name]; exclude
constant gapped; basefreqs
> exclude all; include [partition name]; exclude
gapped; basefreqs
> exclude all; include [partition name]; basefreqs
```

For each partition, both standard- and RY-coded, base frequencies were copied into *Microsoft's* Excel in order to calculate RCV, which is given by the sum of the absolute deviation of each base frequency for each taxon from the average across all taxa, divided by the number of taxa (Phillips et al. 2010). Chi-squared tests for significant base composition bias were also generated and are recorded in Table S6.8.7 alongside RCV estimates.

$\chi^2$  P-values below 0.05 indicate significant base composition bias for that partition, but the power of this test is highly influenced by the number of sites included (Phillips et al. 2010). A reduction in RCV estimates with RY-coding indicates that base composition bias is somewhat mitigated by RY-coding, and this test gives a better indication of base composition bias.

Regardless of whether constant sites and gaps are excluded, or just gaps, or no sites,  $\chi^2$  *p*-values are above 0.05 for all partitions indicating insignificant base composition bias. RY-coding each partition does not decrease the  $\chi^2$  *p*-value. Likewise, RCV does not increase by RY-coding any partition (Table S6.8.7). These results suggest that in order to resolve the relationships among elephant birds no RY coding of any partitions is necessary.

**TABLE S6.8.7** | RCV CALCULATIONS AND CHI-SQUARED TEST ESTIMATES for both nucleotide and RY-coded mitochondrial partitions when constant sites and gaps are excluded, when only gaps are excluded, and when no sites are excluded.

Partition	Exclude constant gapped				Exclude gapped				Exclude none			
	NT		RY		NT		RY		NT		RY	
	RCV	$\chi^2$ <i>p</i> -value	RCV	$\chi^2$ <i>p</i> -value	RCV	$\chi^2$ <i>P</i> -value	RCV	$\chi^2$ <i>p</i> -value	RCV	$\chi^2$ <i>p</i> -value	RCV	$\chi^2$ <i>p</i> -value
m1	0.05	0.87	0.07	0.90	0.01	1.00	0.00	1.00	0.02	1.00	0.00	1.00
m2	0.05	1.00	0.07	1.00	0.01	1.00	0.00	1.00	0.01	1.00	0.00	1.00
m3	0.03	0.73	0.05	0.78	0.02	0.97	0.01	1.00	0.02	0.04	0.01	1.00
stem	0.08	1.00	0.03	1.00	0.02	1.00	0.00	1.00	0.02	1.00	0.01	1.00
loop	0.08	1.00	0.13	0.95	0.02	1.00	0.02	1.00	0.03	1.00	0.01	1.00
all	0.02	0.07	0.04	0.81	0.01	1.00	0.00	1.00	0.01	0.00	0.01	0.03

### S6.8.13 STEMMINESS

“Stemminess is a measure of the amount of phylogenetic signal erosion, and can be used alongside RCV to determine whether RY-coding certain partitions will mitigate biases in the data (i.e., “compositional heterogeneity”) that may lead to incorrect phylogenetic inference (Phillips 2008; Phillips et al. 2010)”. In order to calculate uncorrected stemminess, first a null hypothesis tree was generated in PAUP v.4a150.

```
> execute [filename].nex
> hsearch
> contree
> out DRONO
> showtree
```

The generated tree was converted to Newick format and was added to the NEXUS files above in a TREES block:

```
begin trees;
tree nulltree =
(DRONO, (DRONO2, ((CASBE, CASCA), ((APTOW, APTHA), (APTMA, APTM
A2))), ((2988, ((2973, (2990, 2980)), (3004, (MULAG, 3221)))), ((A
EPHI, (2987, (2997, (2998, 3011)))), ((3025, 3024), (1666, (2986,
(3026, (3220, ((1665, 3219), (3222, 3258))))))))))));
```

Branch lengths were then generated for each partition, both standard- and RY-coded, with: (1) constant sites and gaps were excluded; (2) only gaps were excluded; and (3) no sites were excluded.

```
> execute [filename].nex
> DerootTrees
> out DRONO
> set crit = dist
> dset distance = abs negbrlen = prohibit
```

```

> exclude all; include [partition name]; exclude
constant gapped
> describetrees 1/plot = phylogram brlens = yes

```

Branch lengths were imported into *Microsoft's* Excel in order to calculate uncorrected stemminess per partition, which is defined as the sum of the internal branch lengths of a tree divided by the sum of internal and external branch lengths. Results are recorded in Table S6.8.8.

In order to calculate whether a model will correct for substitution saturation, stemminess was calculated for both standard- and RY-coded partitions after Bayesian tree generation. Standard- and RY-coded partitions were combined into a NEXUS file that was followed by a MrBayes block:

```

Begin MrBayes;

charset m1 = 1-3806;
charset m2 = 3807-7609;
charset m3 = 7610-11412;
charset stems = 11413-13388;
charset loops = 13389-15554;
charset m1ry = 15554-19360;
charset m2ry = 19361-23163;
charset m3ry = 23164-26966;
charset stemsry = 26967-28942;
charset loopsry = 28943-31108;

partition bycodon = 10: m1, m2, m3, stems, loops,
m1ry, m2ry, m3ry, stemsry, loopsry;
set partition=bycodon;
lset applyto=(1-5) nst=6 rates=invgamma;
lset applyto=(6-10) nst=1 rates=invgamma;
outgroup STRCA;

prset applyto=(all) statefreqpr = fixed(empirical);

```

```

unlink statefreq=(all);
unlink revmat=(all);
unlink shape=(all);
unlink pinvar=(all);
unlink brlens=(all);
mcmc ngen=10000000 printfreq=5000 samplefreq=5000
nrns=2 nchains=3 temp=0.1 diagnfreq=50000
savebrlens=yes filename=[filename];
SUMT filename=[filename] relburnin=yes burnin=0.2
contype=allcompat;
SUMP filename=[filename] relburnin=yes burnin=0.2;
end;

```

MrBayes was implemented through the online bioinformatics platform CIPRES (Miller et al. 2010). The output files from MrBayes (.t) were modified such that they could be imported into TreeStat v.1.2 (Rambaut and Drummond 2007). The ‘Treeness’ values generated for each partition and each run were imported into *Microsoft’s* Excel to calculate the average corrected stemminess for each partition (with 95% confidence; Table S6.8.9).

Higher values of stemminess indicate that there is less saturation of phylogenetic signal, which can mitigate some of the base composition bias observed. Stemminess is not increased by RY coding any partitions (Table S6.8.8); in fact, uncorrected stemminess is *decreased* by RY coding, indicating that phylogenetic signal is lost by RY coding. Note that uncorrected stemminess could not be calculated when partitions are RY coded *and* constant sites and/or gaps are excluded as the branch lengths became 0. This result further supports the notion that significant phylogenetic signal is lost by RY coding, suggesting that no partitions should be RY coded. Corrected stemminess is not significantly increased by RY coding any partitions, (Table S6.8.9), suggesting that a model alone may correct for any substitution saturation.

The RCV and stemminess results indicate that RY coding of any mitochondrial positions may result in a significant loss of phylogenetic signal. Nevertheless,

**TABLE S6.8.8** | UNCORRECTED STEMMINESS ESTIMATES for both nucleotide and RY-coded mitochondrial partitions when constant sites and gaps are excluded, when only gaps are excluded, and when no sites are excluded.

Partition	Exclude constant gapped		Exclude gapped		Exclude none	
	NT	RY	NT	RY	NT	RY
	Stemminess	Stemminess	Stemminess	Stemminess	Stemminess	Stemminess
m1	0.64	NA	0.64	NA	0.62	0.03
m2	0.53	NA	0.53	NA	0.53	0.02
m3	0.73	NA	0.73	NA	0.66	0.03
stem	0.60	NA	0.60	NA	0.29	0.03
loop	0.58	NA	0.58	NA	0.28	0.02
all	0.70	NA	0.70	NA	0.58	0.02

**TABLE S6.8.9** | AVERAGE CORRECTED STEMMINESS ESTIMATES for both nucleotide and RY-coded mitochondrial partitions. Upper and lower bounds for a 95% confidence interval for the mean are also shown.

Partition	Stemminess			Stemminess-RY		
	Average	Upper	Lower	Average	Upper	Lower
m1	0.71	0.74	0.68	0.65	0.74	0.56
m2	0.55	0.60	0.49	0.50	0.62	0.38
m3	0.88	0.89	0.87	0.87	0.90	0.84
stems	0.51	0.57	0.45	0.44	0.56	0.32
loops	0.72	0.76	0.68	0.68	0.77	0.59

phylogenetic analyses were performed on two datasets, with: (1) no RY coding of any partitions; and (2) full RY-coding of all partitions.

#### **S6.8.14 MODEL TEST ANALYSIS**

To find the best substitution model for each partition, the standard-coded MSA was executed in PAUP v.4a150, followed by execution of a modelblock.nex file:

```
> execute [filename].nex  
> exclude all; include [partition name]  
> execute modelblock.nex
```

The output file for each partition (model.scores) was placed in the same folder as a jModelTest v.3.7 (Posada and Crandall 1998; Guindon et al. 2003; Darriba et al. 2012) executable, which was evoked through command prompt, specifying the number of characters in the partition (-n) and the number of taxa (-t):

```
> cd c:\[path to executable]  
> modeltest3.7win -n3806 -t39 < m1_model.scores >  
m1.modeltest
```

The best scoring models as determined through hierarchical likelihood ratio tests (hLRT) and corrected Akaike information criterion (AIC) are recorded in Table S6.8.10.

#### **S6.8.15 MAXIMUM LIKELIHOOD PHYLOGENY GENERATION**

An input .phy file where the RY coded partitions were modified to binary characters (i.e., R = 0, Y = 1, ambiguous characters = -) was run in RAxML (Stamatakis 2014) with the following parameters: Data type = DNA or mixed; Analysis / Multiple outgroups; Import partitions.txt (DNA for nt coded, BIN for ry coded); Auto mre = 500; BS brL; GTRGamma or BinGamma (full RY-coding). The output was a 'bipartitions.tre' file.

**TABLE S6.8.10** | MODELTEST RESULTS for each mitochondrial partition. MrBayes parameters used when partitions are standard or RY-coded are also indicated. A GTR+G model was used for all partitions in RAxML.

Partition	hLRT	AIC	Model used	NT		RY	
				nst	rates	nst	rates
m1	TrN+G	GTR+I	GTR+I	6	equal	1	equal
m2	TrN+I	GTR+I	GTR+I	6	equal	1	equal
m3	TrN+I+G	GTR+I+G	GTR+I+G	6	invgamma	1	invgamma
stems	K80+G	K80+I	HKY+I+G	2	invgamma	1	invgamma
loops	GTR+G	GTR+I+G	GTR+I+G	6	invgamma	1	invgamma

### **S6.8.16 BAYESIAN PHYLOGENY GENERATION**

MrBayes (Huelsenbeck and Ronquist 2001) was implemented via the online bioinformatic CIPRES science gateway (Miller et al. 2010). Bayesian inference analyses were run with unlinked substitution models and branch-length rate multipliers among the partitions (Phillips and Pratt 2008). Three MCMC chains for each of two independent runs proceeded for  $10^7$  generations with trees sampled every 2500 generations. The parameters specified in the input NEXUS file were as follows: Ntax = number of taxa; Nchar = number of characters; Format datatype = dna; Gap = -; Interleave = yes; Partition by codon = list of the partition names included; Set partition = bycodon; Lset applyto= [partition #]; nst=1, 2, or 6 (Table S6.8.10); rates=invgamma/gamma [determined from ModelTest]; Outgroup DRONO; Prset applyto = all; statefreqpr = fixed (empirical); Unlink statefreq=all; Unlink revmat=all; Unlink pinvar=all; Unlink brlens=all; Link brlens= [certain partitions]; Link brlens= [other partitions]; Prset applyto = [some partitions]; ratepr= variable; Prset applyto = [other partitions]; ratepr= variable; Mcmc ngen = 10000000; printfreq=5000; samplefreq=2500; nruns=2; nchains=3; temp=0.1; diagnfreq=5000; savebrlens=yes; filename=[filename]; SUMT filename= [filename]; relburnin=yes; burnin=0.2; SUMP filename= [filename]; relburnin=yes; burnin=0.2. The output was a 'con.tre' file.

The burn-in for each MrBayes run ensured that  $-\ln L$  had plateaued, clade frequencies had converged between runs (i.e., the average standard deviation of split frequencies was  $<0.02$ , approaching 0), and that PRSF (potential scale reduction factors, a convergence diagnostic) approached 1 (Phillips and Pratt 2008). Output files were then opened in Tracer v.1.6.1 (Rambaut et al. 2003) to ensure that estimated sample sizes (ESS) for substitution parameter estimates were above 200, and that parameter probability estimates had overlapping distributions, further confirming that runs had converged.

### **S6.8.17 MOLECULAR DATING**

Calibrations used for molecular dating are presented in Table S6.8.11. The input parameters for molecular dating analysis using MCMCtree (Yang 2006) were: seed =

-1, ndata = no. partitions (depending on the dataset), seqtype = 0, usedata = 3, RootAge = <72.3, model = 7, alpha = 1, ncatG =4, cleandata = 0, BDparas = 1 1 0, kappa\_gamma = 6 2, alpha\_gamma = 1 1, rgene+gamma = 1 2, sigma2\_gamma = 1 1, finetune = 1, print = 1, burnin = 12,000, sampfreq = 50, nsample = 1500. The output .BV file was renamed to in.BV and used as the input to rerun the analysis (i.e. usedata = 2). Analyses with both independent (clock = 2) and correlated (clock =3) rates were performed for the standard coded dataset using the phylogenetic tree topology in Figure 6.2. Each analysis was replicated.

### **S6.8.18 ASSESSMENT *of* GENETIC VARIATION *in* BARCODING GENES**

Specimens exhibiting less than 10% missing bases across a 596 bp barcoding region of *cytochrome oxidase I (COI)* were used to calculate the genetic distance within and between elephant birds from each region in Madagascar. An alignment of 13 sequences was imported into MEGA v.6.06 (Tamura et al. 2013). The within-group and between-group mean Kimura 2-parameter distance (Kimura 1980) was estimated, implementing the following options: variance estimation method = none, substitution type = nucleotide, model/method = Kimura 2-parameter model, substitutions to include = d: Transitions + Transversions, rates among sites = uniform rates, pattern among lineages = same (homogeneous), gaps/missing data treatment = pairwise deletion, select codon positions = 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, noncoding sites. To gauge the limits of intra- and inter-specific variation in this barcoding region, the distance within and between species of moa, rhea, emu, cassowary, and kiwi were also estimated in the same way using published sequences obtained from GenBank (Accessions: *Euryapteryx curtus* KF888653, AY833118; *Euryapteryx geranoides* AY833116, AY833117, AY833114, AY833115, AY833119; *Emeus crassus* AY833120, AF338712, NC\_002673, AY833121; *Anomalopteryx didiformis* AY833122, AF338714, NC\_002779; *Dinornis robustus* AY833123, AY833124, AY822127, AY833128, AY8833125, AY833126; *Dinornis giganteus* AY016013, NC\_003672; *Dinornis novaezealandiae* AY833129; *Megalapteryx didinus* AY833130; *Pachyornis elephantopus* AY833105, AY833106, AY833107, AY833108; *Pachyornis mappini* AY833109, AY833110, AY833111; *Pachyornis australis* AY833112, AY833113; *Rhea americana* AF090339, JN801968,

**TABLE S6.8.11** | CALIBRATIONS USED FOR MOLECULAR DATING.

Node	Minimum age (Ma)	Maximum age (Ma)	Justification	Reference
Emu/Kiwi	24.5	72.5	Minimum= <i>Emuarius</i> = oldest generally agreed crown member. Maximum covers absence of non-ostrich (or any) palaeognathae from global Maastrichtian avifaunas.	Worthy TH, Hand SJ, Archer M. 2014. Phylogenetic relationship of the Australian Oligo-Miocene ratite <i>Emuarius gidju</i> Casuariidae. <i>Integr. Zool.</i> <b>9</b> : 148-166.
Elephant bird/Kiwi	47.4	61.8	95% HPDs of the best current available estimate for the timing of divergence	Grealy et al. (2017)

JN801969, NC\_000846, Y16884; *Rhea pennata* (*Pterocnemia pennata*) AF338709, JN801970, JN801971, JN801972, NC\_002783; *Casuarius casuarius* AF338713, NC\_002778; *Casuarius bennetti* AY016011, CBU76058; *Dromaius novaehollandiae* AY016014, NC\_002784; *Apteryx australis* AAU76057; *Apteryx australis mantelli* EU525309, EU525313, EU525314, EU525315, EU525318, EU525316, EU525317, AY016010, EU525308, 525311, 525312, 525310; *Apteryx rowi* EU525322, EU525323, EU525324; *Apteryx haastii* AF338708, NC\_002782, EU525306, EU525307, EU525305; *Apteryx owenii* EU525319, EU525320, GU071052, NC\_013806, EU525321). Analyses were also conducted using the full *COI* gene with (i) complete deletion of missing bases, (ii) pairwise deletion of missing bases, and (iii) partial deletion of missing bases (85%), as well as with (a) data from all 20 taxa sequenced and (b) data from the 13 taxa exhibiting less than 10% missing bases. The overall pattern of intra- versus inter-specific variation in *COI* obtained from these alternative analyses was not significantly different from the results reported in the main manuscript (data available upon request).

#### **S6.8.19 CHARACTERISATION of EGG SHELL MICROSTRUCTURE by MICRO-CT**

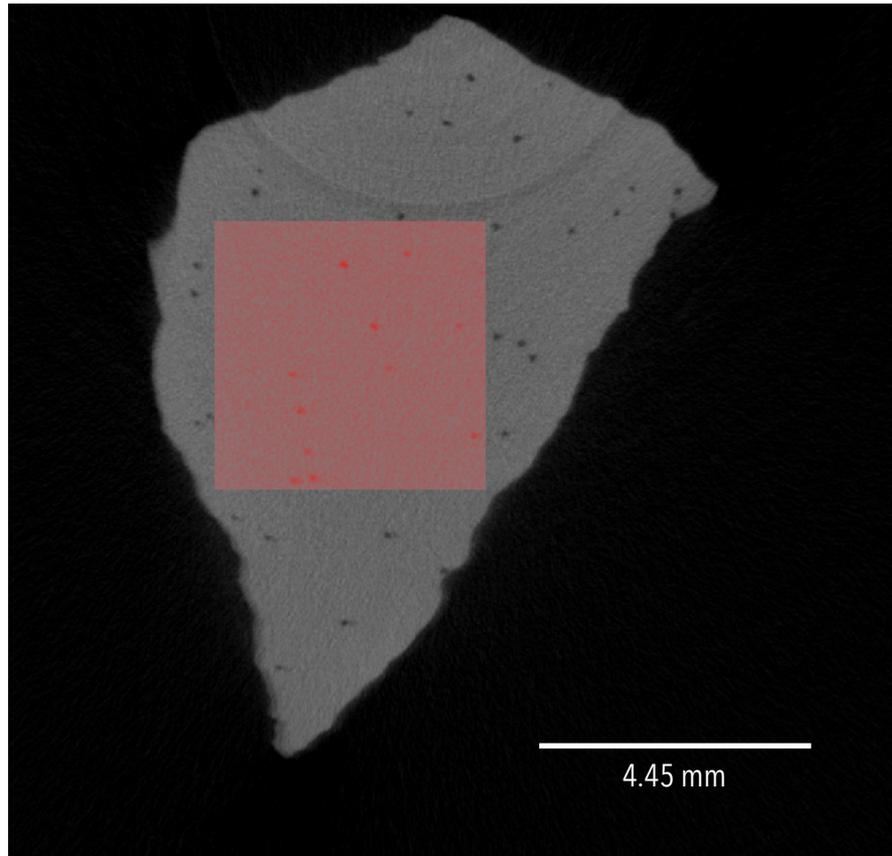
20 eggshell specimens (Table S6.8.12) were submitted to the Centre for Microscopy, Characterisation and Analysis at the University of Western Australia (WA, Australia) for imaging on the Skyscan 1175 micro-CT (*Bruker-microCT*, Kontich, Belgium). The operating conditions of the CT include: a 90 kV voltage, a 278  $\mu$ A current, a 0.1 mm thick copper filter, an exposure time of 2092 ms, a resolution setting with pixel size set to 9 microns, a rotation angle and range of 0.7° of a 360° rotation, and a frame averaging of two.

**TABLE S6.8.12** | EGGSHELL SPECIMENS IMAGED WITH MICRO-CT.

Field ID	AD#	DNA?	Location	Co-ordinates	Thickness (mm)	Amino acid racemisation	Estimated age (ya)
M06-M028	1666	Yes	Talaky	-25.4, 45.7	4.12	0.0450	1500+/- 500
NA	2409	Yes	S. Salary	-22.6, 43.3	3.5	NA	NA
M07-M050	2117	No	E. Mandrare River	-25.1, 46.4	3.71	0.0470	NA
M06-M164	2118	Yes	Itampolo	-24.7, 43.9	3.15	0.0645	NA
TONY2 A1-1-3	1739	Yes	Lakatom-Bato	-22.1, 43.2	4.01	NA	NA
M06-M112	2135	Yes	Faux Cap	-25.6, 45.5	2.97	0.0615	NA
NA	2384	Yes	Ampasilava	NA	3.46	NA	NA
NA	2387	No	Andavadoaka	-22.0, 43.2	2.90	NA	NA
M07-M095	2134	No	Big Dune Near Lake	-25.1, 46.5	2.61	0.070	1830+/-15
CHANTAL2	1662	Yes	Segeny	-23.5, 43.8	2.16	0.035	1855+/-15
M06-M089	2125	Yes	Cap St. Marie	-25.6, 45.1	1.14	0.060	NA
M06-M124	2109	No	Lavanovo	-25.4, 44.8	1.11	0.0805	NA
M07-M055	2111	No	E. Mandrare River	-25.2, 46.4	1.08	0.0560	NA
M06-M164	2112	Yes	Itampolo	-24.7, 43.9	1.51	0.0955	NA
M06-M229	2113	Yes	Anakao	-23.7, 43.6	1.4	0.0690	NA
M07-M199	1293	Yes	Isle E. Ambolobozokely	-12.4, 49.6	1.94	0.0605	NA
M07-M192	1292	No	Isle E. Ambolobozokely	-12.4, 49.6	2.19	0.0680	NA
M07-M167	1295	No	Isle E. Ambolobozokely	-12.5, 49.6	2.10	NA	NA
M07-M217	2139	Yes	E. Ramena Beach	-12.3, 49.4	2.69	0.0905	NA
M07-M125	2140	No	Irodo	-12.6, 49.5	1.89	0.0840	NA

*Bruker Nrecon* software was used to reconstruct projection images using a modified Feldkamp cone-beam algorithm. Image compensation settings were set to: smoothing = 3, beam hardening correction = 30%, ring artifact reduction = 20, thresholding = 0.00-0.06. *Bruker CTAn v1.16.4.1+* (SkyScan 2003-2011, Bruker microCT 2012-2016) software was used for analysis of the pores within the eggshells. These analyses included a 2D analysis of pore density and pore area, as well as a 3D analysis of pore volume and percentage porosity. All settings were optimised. An approximately 20.07 mm<sup>2</sup> (4.45 x 4.45 mm) region of interest was selected over a 1 mm thick volume of interest (VOI; approximately 100 slices). The region of interest (ROI) was placed in the centre of the specimen (Figure S6.8.2) in order to minimise the effects of imaging artifacts, and potential weathering or erosion of pores close to the surfaces of the eggshell, on the analysis. The volume of interest was spread approximately 50 slices above and below the centre slice. For segmentation and analysis, a task list comprised of the following plug-ins was used: (1) global thresholding = 0-65, (2) despeckle (remove white speckles in 3D less than  $x$  voxels) = 400 voxels, (3) bitwise operation = Image + ROI, (4) despeckle (remove white speckles in 2D less than  $x$  pixels) = 15 pixels, (5) save bitmaps (save all 2D slices), (6) 3D analysis, (7) bitwise operation = NOT Image, and (8) 3D Analysis. Pore density was calculated as the total number of objects divided by the total area for the ROI. Average pore area was calculated as the mean of area of pores within the ROI. Total pore area was calculated as the sum of the pore areas within the ROI. Average pore volume was calculated as the total pore volume divided by the number of pores within the VOI. Percent porosity was calculated as the total pore volume as a percentage of the total VOI. Results are recorded in Table S6.8.13.

Images were visualised and rendered in the software FEI Avizo Fire v.8.1.1 (*Konrad-Zeuse-Zentrum* Berlin 1995-2014; FEI, SAS 1999-2014). The outer and inner surface views were volume rendered directly over the reconstructed slices. The pore structure surface view was performed on a 3D Model (.stl), which was created using the Marching Cubes 33 algorithm in the Bruker CTAn software. Raw data, reconstructed projection images, and rendered images are available online at DataDryad, doi: TBA.



---

**FIGURE S6.8.2** | AN EXAMPLE OF THE 20.07 mm<sup>2</sup> (4.45 X 4.45 MM) REGION OF INTEREST (ROI) selected for both 2D and 3D analysis of pore density, area, volume, and porosity. The sample shown is AD#1666. The rings visible on the slice are artifacts created by the scanning process and are removed by the ‘despeckling’ plugin.

**TABLE S6.8.13** | FEATURES OF EGGHELL MICROSTRUCTURE including various pore dimensions determined through micro-CT analysis.

Genus	Region	Sample	Thickness (mm)	Pore density (pores/mm <sup>2</sup> )	Average pore area (mm <sup>2</sup> )	Total pore area (mm <sup>2</sup> )	Pore area %	Average Pore volume (mm <sup>3</sup> )	Porosity %
<i>Aepyornis</i>	South	AD#1666	4.120	0.149	0.004	0.012	0.060	0.000	0.116
<i>Aepyornis</i>	South	AD#2118	3.150	0.698	0.008	0.116	0.576	0.003	0.597
<i>Aepyornis</i>	South	AD#2117	3.710	0.947	0.007	0.128	0.637	0.002	0.696
<i>Aepyornis</i>	South West	AD#2409	3.500	0.997	0.004	0.087	0.432	0.001	0.486
<i>Aepyornis</i>	South West	AD#1739	4.010	1.644	0.019	0.628	3.131	0.003	3.157
<i>Aepyornis</i>	South	AD# 2135	2.970	0.897	0.015	0.263	1.311	0.012	1.509
<i>Aepyornis</i>	South	AD#2134	2.610	0.847	0.009	0.158	0.785	0.003	0.725
<i>Aepyornis</i>	South West	AD#1662	2.160	0.747	0.008	0.127	0.631	0.008	4.759
<i>Aepyornis</i>	South West	AD#2387	2.900	0.698	0.018	0.248	1.236	0.003	1.268
<i>Aepyornis</i>	South West	AD#2384	2.950	0.548	0.006	0.070	0.351	0.001	0.331
<i>Aepyornis</i>	North	AD#1293	1.940	1.046	0.007	0.139	0.692	0.003	0.676
<i>Aepyornis</i>	North	AD#2140	1.890	0.947	0.010	0.183	0.912	0.002	0.930
<i>Aepyornis</i>	North	AD#1295	2.100	0.648	0.004	0.057	0.285	0.001	0.293
<i>Aepyornis</i>	North	AD#1292	2.190	0.698	0.009	0.121	0.601	0.005	0.750
<i>Aepyornis</i>	North	AD#2139	2.690	0.797	0.019	0.310	1.546	0.015	1.521
<i>Mullerornis</i>	South	AD# 2125	1.140	0.100	0.006	0.012	0.062	0.001	0.059
<i>Mullerornis</i>	South	AD#2111	1.080	0.199	0.004	0.015	0.074	0.000	0.048
<i>Mullerornis</i>	South	AD#2109	1.110	0.448	0.006	0.051	0.256	0.001	0.230
<i>Mullerornis</i>	South	AD#2112	1.510	0.698	0.011	0.155	0.774	0.004	0.728
<i>Mullerornis</i>	South West	AD#2113	1.400	0.100	0.002	0.004	0.021	0.000	0.042

### S6.8.20 STABLE ISOTOPE ANALYSIS

Prior to preparing eggshell for isotopic analyses the fragments are mechanically cleaned by grinding, then reduced in mass by one third with the stoichiometric addition of 2N HCl *in vacuo* following procedures outlined in (Miller et al. 2005). For  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}_{\text{org}}$ , an approximately 5-mg subsample was placed in a silver capsule and dissolved with three sequential 20-microliter aliquots of 6 N HCl-sequanal grade. The residual acid was evaporated at room temperature under a fume hood for approximately 24 hr; remaining water was removed in an oven at 85°C for 24 to 36 hr and stored at 50°C under  $\text{N}_2$  until analysis.  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}_{\text{org}}$  were determined using an elemental analyzer (NC 2500; CE Elantech, Lakewood, NJ) interfaced with *Thermo Finnigan* (San Jose, CA) Delta Plus XL or Delta V Plus mass spectrometers (Carnegie Institution of Washington, Washington, DC). For  $\delta^{13}\text{C}_{\text{carb}}$  and  $\delta^{18}\text{O}$ , approximately 0.1 to 0.2 mg of powdered eggshell was placed in an exetainer vial and flushed with helium to remove air in a Thermo Fisher Gas Bench. Six drops of 100%  $\text{H}_3\text{PO}_4$  were added to each vial manually, then placed into a heating block at 70°C. For each batch of 40 to 80 eggshell powders we analyzed several blank vials and multiple standards, including internal lab standards along with NBS-18, NBS-19, and RO-22 (Isoanalytical). Evolved  $\text{CO}_2$  was analyzed on *Thermo Finnigan* Delta XL Plus or Delta V Plus mass spectrometer. The  $\delta^{18}\text{O}$  is measured in ‰ with respect to Standard Mean Ocean Water (SMOW) and converted to PDB; an internal carbonate standard analyzed six times had standard deviations of  $\pm 0.15\text{‰}$  for  $\delta^{18}\text{O}$  and  $\pm 0.25$  for  $\delta^{13}\text{C}_{\text{carb}}$ .

### S6.8.21 SUPPLEMENTARY REFERENCES

- Brooks AS, Hare PE, Kokis JE, Miller GH, Ernst RD, et al. 1990. Dating Pleistocene archaeological sites by protein diagenesis in ostrich eggshell. *Science* **248**: 60-64.
- Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. 2001. Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**: 704-707.

- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**.
- De Los Monteros AE. 2003. Models of the primary and secondary structure for the 12S rRNA of birds: a guideline for sequence alignment. *DNA sequence* **14**: 241-256.
- do Amaral FR SF, Wajntal A. 2010. Towards and assessment of character interdependence in avian RNA phylogenetics: a general secondary structure model for the avian mitochondrial 16s rRNA. *Molecular Phylogenetics and Evolution* **56**: 498-506.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
- Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**: 2153-2155.
- Grealy A, Phillips M, Miller G, Gilbert MTP, Rouillard J-M, et al. 2017. Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Molecular Phylogenetics and Evolution* **109**: 151-163.
- Guindon S and Gascuel O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systems Biology* **52**: 696-704.
- Haddrath O and Baker AJ. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. *Proceedings of the Royal Society B-Biological Sciences* **268**: 939-945.

- Huson DH, Auch AF, Qi J, Schuster SC. 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377-386.
- Jonsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**: 1682-1684.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059-3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.
- Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**:111-120.
- Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA. 2012. Setting the stage—Building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger* **194**: 3-6.
- Laslett D and Canback B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**: 172-175.
- Liu J DQ and Gao L. 2016. The complete mitochondrial genome of *Apteryx mantelli*. Unpublished.
- Miller GH. 2000. Isoleucine epimerization in eggshells of the flightless Australian birds, *Genyornis* and *Dromaius*. In: Goodfriend GA, Collins MJ, Fogel ML, Macko SA, Wehmiller JF, eds. *Perspectives in amino acid and protein geochemistry*. Oxford University Press: New York.

- Miller GH, Fogel ML, Magee JW, Gagan MK, Clarke SJ, Johnson BJ. 2005. Ecosystem collapse in Pleistocene Australia and a human role in megafaunal extinction. *Science* **309**: 287-290.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*: 1-8.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Phillips MJ, Gibb GC, Crimp EA, Penny D. 2010. Tinamous and Moa Flock Together: Mitochondrial Genome Sequence Analysis Reveals Independent Losses of Flight among Ratites. *Systems Biology* **59**: 90-107.
- Phillips MJ and Pratt RC. 2008. Family-level relationships among the Australasian marsupial "herbivores" (Diprotodontia: Koala, wombats, kangaroos and possums). *Molecular Phylogenetics and Evolution* **46**: 594-605.
- Posada D and Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.
- Rambaut A and Drummond AJ. 2007. TreeStat v1.2: Tree statistic calculation tool. Version 1.2.
- Ruokonen M KL. 2002. Structure and evolution of the avian mitochondrial control region. *Molecular Phylogenetics and Evolution* **23**: 422-432.
- Swofford DL. 2003. PAUP\*. Version 4.0b. Sunderland, Massachusetts: Sinauer Associates.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725-2729.

Wehmiller J. 2013. Interlaboratory comparison of amino acid enantiomeric ratios in Pleistocene fossils. *Quaternary Geochronology* **16**: 173-182.

Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.

## 6.9 EPILOGUE

---

In this chapter, we applied and improved upon the methods cultivated in Chapter 5 (specifically, we further optimised aDNA extraction from eggshell and employed improved hybridisation enrichment protocols) to reconstruct multiple additional mitogenomes from aDNA extracted from elephant bird eggshell. These genomes were used to investigate elephant bird phylogeography for the first time. Using state-of-the-art microscopy techniques, the micromorphology of elephant bird eggshell was also characterised and used alongside the genetic results to revisit current taxonomic classifications of elephant bird species. In light of the morphological and molecular evidence gleaned from their eggshell, we advocate for a major revision in elephant bird taxonomy. The information that we have obtained about elephant bird biology and evolution would not have been possible without the genetic information retrieved from their eggshell.

Not presented in this study, we also captured nuclear DNA from the same extracts. These sequences represent over 100 nuclear loci that have been shown to be of functional importance: that is, we targeted genes that are involved in the development or regulation of multiple adaptive phenotypes, including body size, feather development, eggshell development, olfactory reception, wing development, and beak development. In the following months, these sequences will be compared with those from other birds in order to identify functional mutations that may be responsible for elephant bird adaptations such as gigantism, flightlessness, cursoriality, and of course, their gigantic eggs.

The addition of phylogenetic information from nuclear loci would also strengthen our understanding of elephant bird taxonomy. Additionally, proteins have been found to preserve well in eggshell<sup>4 5</sup>, which is particularly important in warm, tropical climates where nuclear DNA may not preserve well, and would offer additional molecular evidence regarding their phylogenetic relationships and taxonomy, as well

---

<sup>4</sup> Brooks AS, Hare PE, Kokis JE, Miller GH, Ernst RD, et al. 1990. Dating Pleistocene archaeological sites by protein digenesis in ostrich eggshell. *Science* **248**, 60-64.

<sup>5</sup> Demarchi B, Hall S, Roncal-Herrero T, Freeman CL, Woolley J, et al. 2016. Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**; e17092.

as provide insight into phenotype and adaptation. Finally, we are currently re-analysing the results presented here with the inclusion of an additional two mitochondrial genomes from skeletal type specimens described in Yonezawa et al. (2016). These up-dated results can be found in the published version of the article.

These studies further demonstrate that genomic information can be obtained from the eggshell of extinct birds living in warm-climate ecosystems, and highlight how important this information can be for understanding the evolution and adaptation of extinct organisms. As such, future efforts will focus on using the methods developed here to retrieve aDNA from other extinct birds warm-climates, such as the Australian megafaunal bird *Genyornis*. How these research directions might develop is further explored in the next chapter (Chapter 7).

— CHAPTER 7 —

---

GENERAL DISCUSSION

---

*What we call the beginning is often the end. And to make an end is to make a beginning. The end is where we start from.*

- TS Eliot

## 7.1 GENERAL DISCUSSION

---

**T**HROUGHOUT this thesis, methods to extract DNA from novel ancient substrates within a variety of warm-climate settings have been further developed, and the work has demonstrated how this genetic information can be used to explore a multitude of evolutionary questions. This final chapter draws together the concepts examined in each of the previous chapters, reviews recent literature, places this body of work within the context of current aDNA knowledge, and discusses future research possibilities.

### 7.1.1 FILLING THE GAP:

#### WHAT DID WE LEARN?

At the outset of this thesis, scientists within the field of ancient DNA were just beginning to explore the use of novel substrates, such as mixed-source samples, as sources of aDNA. Those studies that did retrieve aDNA from environmental samples such as lake sediments, or other mixed-source samples such as coprolites, still focused their attention on samples from frozen or permafrost environments. As such, until 2013, the aDNA literature was heavily biased towards the study of single-source samples from cool-climates, resulting in a heavily euro-centric view of the past (albeit necessarily given the limits of technology at the time). Here, we aimed to optimise methods for retrieving and analysing aDNA from various novel substrates, both multi-sourced (bulk bone) and single-sourced (eggshell), in order to characterise past biodiversity, estimate aDNA preservation, and reconstruct the evolutionary history of vertebrates that once inhabited warm-climate ecosystems (Australia and Madagascar). What was ultimately achieved is summarised below:

**THE BULK BONE METABARCODING METHOD IS LARGELY ACCURATE AND REPRODUCIBLE BUT ASPECTS REQUIRE CAREFUL CONSIDERATION (CHAPTER 2).**

Chapter 2 further developed the bulk bone metabarcoding method, testing how accurately the method is able to genetically reproduce morphological taxonomic identifications from well-characterised fossil assemblages in Australia and New

Zealand, as well as how precisely those genetic IDs could be reproduced from one subsample or extract to the next; something that had not been previously quantified. Achieving this enabled practical guidelines for best-practice bulk bone workflows for both palaeontologists/archaeologists and geneticists to follow when implementing the BBM method to be suggested. Although the method is largely reproducible and accurate, several caveats need to be taken into consideration at every stage of the experimental design in order to mitigate their effects. The results of this study formed the foundation of the chapters that follow it.

**BBM CAN BE USED TO ASSESS aDNA PRESERVATION AND IDENTIFY FOSSIL ASSEMBLAGES FROM WITHIN WARM-CLIMATES (CHAPTER 3).** As the Naracoorte Caves are a rich source of fossils that have yielded so much information about the ecological history of South Australia, Chapter 3 set out to retrieve aDNA from fossil bulk bone within this World Heritage site for the first time, and characterise its preservation. In this published chapter, a demonstration of how the BBM method can be used to assess aDNA preservation without the destruction of rare specimens was conducted for the first time. The BBM data was also used to identify small vertebrate assemblages that may be difficult to morphologically distinguish from post-cranial remains (including bird, amphibian, and small marsupials), even despite significant aDNA degradation. We identified several families that were not previously known from the fossil record at Naracoorte, further supporting the use of aDNA to complement traditional morphological approaches to taxonomic identification. In this way, the BBM method was shown to be versatile in its applications, and succeeded where previous attempts to retrieve aDNA from single-source bones did not. These data suggest that BBM can contribute to a more thorough understanding of past biodiversity and to conservation efforts, particularly when establishing former ranges for currently extirpated and endangered species in warm climates like Australia.

**BBM CAN BE USED TO ESTABLISH A ZOOARCHAEOLOGICAL RECORD WHERE MORPHOLOGY CANNOT (CHAPTER 4).** Can bulk bone metabarcoding be used to identify morphologically indistinguishable archaeological fish bones from a tropical environment? Applying the core concepts laid-out in the previous chapters, the BBM method was used to identify bulk archaeological fish bone from an exposed coastal

site in Madagascar. Although degraded, aDNA was retrieved and provided the first description of an archaeological fish assemblage from Southwest Madagascar with 23 families of fish identified. The ecological affiliations and behaviors of the taxa identified provided clues about the subsistence practices of past peoples, including their primary reliance on reef fishes, possible engagement in night fishing, and potential over-exploitation of large-bodied, high-trophic level species. Chapter 4 illustrates that aDNA information from bulk bone can provide insights the past biodiversity even in poor preservation conditions, and despite the absence of morphological data. In this case, knowledge of how the marine ecosystem has been potentially altered by human activity is vital if today's people are to live sustainably from the environment.

**PALAEOGENOMES CAN BE RETRIEVED FROM FOSSIL AVIAN EGGSHELL IN WARM-CLIMATE ECOSYSTEMS AND BE USED TO INFORM EVOLUTIONARY HISTORY (CHAPTER 5).** Chapter 5 aimed at retrieving aDNA from fossil eggshell of the extinct Madagascan elephant bird (*Aepyornis* sp.) in order to reconstruct a complete mitochondrial genome that would be used to place them within the avian phylogeny. In addition to achieving this aim, nuclear DNA was also recovered from the eggshell for the first time, whereas previously, nuclear DNA was not retrieved from bone due to poor DNA preservation caused by Madagascar's hot climate. The complete mitogenome and nuclear DNA were used to confirm the recent finding that elephant birds are most closely related to kiwi, and to refine dates for the evolutionary divergences among palaeognaths. The results indicated that palaeognaths rapidly diverged after the extinction of the dinosaurs, and became flightless independently after dispersing to isolated but ecologically similar niches. Thus, aDNA from a sub-Saharan extinct bird helped change long-held views on avian evolution, emphasising that we should not neglect studying past organisms that inhabited warm climates, regardless of the challenges. Eggshell was further shown to contain a high proportion of endogenous aDNA relative to bacterial aDNA, indicating that it is an ideal substrate for obtaining aDNA from extinct birds, especially in warm environments.

**EGGSHELL ADNA CAN BE USED TO DESCRIBE PHYLOGEOGRAPHY AND REASSES TAXONOMY IN EXTINCT BIRDS (CHAPTER 6).** The methods used for the first time with eggshell in Chapter 5 were built upon in Chapter 6, with further optimisation of

protocols for aDNA extraction from eggshell and hybridisation enrichment, that allowed multiple full mitochondrial genomes to be obtained from elephant bird eggshell specimens throughout several regions of Madagascar, as aimed. Using these genomes, the evolutionary relatedness of elephant birds from across Madagascar could be reconstructed for the first time. This chapter also combined genetic and morphological approaches (including the use of micro-CT technology to examine eggshell microstructure) to distinguish elephant bird taxa, which lead to a revision in elephant bird taxonomy and better understanding of the processes that shaped their diversification. Again, this study supports the idea that eggshell is a reservoir for the preservation of high-quality aDNA in warm-climates. The results of this study add to the growing body of research into the history of Madagascar, and the evolution and extinction of its native fauna and flora that remain under considerable threat today.

### **7.1.2 A GRAIN *of* SALT:**

#### **WHAT COULDN'T WE ACHIEVE?**

Overall, the aims we set out to achieve were met and provide much-needed information about the palaeontological and archaeological history of several warm-climate environments for the first time. However, the methods developed herein are not without their limitations.

As discussed extensively in Chapter 2 (S2.8.13), there are several limitations to the bulk bone metabarcoding method that remain. For one, PCR bias is still a problem in metabarcoding workflows and we did not quantify the individual contribution of PCR bias to variance in the results. In hindsight, this would have been a wise direction to take, though it would have tripled the number of index tag combinations required, which would have significantly increased the cost of the experiment. Another limitation of the BBM method is that species-level resolution could not be achieved in the taxonomic identifications made throughout Chapters 2, 3, and 4, which more-often-than-not could only be made to the family-level with confidence. This limitation comes from the incompleteness of the genetic reference database (GenBank) and instability in the taxonomic framework, as well as the short length of the mitochondrial barcoding regions targeted. With new genetic references constantly being added to GenBank, this limitation will surely diminish with time but

will require scientists to devote specific energy to helping build upon the existing databases; for instance, some are sequencing museum specimens in order to add to the database of modern and historical species barcodes (Staats et al. 2013; Burrell et al. 2015; Xu et al. 2015). One positive is that the data presented here can be easily accessed remotely from around the world and reanalysed at any time in the future. However, the use of short barcoding regions will remain a problem without an obvious solution because the level of aDNA preservation will always dictate the maximum length that can be amplified, and at least for now, this is beyond our control.

Some studies have used shotgun data as an alternative to metabarcoding for identifying taxa; however, we weren't able to achieve this in Chapter 3. Again, this is a limitation of the reference database: most sequences, many of which would have been nuclear, did not have a reference in GenBank, resulting in, for example, the closest match to many Australian marsupials being South American marsupials. To be able to identify organisms through shotgun sequencing of bulk bone would be desirable as it could add extra support to the IDs made through metabarcoding, but it requires a fairly complete full-genome genetic reference database and a large budget to generate enough coverage of the endogenous DNA that is in extremely low proportion relative to exogenous DNA. 'Phylogenetic intersection analysis' (PIA) has been developed to extract taxonomic IDs from shotgun data (Smith et al. 2015), and may be a promising method to explore in the future; however, the authors estimate that this analysis only has an accuracy of 81%. Furthermore, while we were able to show that general aDNA preservation at a site can be assessed through shotgun sequencing of bulk bone, we cannot say with certainty the factors that may have influenced the rate of DNA decay, as data such as sediment pH and moisture content were not recorded. Although this was not an aim of the study, it may have allowed us to better predict or explain the limits of aDNA preservation at the site.

In Chapters 5 and 6, a large proportion of eggshells tested did not yield informative, endogenous aDNA. These eggshells were generally the oldest, or were not found *in situ* (i.e., they were collected from the surface). This suggests that while the preservation of aDNA in eggshell might be superior to that of bone and other substrates from warm climates, there are limitations imposed by the level of

exposure to the environment and the age of the sample. It is also difficult to determine how ubiquitous DNA preservation in fossil avian eggshell is. To date, aDNA has only been isolated from fossil ratite eggshell, which is significantly thicker and more robust in comparison to the eggshell of most other birds. Whether DNA can be retrieved from the thinner eggshell of other extinct birds remains to be tested but is nevertheless promising. Regardless, scientists continue to optimise protocols for aDNA extraction from even well-studied substrates (Benoit et al. 2013; Hoffmann and Griebeler 2013; Damgaard et al. 2015), suggesting that aDNA extraction from any substrate probably will always need to be optimised on a per-sample basis, and that such optimisation should be an essential component of every aDNA study.

### **7.1.3 THE ADNA SPHERE:**

#### **WHERE DO WE FIT IN?**

Despite the limitations of the methods, our research has made a significant contribution to the field of ancient DNA, with some of the published manuscripts presented in this thesis being cited by others (Bolohan 2016; Douglass 2016; Giovas et al. 2016; Leonardi 2016). Table S7.4.1 shows that in the three years since the commencement of the work presented here, there still remains a bias toward studies focusing on amplifying short fragments of mitochondrial DNA from single-source ancient bones. In particular, most of what is known about human prehistory through aDNA comes from the Northern Hemisphere (Slatkin and Racimo 2016; Table S7.4.1). However, there has been an increase in the number of studies published addressing metabarcoding workflows (Murray et al. 2015), and extracting aDNA from warm-climate settings, including Australia (Pacioni et al. 2015; Haouchar 2016), the Caribbean (Fabre et al. 2014; Gutiérrez-García 2014; Mendisco et al. 2015; Schroeder et al. 2015), the Pacific (Matisoo-Smith 2015), and Africa (Kistler 2014; Gallego Llorente et al. 2015). For example, aDNA was extracted from 400-5000 ya North African dromedary (*Camelus dromedarius*) archaeological specimens to investigate the history of their domestication (Almathen et al. 2016). In another study, the mitochondrial control region of ancient chicken specimens from across Polynesia and Southeast Asia revealed that dispersal route of chickens through the Pacific region (Thomson et al. 2014). These studies, and those presented in this

thesis, reinforce that aDNA from warm, arid, or tropical environments is a largely untapped resource of information about the past biodiversity and history of these regions, and are paving the way for future research to be conducted within this sphere that is rapidly growing.

The approaches taken here also reflect the growing trend in developing protocols for the extraction of aDNA from other novel substrates. This year, a study was published by Villanea and others that developed a method for the retrieval of aDNA from the shells of an extinct land snail (*Naesiotus* sp.; Villanea et al. 2016). aDNA extraction methods have also recently been optimised for the retrieval of aDNA from the seeds of grapes (*Vitis vinifera* ssp. *Sativa*; Gismondi et al. 2016), as well as ancient seeds from other economically important crop species, such as rice (*Oryza sativa*; Mutou et al. 2014). There has also been an increase in the number of studies that aim to retrieve nuclear DNA (Table S7.4.1), and utilise shotgun sequencing and enrichment methods; thus, our methods and techniques are both current and innovative.

#### **7.1.4 WHAT IS IT GOOD FOR?:**

##### **POTENTIAL APPLICATIONS**

This research may have potential applications for the conservation and management of threatened species: knowledge of past genetic diversity, geographic range expansion and contractions, and the biotic and abiotic factors that lead to population decline, fragmentation and extinction is critical for informing conservation decisions, as well as for the sustainable management of certain industries (e.g., fisheries), economies, and local communities. Such knowledge can be revealed through the analysis of aDNA retrieved from fossil bulk bone and eggshell. “Conservation archaeogenomics” or “conservation paleobiology” (cf., Rick and Lockwood 2013; Dietl et al. 2015; Hofman et al. 2015 for reviews on this topic) will play an increasingly important role in conservation as we enter the ‘anthropocene’. Recently, for example, aDNA helped track the extinction of *Megadyptes waitaha* and subsequent colonisation of their habitat by *Megadyptes antipodes* penguins following human arrival in New Zealand (Rawlence et al. 2015). Likewise, anthropogenic biodiversity loss was explored here in Chapter 4. Restoration and conservation of

today's biodiversity can only be achieved with a thorough understanding of how and why it has changed over time.

The concepts developed throughout this thesis have additional applications for other disciplines, particularly in relation to those that implement metabarcoding workflows. The biases and limitations of the bulk bone metabarcoding method explored here are largely the same as those that investigate biodiversity within unknown, degraded, and complex mixtures of DNA (Pedersen et al. 2015), such as bacterial metagenomics and microbiome reconstruction (Zieseimer et al. 2015). For example, ecologists are increasingly looking toward using environmental DNA (such as DNA extracted from sea water, sediment, and faecal matter) for monitoring biodiversity and conservation (Bohmann et al. 2014; Thomsen and Willerslev 2015; Barnes and Turner 2016). These fields may benefit from the methods explored throughout this thesis.

#### **7.1.5 WHAT'S NEXT?**

##### **RECOMMENDATIONS FOR FURTHER STUDY**

Although the information gathered in this thesis significantly contributes to our current body of knowledge within the field of palaeogenetics, the work presented herein will continue to be built upon as several avenues of research remain and new questions have arisen. These are discussed in detail below.

**GENETIC DIVERSITY CHANGE OVER TIME AND SPACE.** Not addressed in Chapter 2 was the collection and aDNA extraction of numerous bulk bone samples from another site on Kangaroo Island, Kelly Hill Cave, that range from the Holocene to as far back as the mid-Pleistocene, 50,000 years ago. As a land-bridge island, Kangaroo Island is especially significant because it can act as a natural control for studying the roles that evolutionary processes such as natural selection and genetic drift play in the evolution of mainland and island fauna, and we can examine how these roles change during times of connection and isolation. The Naracoorte site studied in Chapter 3 offers a mainland comparison because these sites overlap temporally, which is important for testing hypotheses about the evolutionary events that have shaped the biodiversity on KI and mainland SA. For example, the extirpated

bearded-dragon genus *Rankinia* has recently been discovered in the fossil record at Kelly Kill Cave, KI; specifically targeting hypervariable regions within the mitochondrial genome of *Rankinia*, aDNA from bulk bone specimens could be used to characterise the genetic variation that existed in KIs historical populations of *Rankinia* and how it changed over time, as well as compare this variation to both modern and historical mainland populations of *Rankinia*. This would help us investigate the processes responsible for the extirpation of *Rankinia* from KI and mainland SA, which may have implications for the conservation of remaining *Rankinia* populations in New South Wales, Victoria, and Tasmania.

**FURTHER OPTIMISATION OF METHODS.** Further experimentation could be carried-out to optimise the bulk bone metabarcoding method. For example, multiplexing barcoding primers such that several barcode products are amplified simultaneously could be a way to reduce the time, labor, and cost of metabarcoding workflows; however, this has yet to be tested and may result in additional PCR bias. Another way to potentially reduce PCR bias could be to perform emulsion PCRs (emPCRs), where the PCR reaction is partitioned into thousands of ‘micro-reactions’ within oil droplets such that one template is being amplified per droplet. Partitioning a PCR would mean that low-copy number templates, such as low-abundant species, have an equal opportunity to be amplified without being ‘swamped’ by more abundant templates, and are less likely to be underrepresented. Bulk bone aDNA shotgun libraries could also be enriched for endogenous DNA through the use of hybridisation capture, and potentially allow whole vertebrate mitochondrial genomes to be reconstructed from the mix.

As discussed in Chapters 1 and 2, being able to quantify aDNA is particularly important when dealing with mixed aDNA, such as DNA extracted from bulk bone. However, qPCR cannot accurately quantify very low starting amounts of DNA. The recent release of a new library quantification method, digital droplet PCR (ddPCR), may provide a solution to this problem. In one approach, ddPCR partitions template molecules into thousands of oil droplet micro-reactors, and counts the number of template-positive reactions to absolutely calculate the number of starting DNA templates solution prior to amplification; in another approach, a fluorescent probe is hybridised to PCR products following emPCR, which is then detected by flow

cytometry (Dressman et al. 2003; Hindson et al. 2011). ddPCR method is preferable over qPCR because its accuracy is not influenced by PCR efficiency. Therefore, inhibition caused by the presence of contaminants in aDNA will not influence quantification (Hoshino and Inagaki 2012), and the reduction in PCR stochasticity improves precision (Lin and Yao 2012). In addition, ddPCR does not rely on the presence of known standards for absolute quantification. The available ddPCR platforms are QuantStudio 3D (*Life Technologies*), RainDrop (*RainDance Technologies*), and QX100 (*Biorad*). Currently, these platforms only provide quantification using probes (*TaqMan*) designed to hybridise to a specific DNA sequence; however, this is only useful for quantifying single-source amplicon libraries, and not useful for quantifying shotgun libraries or metabarcoding libraries. The recently released QX200 (*Biorad*) is the only platform that provides this option. The more accurately templates can be quantified, the better the performance of the PCR, and the more reproducible the results will be (White et al. 2009). Furthermore, accurate quantification of template copy numbers could facilitate prioritisation of only the most promising samples for sequencing (for example, when screening eggshell aDNA samples). However, these platforms have not been optimised for use on aDNA templates; thus, this is a particularly important avenue for future exploration.

The development of third-generation sequencing technologies (TGS) may also help overcome some of the challenges that still face the field of aDNA. TGS technologies involve sequencing-by-synthesis in that is detected in real-time. The currently available TGS platforms include SMRT (single-molecule real-time) sequencing (*Pacific Biosciences*), and MinION (*Oxford Nanopore*; see Xu et al. 2009, Pareek et al. 2011, and Liu 2012 for in-depth comparisons of all TGS platforms and detailed explanation of the chemistries). The central tenet of TGS technologies is that DNA can be sequenced without prior amplification, thereby eliminating biases and errors that are introduced during PCR (Shokralla et al. 2012), as well as reducing the time, labor, and cost of library preparation (Xu et al. 2009; Pareek et al. 2011). TGS methods also claim to be faster and more accurate, require less starting template, and, depending on the platform, produce much longer read lengths than any of the current platforms (Table 1.2); this is particularly useful for *de novo* assembly of genomes where long read length is an advantage. For aDNA applications, TGS technology has

**TABLE 7.1** | SUMMARY OF SEQUENCING TECHNOLOGIES (including third-generation sequencing technologies in development).

<b>Sequencing technology</b>	<b>Key references</b>	<b>Generation</b>
Sanger sequencing	(Sanger et al. 1977)	First
MPSS	(Brenner et al. 2000)	Second
Multiplex polony sequencing	(Mitra et al. 2003; Shendure et al. 2005)	Second
454 Pyrosequencing	(Ronaghi et al. 1998; Margulies et al. 2007)	Second
Illumina (Solexa) sequencing-by-synthesis	(Bentley et al. 2008)	Second
SOLiD sequencing-by-ligation		Second
Ion Torrent sequencing by ion semiconductor	(Rusk 2011)	Second
DNA nanoball sequencing	(Drmanac et al. 2010; Porreca 2010)	Third
Heliscope single-molecule sequencing	(Braslavsky et al. 2003; Harris et al. 2008; Bowers et al. 2009; Thompson and Steinmann 2010)	Third
Single-molecule Real-Time sequencing	(Levene et al. 2003; Ju et al. 2006; Korlach et al. 2008; Eid et al. 2009)	Third
VisiGen Modified DNA polymerase and fluorescence resonance sequencing	(Selvin 2000)	Third
Nanopore DNA sequencing	(Deamer and Akeson 2000; Howorka et al. 2001; Lagerqvist et al. 2006; Soni and Meller 2007; Clarke et al. 2009; Stoddart et al. 2009; Derrington et al. 2010; Timp et al. 2010)	Third

Scanning tunneling microscopy (STM) sequencing	(Tao et al. 1993; Tanaka and Kawai 2003; Tanaka and Kawai 2009)	Third
Nano-knife edge probes Sequencing	(Xu et al. 2009)	Third
Sequencing by hybridization	(Hanna et al. 2000)	Third
Sequencing by mass spectrometry	(Monforte and Becker 1997; Edwards et al. 2005)	Third
Microfluidic Sanger sequencing	(Kan et al. 2004; Blazej et al. 2006; Chen et al. 2010)	Third
Transmission electron microscopy (TEM)-based DNA sequencing	(Bell et al. 2012)	Third
RNAP sequencing	(Greenleaf and Block 2006)	Third
<i>In vitro</i> virus high throughput sequencing	(Fujimori et al. 2012)	Third

proved useful for maximizing the amount of endogenous aDNA sequences retrieved (Orlando et al. 2011; Ginolhac et al. 2012). Orlando et al. (2011) showed that more Pleistocene horse DNA sequences could be retrieved using TGS over NGS. However, TGS approaches too have disadvantages: currently they are prohibitively expensive, and because they sequence single-molecules, the coverage and error profile will differ greatly from NGS technologies. Other TGS technologies currently in development are summarised in Table 7.1.

**CHARACTERISING THE FACTORS THAT INFLUENCE aDNA PRESERVATION.** Chapter 3 demonstrated how bulk bone can be used to evaluate and model aDNA preservation at a site; however, the extent to which different factors influence aDNA preservation remains to be fully investigated. Recent studies found a correlation between the degree of bone preservation determined by spectroscopic analysis, electron microscopy, and immunohistochemistry, and the presence of aDNA in bone (Sosa et al. 2013; Barta et al. 2014; Coulson-Thomas et al. 2015; Scorrano et al. 2015). Other studies have investigated the effect of factors such as depositional context, storage time, microbial profile and age on the success of aDNA amplification in various ancient specimens (Der Sarkissian et al. 2014; Heintzman 2014; Elsner et al. 2015). Specific measurements of the pH, salinity, humidity, microbial content, temperature, etc. of the depositional environment, along with shotgun sequencing of bulk bone aDNA, would allow us to examine whether there is a correlation between these factors and aDNA survival; this would then allow researchers to better predict the probability of aDNA preservation based on these metrics.

**FURTHER COMPARISON WITH MORPHOLOGICAL DATA.** As discussed in the epilogue to Chapter 4, the identifications of the ichthyofauna within the archaeological assemblage as determined through BBM should be compared to other archaeological sites in the same area where morphological identification of the fish bones is possible. In addition, another primer set targeting a short region of the fish *16S rRNA* gene is currently being used to amplify the same samples (as well as eight additional samples), in order to confirm and add support to our taxonomic IDs. In the future, it would be interesting to compare archaeological fish assemblages across both time

and space to examine how Malagasy subsistence practices differ between regions and throughout time.

**FULL GENOME SEQUENCING AND MAPPING FUNCTIONAL MUTATIONS.** Finally, the next step in the study of elephant birds will be to reconstruct their genome in full. This will require sequencing of several independently constructed shotgun libraries on the high-throughput NGS platform, HiSeq (*Illumina*). Having a full genome of the elephant bird will allow us to perform genome-wide screens for regions that may be under selection, as these regions are likely to be responsible for adaptations, most of which are encoded in the nuclear genome. Full-genome studies are becoming increasingly common in ancient DNA (see Harkins and Stone 2015, Marciniak et al. 2015, Parks et al. 2015, Sanchez-Quinto and Lalueza-Fox 2015, Sarkissian et al. 2015, and Leonardi 2016 for in depth reviews on palaeogenomics), and allow detailed analysis of population admixture (Sanchez-Quinto and Lalueza-Fox 2015; Schaefer et al. 2016), dispersal (Schroeder et al. 2015) and demography (Sams et al. 2015), and selective sweeps (Malaspinas 2016), particularly for humans (Paabo 2014; Allentoft et al. 2015; Brandt et al. 2015; Ermini et al. 2015; Paabo 2015; Fu et al. 2016; Table S7.4.1). As mentioned in Chapter 5 and 6, capture baits based on a number of loci that play a role in the generation of phenotype have been designed and used to enrich elephant bird shotgun libraries for these loci; in the near future, mutations that may have a functional impact on phenotypes such as gigantism and flightlessness will be mapped (this has been done for other ratites such as the moa; Huynen et al. 2014). Additionally, a full genome will allow researchers to further characterise the genetic diversity within elephant birds populations by genotyping polymorphic loci such as SNPs and microsatellites, as has been done for the moa (Allentoft et al. 2011). Genetic diversity can be measured by determining the heterozygosity by comparing how much the frequency of an allele differs between populations, and by estimating the number of shared haplotypes. These measures of genetic diversity can then be used to reconstruct the evolutionary history of elephant birds, including the direction of gene flow.

### 7.1.6 THE FUTURE *of* THE PAST:

#### WHERE ELSE IS THE FIELD HEADED?

Over the tenure of this thesis, the field of ancient DNA has advanced considerably. In terms of the types of questions that aDNA will be used to address in the future, the field is headed towards applying aDNA in a way that will achieve more practical outcomes. Some of these areas are briefly explored below.

**PALAEO-FUNCTIONAL GENOMICS.** One area of focus is on palaeo-functional genomics, where the functions and interactions of genes from extinct organisms are compared to their extant relatives. This includes investigating how gene expression is regulated by interaction with other genes, proteins, and the environment. It has been shown that often “spatial, temporal, and quantitative” changes in gene expression, caused by differences in non-coding regulatory elements (Prabhakar et al. 2008), microRNAs (Chen and Rajewsky 2007), or epigenetic modifications (Llamas et al. 2012), are responsible for evolution—not necessarily changes in genes themselves (Pask et al. 2008). For example, mutations in a non-coding sequence of human DNA (*HACNS1*) resulted in an adaptive increase in expression of limb development genes (Prabhakar et al. 2008). Similarly, an up-regulation of *Prx1* in the limbs of bats contributed to the evolution of the bat wing (Cooper and Tabin 2008), and SNP genotyping of a gene that regulates the expression of lactase in ancient Europeans suggests that the ability of Europeans to digest lactose arose earlier than previously thought (Krutli et al. 2014). “Therefore, the non-coding regions of extinct genomes are likely to hold the most important genetic information that defined a species” (Pask et al. 2008). Functional exploration of ancestral versus derived changes in coding regions and non-coding regions can reveal whether such changes were, or are, adaptive.

While the function of genes from extinct organisms can be inferred by protein modeling *in silico* (Fortes et al. 2013; Campbell and Hofreiter 2015), ancient gene function has also been examined *in vitro* (Campbell et al. 2010), and even *in vivo* (Pask et al. 2008). Campbell *et al.* (2010) examined the structure and function of mammoth haemoglobin relative to modern elephants by transforming the mammoth hemoglobin gene into a plasmid vector and assessing the properties of the expressed

protein. They identified an amino acid substitution that would allow haemoglobin to bind more oxygen at low temperatures, suggesting that these changes were under selection due to the cold climate environment of mammoths (Campbell et al. 2010). Changes observed in various other genes may account for other cold-adapted phenotypes such as skin and hair development (Lynch et al. 2015). Similarly, Pask et al. (2008) transformed a transcriptional enhancer from the extinct Tasmanian tiger (*Thylacinus cynocephalus*) into a mouse model to investigate how this element would change expression of the mouse gene *in vivo*. They found that this element regulated gene expression in the Tasmanian tiger the same way as its mouse orthologue (Pask et al. 2008). In this way, the function of protein coding and non-coding sequences from extinct animals can be observed directly.

**GENE EXPRESSION.** Gene expression can be examined by sequencing RNA or the ‘transcriptome’. While RNA is readily retrievable from fresh tissues, its notorious instability and propensity to degrade rapidly makes the possibility of obtaining ancient RNA (aRNA) highly unlikely, but not inconceivable (Ozsolak and Milos 2011). To date, aRNA has only been extracted and sequenced from viruses (Castello et al. 1999; Zhang et al. 2006; Smith et al. 2014), where the genome is largely made of RNA, and seeds (Rollo et al. 1991), where RNA is essential for germination. For example, peach latent mosaic virus RNA was extracted from 50-year-old dry leaf material (Comas et al. 2013), and RNA has been recovered from ancient maize kernels (Rollo 1985; Fordyce et al. 2013a). This is feasible because some viruses can ‘survive’ for decades *ex vivo*, and seeds can germinate after thousands of years lying dormant (Sallon et al. 2008). The circular secondary structure of viroid RNA may prevent it from degradation (Guy 2013), while the long-term survival of RNA in seeds may be due to increased stability caused by an association of the RNA with proteins (see Fordyce et al. 2013a, Guy 2013, see Guy 2014, and Fordyce et al. 2013b for in-depth reviews on aRNA).

**EPIGENETICS.** Alternatively, gene expression may be regulated by epigenetic modifications, such as cytosine methylation. The environment can trigger epigenetic modifications that may “participate in the rapid adaptation of species to changes in the environment” (Llamas et al. 2012). Thus, mapping such modifications in extinct organisms can add a new dimension to understanding the complex gene-environment

interactions that play a role in evolution. Cytosine methylation patterns can be assessed on a genome wide level using bisulphite treatment of DNA combined with NGS (Cokus et al. 2008; Smith et al. 2014a; Smith et al. 2014 b; see Gokhman et al. 2016 for a review on ancient epigenetics). For example, the cytosine methylation patterns of the extinct Pleistocene bison, *Bison priscus* were mapped using bisulphite sequencing (Llamas et al. 2012), and Pedersen et al. (2014) examined methylation patterns in 4000 ya palaeo-Inuit hair. In another example, Gokhman et al. (2014) reconstructed methylation patterns in Neanderthals and Denisovans and identified several regions of the genome that are differentially methylated in comparison to modern human that could account for the anatomical differences and susceptibility to disease (Gokhman et al. 2014).

**RESURRECTING EXTINCT ORGANISMS.** Being able to express the genes encoded by aDNA may be one step toward ‘resurrecting’ extinct organisms, a prospect that has always been appealing yet controversial. 2009 saw the first ‘revival’ of an extinct organism—a subspecies of the Pyrenean ibex (*Capra pyrenacia pyrenacia*) that went extinct nine years earlier was cloned from a cryopreserved tissue of the last living specimen (Folch et al. 2009; Pina-Aguilar et al. 2009). However, the prospect of bringing back to life long extinct organisms remains grim (Nicholls 2008). Several approaches for resurrecting extinct organisms have been proposed, including transfer of embryonic stem cells with undamaged ancient nuclei (Wakayama et al. 2008; Wakayama et al. 2011), synthesis of extinct genomes (Gibson et al. 2008) and recombinase-mediated genomic replacement, where large segments of an extant genome is replaced with segments from a related extinct genome (Wallace et al. 2007). In fact, Kato et al. (2009) recovered somatic nuclei from mammoth tissue and transferred them into mouse embryonic stem cells, and found that more than half were able to survive (Kato et al. 2009). Although this approach appears to be the most promising, resurrecting a live animal not only requires undamaged nuclei, but also requires a “suitable recipient oocyte” (Loi et al. 2011). In addition, much of an organism’s development is controlled by interactions between the genome and the environment, as well as interactions between the nuclear genome and organelle genome(s); even with a complete genome or nucleus, recreating the conditions for the correct development of an extinct organism would be prohibitively difficult (Huynen et al. 2012). Nevertheless, any developments in this area of aDNA research

will have no doubt have dramatic implications for the conservation of endangered or recently extinct organisms.

**HEALTH AND DISEASE.** Scientists are also looking to understand the evolution of infectious diseases in order to elucidate how diseases originate and spread, which has implications for modern health as new epidemics arise. Five whole genomes of *Yersinia pestis* (the Black Death) were reconstructed from bacterial DNA extracted from 18<sup>th</sup> century human teeth in Europe (Bos et al. 2016), and the genetic history of *Ascaris* was examined through the extraction of aDNA from parasite eggs inside a human coprolite (Oh et al. 2015). Other diseases that have been investigated include leprosy (*Mycobacterium leprae*; e.g., Suzuki et al. 2014), tuberculosis (*Mycobacterium tuberculosis*, e.g., Muller et al. 2014; Spigelman et al. 2015), and more (Table S7.4.1; Anastasiou and Mitchell 2013; Donoghue 2013; Frias et al. 2013; Darling and Donoghue 2014; Theves et al. 2014; Donoghue et al. 2015; Harkins and Stone 2015). Even vectors of disease can be uncovered through aDNA; 1000-year old *Mycobacterium tuberculosis* genomes were reconstructed from Peruvian human skeletons and were found to be more similar to tuberculosis strains hosted by seals and sea lions, indicating that these animals may have helped transmit the disease (Bos et al. 2014). As microbiomes influence human susceptibility to diseases such as obesity, ancient gut and oral microbiomes have also been examined through aDNA analysis of human coprolites and dental calculus (Warinner et al. 2014; Warinner et al. 2015a; Warinner et al. 2015b; Weyrich et al. 2015). For instance, Appelt et al. (2014) used a shotgun sequencing approach to characterise the gut virome of a 14<sup>th</sup>-century human from a coprolite. The role that aDNA can play in understanding modern human health is rapidly being recognised.

**AGRICULTURE AND DOMESTICATION.** The advent of agriculture has been shown to go hand-in-hand with the emergence of certain diseases and changes to the human microbiome, both of which have a significant impact on human health (Harper and Armelagos 2013). As such, many studies are examining the evolution of important agricultural species (e.g., legumes; Mikic 2015), and domesticates (e.g., pigs; Sapir-Hen et al. 2015) through aDNA analysis (Allaby et al. 2015). Delye et al. (2013) found through aDNA analysis of grass weed *Alopercurus myosuroides* herbarium specimens that historic populations of the plant contained herbicide resistance

genotypes that pre-date the invention of herbicides, which provides information about the evolution of herbicide resistance. In another example, the genomes of historical strains of potato blight (*Phytophthora infestans*) were reconstructed using aDNA extracted from Nineteenth-century potato and tomato herbarium specimens (Yoshida et al. 2013). Recently, the first archaeogenome of plants was published that allowed an exploration into the domestication of barley (Mascher et al. 2016). The results of these studies may have agricultural and economic implications.

**“OUT-OF-THIS-WORLD” APPLICATIONS.** There has been some suggestion that aDNA techniques may be necessarily applied to, for example, test for the presence of life on Mars (Willerslev and Cooper 2005). Alien substrates may provide an important opportunity to investigate the origins of life on Earth, including the hypothesis of panspermia (that life on Earth arose first on other planets; Willerslev and Cooper 2005). Any reports of nucleic acids found in space will need to be subjected to the same rigorous authentication criteria as aDNA to rule out contamination (Willerslev and Cooper 2005).

## 7.2 CONCLUSION

---

Since its inception, the field of ancient DNA has been fraught with many challenges; overcoming these challenges has helped mould the field into a legitimate and rapidly growing area of scientific inquiry that complements the traditional fields of palaeontology, archaeology, geology, and phylogenetics, and provides unique insights into the past. Although some of the challenges of working with aDNA have been addressed throughout this thesis, new challenges arise while others remain (and may continue to remain) unsolvable. Nevertheless, recent advances in aDNA methods—including those developed here for the retrieval of aDNA from novel substrates in warm-climate ecosystems—alongside next-generation sequencing technology have made it more feasible than ever to obtain high-quality genetic and genomic data from ancient substrates, allowing new questions to be explored and old question to be resolved. Ancient DNA continues to be proven a useful tool for studying past biodiversity, population history, phylogeography, palaeoenvironment, and evolutionary process, having implications for conservation and interdisciplinary

applications. However, as the technology continues to improve exponentially, the field of ancient DNA becomes less about how to obtain aDNA sequences, and more about deciphering the function of those sequences, assessing how they have changed, and reconstructing the role they may have played in evolution across time and space. The challenge that remains now is meeting the immense computational and bioinformatics demands of sequence data; to do this, it will become essential to have well designed experiments to test *a priori* hypotheses, and empirically optimised protocols. Nevertheless, new technologies allow us to push the boundaries of possibility, stepping the field into an exciting future of deeper exploration of the past.

### 7.3 REFERENCES

---

Allaby RG, Gutaker R, Clarke AC, Pearson N, Ware R, et al. 2015. Using archaeogenomic and computational approaches to unravel the history of local adaptation in crops. *Philosophical Transactions of the Royal Society B* **370**.

Allentoft ME, Oskam C, Houston J, et al. 2011. Profiling the Dead: Generating Microsatellite Data from Fossil Bones of Extinct Megafauna-Protocols, Problems, and Prospects. *PLOS One* **6**.

Allentoft ME, Sikora M, Sjogren KG, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* **522**: 167-172.

Almathen F, Charruau P, Mohandesan E, et al. 2016. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 6707-6712.

Anastasiou E and Mitchell PD. 2013. Palaeopathology and genes: Investigating the genetics of infectious diseases in excavated human skeletal remains and mummies from past populations. *Gene* **528**: 33-40.

- Barnes MA and Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* **17**: 1-17.
- Barta JL, Monroe C, Crockford SJ and Kemp BM. 2014. Mitochondrial DNA preservation across 3000-year-old northern fur seal ribs is not related to bone density: Implications for forensic investigations. *Forensic Science International* **239**: 11-18.
- Bell DC, Thomas WK, Murtagh KM, Dionne CA, Graham AC, Anderson JE and Glover WR. 2012. DNA Base Identification by Electron Microscopy. *Microscopy and Microanalysis* **18**: 1049-1053.
- Benoit JN, Quatrehomme G, Carle GF and Pognonec P. 2013. An alternative procedure for extraction of DNA from ancient and weathered bone fragments. *Medicine Science and the Law* **53**: 100-106.
- Bentley DR, Balasubramanian S, Swerdlow HP, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.
- Blazej RG, Kumaresan P and Mathies RA. 2006. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 7240-7245.
- Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, et al. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology Evolution* **29**: 358-367.
- Bolohan N, Ciorpac M, Matau F, Gorgan DL. 2016. Ancestral DNA - an incontestable source of data for Archaeology. *Studia Antiqua et Archaeologica* **21**: 157-188.

- Bos KI, Harkins KM, Herbig A, et al. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**: 494-497.
- Bos KI, Herbig A, Sahl J, et al. 2016. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* **5**.
- Bowers J, Mitchell J, Beer E, et al. 2009. Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods* **6**: 593-U560.
- Brandt G, Szecsenyi-Nagy A, Roth C, Alt KW and Haak W. 2015. Human paleogenetics of Europe - The known knowns and the known unknowns. *Journal of Human Evolution* **79**: 73-92.
- Braslavsky I, Hebert B, Kartalov E and Quake SR. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 3960-3964.
- Brenner S, Johnson M, Bridgham J, et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* **18**: 630-634.
- Burrell AS, Disotell TR and Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* **79**: 35-44.
- Campbell KL and Hofreiter M. 2015. Resurrecting phenotypes from ancient DNA sequences: promises and perspectives. *Canadian Journal of Zoology* **93**: 701-710.
- Campbell KL, Roberts JEE, Watson LN, et al. 2010. Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genetics* **42**: 536-540.

- Castello JD, Rogers SO, Starmer WT, Catranis CM, Ma LJ, et al. 1999. Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biology* **22**: 207-212.
- Chen K and Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics* **8**: 93-103.
- Chen Y-J, Roller EE and Huang X. 2010. DNA sequencing by denaturation: experimental proof of concept with an integrated fluidic device. *Lab on a Chip* **10**: 1153-1159.
- Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S et al. 2009. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* **4**: 265-270.
- Cokus SJ, Feng S, Zhang X, et al. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215-219.
- Comas I, Coscolla M, Luo T, et al. 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nature Genetics* **45**: 1176-1182.
- Cooper KL and Tabin CJ. 2008. Understanding of bat wing evolution takes flight. *Genes & Development* **22**: 121-124.
- Coulson-Thomas YM, Norton AL, Coulson-Thomas VJ, et al. 2015. DNA and bone structure preservation in medieval human skeletons. *Forensic Science International* **251**: 186-194.
- Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E et al. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports* **5**.

- Darling MI and Donoghue HD. 2014. Insights from paleomicrobiology into the indigenous peoples of pre-colonial America - A Review. *Memorias Do Instituto Oswaldo Cruz* **109**: 131-139.
- Deamer DW and Akeson M. 2000. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in Biotechnology* **18**: 147-151.
- Der Sarkissian C, Ermini L, Jonsson H, Alekseev AN, Crubezy E, et al. 2014. Shotgun microbial profiling of fossil remains. *Molecular Ecology* **23**: 1780-1798.
- Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, et al. 2010. Nanopore DNA sequencing with MspA. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 16060-16065.
- Dietl GP, Kidwell SM, Brenner M, Burney DA, Flessa KW, et al. 2015. Conservation Paleobiology: Leveraging Knowledge of the Past to Inform Conservation and Restoration. In: Jeanloz R, Freeman KH, editors. *Annual Review of Earth and Planetary Sciences* **43**: 79-103.
- Donoghue HD. 2013. Insights into ancient leprosy and tuberculosis using metagenomics. *Trends in microbiology* **21**: 448-450.
- Donoghue HD, Spigelman M, O'Grady J, Szikossy I, Pap I, et al. 2015. Ancient DNA analysis - An established technique in charting the evolution of tuberculosis and leprosy. *Tuberculosis* **95**: 140-144.
- Douglass K. 2016. The diversity of late Holocene shellfish exploitation in Velondriake, Southwest Madagascar. *The Journal of Island and Coastal Archaeology*: doi:10.108/15564894.2016.1216480.
- Dressman D, Yan H, Traverso G, Kinzler KW and Vogelstein B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for

detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 8817-8822.

Drmanac R, Sparks AB, Callow MJ, et al. 2010. Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays. *Science* **327**: 78-81.

Edwards JR, Ruparel H and Ju JY. 2005. Mass-spectrometry DNA sequencing. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **573**: 3-12.

Eid J, Fehr A, Gray J, et al. 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**: 133-138.

Elsner J, Schibler J, Hofreiter M and Schlumbaum A. 2015. Burial condition is the most important factor for mtDNA PCR amplification success in Palaeolithic equid remains from the Alpine foreland. *Archaeological and Anthropological Sciences* **7**: 505-515.

Ermini L, Sarkissian CD, Willerslev E and Orlando L. 2015. Major transitions in human evolution revisited: A tribute to ancient DNA. *Journal of Human Evolution* **79**: 4-20.

Fabre PH, Vilstrup JT, Raghavan M, Sarkissian CD, Willerslev E, et al. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters* **10**.

Folch J, Cocero MJ, Chesne P, et al. 2009. First birth of an animal from an extinct subspecies (*Capra pyrenaica pyrenaica*) by cloning. *Theriogenology* **71**: 1026-1034.

Fordyce SL, Avila-Arcos MC, Rasmussen M, et al. 2013a. Deep Sequencing of RNA from Ancient Maize Kernels. *PLOS One* **8**.

- Fordyce SL, Kampmann ML, van Doorn NL, Gilbert MT. 2013b. Long-term RNA persistence in postmortem contexts. *Investigative Genetics* **4**.
- Fortes GG, Speller CF, Hofreiter M, King TE. 2013. Phenotypes from ancient DNA: Approaches, insights and prospects. *Bioessays* **35**: 690-695.
- Frias L, Leles D, Araujo A. 2013. Studies on protozoa in ancient remains--a review. *Mem Inst Oswaldo Cruz* **108**: 1-12.
- Fu Q, Posth C, Hajdinjak M, et al. 2016. The genetic history of Ice Age Europe. *Nature* **534**: 200-205.
- Fujimori S, Hirai N, Ohashi H, et al. 2012. Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Scientific reports* **2**: 691-691.
- Gallego Llorente M, Jones ER, Eriksson A, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**: 820-822.
- Gibson DG, Benders GA, Andrews-Pfannkoch C, et al. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215-1220.
- Ginolhac A, Vilstrup J, Stenderup J, et al. 2012. Improving the performance of true single molecule sequencing for ancient DNA. *BMC Genomics* **13**.
- Giovas C, Lambrides AB, Fitzpatrick SM, Kataoka O. 2016. Reconstructing prehistoric fishing zones in Palau, Micronesia using fish remains: a blind test of inter-analyst correspondence. *Archaeology in Oceania*, doi:10.1002/arco.5119.
- Gismondi A, Di Marco G, Martini F, Sarti L, Crespan M, et al. 2016. Grapevine carpological remains revealed the existence of a Neolithic domesticated *Vitis*

vinifera L. specimen containing ancient DNA partially preserved in modern ecotypes. *Journal of Archaeological Science* **69**: 75-84.

Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, et al. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**: 523-527.

Gokhman D, Meshorer E, Carmel L. 2016. Epigenetics: It's Getting Old. Past Meets Future in Paleoepigenetics. *Trends in Ecology and Evolution* **31**: 290-300.

Greenleaf WJ and Block SM. 2006. Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science* **313**: 801-801.

Gutiérrez-García TA V-DE, Arroyo-Cabrales J, Kuch M, Enk J, King C, et al. 2014. Ancient DNA and the tropics: a rodent's tale. *Biology Letters* **10**: 20140224.

Guy PL. 2013. Ancient RNA? RT-PCR of 50-year-old RNA identifies peach latent mosaic viroid. *Archives of Virology* **158**: 691-694.

Guy PL. 2014. Prospects for analyzing ancient RNA in preserved materials. *Wiley Interdisciplinary Reviews-RNA* **5**: 87-94.

Hanna GJ, Johnson VA, Kuritzkes DR, Richman DD, Martinez-Picado J, et al. 2000. Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase. *Journal of Clinical Microbiology* **38**: 2715-2721.

Haouchar D, Pacioni C, Haile J, McDowell MC, Baynes A, et al. 2016. Ancient DNA reveals complexity in the evolutionary history and taxonomy of the endangered Australian brush-tailed bettongs (Bettongia: Marsupialia: Macropodidae: Potoroinae). *Biodiversity and Conservation*.

Harkins KM and Stone AC. 2015. Ancient pathogen genomics: insights into timing and adaptation. *Journal of Human Evolution* **79**: 137-149.

- Harper KN and Armelagos GJ. 2013. Genomics, the origins of agriculture, and our changing microbe-scape: time to revisit some old tales and tell some new ones. *American Journal of Physical Anthropology* **57**: 135-152.
- Harris TD, Buzby PR, Babcock H, et al. 2008. Single-molecule DNA sequencing of a viral genome. *Science* **320**: 106-109.
- Heintzman PD, Elias, Scott A, Moore K, Paszkiewicz K, Barnes I. 2014. Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera) *Molecular Ecology Resources* **14**.
- Hindson BJ, Ness KD, Masquelier DA, et al. 2011. High-Throughput Droplet Digital PCR System for Absolute Quantitation of DNA Copy Number. *Analytical Chemistry* **83**: 8604-8610.
- Hoffmann GS and Griebeler EM. 2013. An improved high yield method to obtain microsatellite genotypes from red deer antlers up to 200 years old. *Molecular Ecology Resources* **13**: 440-446.
- Hofman CA, Rick TC, Fleischer RC, Maldonado JE. 2015. Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology and Evolution* **30**: 540-549.
- Hoshino T and Inagaki F. 2012. Molecular quantification of environmental DNA using microfluidics and digital PCR. *Systematic and Applied Microbiology* **35**: 390-395.
- Howorka S, Cheley S, Bayley H. 2001. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature Biotechnology* **19**: 636-639.
- Huynen L, Millar CD, Lambert DM. 2012. Resurrecting ancient animal genomes: The extinct moa and more. *Bioessays* **34**: 661-669.

- Huynen L, Suzuki T, Ogura T, Watanabe Y, Millar CD, et al. 2014. Reconstruction and in vivo analysis of the extinct *tbx5* gene from ancient wingless moa (Aves: Dinornithiformes). *BMC Evolutionary Biology* **14**.
- Ju J, Kim DH, Bi L, et al. 2006. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 19635-19640.
- Kan CW, Fredlake CP, Doherty EAS, Barron AE. 2004. DNA sequencing and genotyping in miniaturized electrophoresis systems. *Electrophoresis* **25**: 3564-3588.
- Kato H, Anzai M, Mitani T, et al. 2009. Recovery of cell nuclei from 15,000 years old mammoth tissues and its injection into mouse enucleated matured oocytes. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences* **85**: 240-247.
- Kistler L RA, Godfrey LR, Crowley BE, Hughes CE, Lei R, et al. 2014. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *Journal of Human Evolution* **79**: 45-54.
- Korlach J, Bibillo A, Wegener J, Peluso P, Pham TT, et al. 2008. Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides Nucleotides & Nucleic Acids* **27**: 1072-1083.
- Krutli A, Bouwman A, Akgul G, Della Casa P, Ruhli F et al. 2014. Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central europe. *PLOS One* **9**: e86251.
- Lagerqvist J, Zwolak M, Di Ventra M. 2006. Fast DNA sequencing via transverse electronic transport. *Nano Letters* **6**: 779-782.

- Leonardi M, Librado P, Der Sarkissian C, Schubert M, Alfarhan AH, et al. 2016. Evolutionary Patterns and Processes: Lessons from Ancient DNA. *Systematic Biology*.
- Levene MJ, Korlach J, Turner SW, Foquet M, Craighead HG, et al. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**: 682-686.
- Lin C and Yao B. 2012. Recent Advance in Digital PCR. *Progress in Chemistry* **24**: 2415-2423.
- Liu L, Li Y, Siliang L, Hu N, et al. 2012. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology*.
- Llamas B, Holland ML, Chen K, Cropley JE, Cooper A, et al. 2012. High-Resolution Analysis of Cytosine Methylation in Ancient DNA. *PLOS One* **7**.
- Loi P, Wakayama T, Saragustry J, Fulka J Jr, Ptak G. 2011. Biological time machines: a realistic approach for cloning an extinct mammal. *Endangered Species Research* **14**: 227-233.
- Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, et al. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell reports* **12**: 217-228.
- Malaspinas AS. 2016. Methods to characterize selective sweeps using time serial samples: an ancient DNA perspective. *Molecular Ecology* **25**: 24-41.
- Marciniak S, Klunk J, Devault A, Enk J, Poinar HN. 2015. Ancient human genomics: the methodology behind reconstructing evolutionary pathways. *Journal of Human Evolution* **79**: 21-34.

- Margulies M, Jarvie TP, Knight JR, Simons JF. 2007. The 454 Life Sciences Picoliter Sequencing System. In: Mitchelson KR, editor. *New High Throughput Technologies for DNA Sequencing and Genomics*: 153-186.
- Matisoo-Smith E. 2015. Ancient DNA and the human settlement of the Pacific: A review. *Journal of Human Evolution* **79**: 93-104.
- Mendisco F, Pemonge MH, Leblay E, Romon T, Richard G, et al. 2015. Where are the Caribs? Ancient DNA from ceramic period human remains in the Lesser Antilles. *Philosophical Transactions of the Royal Society B* **370**.
- Mikic AM. 2015. The First Attested Extraction of Ancient DNA in Legumes (Fabaceae). *Frontiers in Plant Science* **6**.
- Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE et al. 2003. Digital genotyping and haplotyping with polymerase colonies. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 5926-5931.
- Monforte JA and Becker CH. 1997. High-throughput DNA analysis by time-of-flight mass spectrometry. *Nature Medicine* **3**: 360-362.
- Muller R, Roberts CA, Brown TA. 2014. Genotyping of ancient Mycobacterium tuberculosis strains reveals historic genetic diversity. *Proceedings of the Royal Society B-Biological Sciences* **281**.
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS One* **10**.
- Mutou C, Tanaka K, Ishikawa R. 2014. DNA Extraction from Rice Endosperm (Including a Protocol for Extraction of DNA from Ancient Seed Samples). In: Henry RJ, Furtado A, editors. *Cereal Genomics: Methods and Protocols*: 7-15.

- Nicholls H. 2008. Darwin 200: Let's make a mammoth. *Nature* **456**: 310-314.
- Oh CS, Seo M, Hong JH, Chai JY, Oh SW, et al. 2015. Ancient Mitochondrial DNA Analyses of *Ascaris* Eggs Discovered in Coprolites from Joseon Tomb. *Korean Journal of Parasitology* **53**: 237-242.
- Orlando L, Ginolhac A, Raghavan M, et al. 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research* **21**: 1705-1719.
- Ozsolak F and Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* **12**: 87-98.
- Paabo S. 2014. The human condition-a molecular approach. *Cell* **157**: 216-226.
- Paabo S. 2015. The diverse origins of the human gene pool. *Nature Reviews Genetics* **16**: 313-314.
- Pacioni C, Hunt H, Allentoft ME, Vaughan TG, Wayne AF, et al. 2015. Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular Ecology* **24**: 5813-5828.
- Pareek CS, Smoczynski R, Tretyn A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics* **52**: 413-435.
- Parks M, Subramanian S, Baroni C, Salvatore MC, Zhang G, et al. 2015. Ancient population genomics and the study of evolution. *Philosophical Transactions of the Royal Society B* **370**.
- Pask AJ, Behringer RR, Renfree MB. 2008. Resurrection of DNA Function *In Vivo* from an Extinct Genome. *PLOS One* **3**.

- Pedersen MW, Overballe-Petersen S, Ermini L, et al. 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B* **370**.
- Pina-Aguilar RE, Lopez-Saucedo J, Sheffield R, Ruiz-Galaz LI, Barroso-Padilla JdJ et al. 2009. Revival of Extinct Species Using Nuclear Transfer: Hope for the Mammoth, True for the Pyrenean Ibex, But Is It Time for "Conservation Cloning"? *Cloning and Stem Cells* **11**: 341-346.
- Porreca GJ. 2010. Genome sequencing on nanoballs. *Nature Biotechnology* **28**: 43-44.
- Prabhakar S, Visel A, Akiyama JA, et al. 2008. Human-specific gain of function in a developmental enhancer. *Science* **321**: 1346-1350.
- Rawlence NJ, Perry GLW, Smith IWG, Scofield RP, Tennyson AJD, et al. 2015. Radiocarbon-dating and ancient DNA reveal rapid replacement of extinct prehistoric penguins. *Quaternary Science Reviews* **112**: 59-65.
- Rick TC and Lockwood R. 2013. Integrating paleobiology, archeology, and history to inform biological conservation. *Conservation biology: the journal of the Society for Conservation Biology* **27**: 45-54.
- Rollo F. 1985. Characterization by molecular hybridization of RNA fragments isolated from ancient (1400 BC) seeds. *Theoretical and Applied Genetics* **71**: 330-333.
- Rollo F, Venanzi FM, Amici A. 1991. Nucleic-acids in mummified plant seeds--biochemistry and molecular-genetics of pre-columbian maize. *Genetic Resources* **58**: 193-201.
- Ronaghi M, Uhlen M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* **281**: 363-365.
- Rusk N. 2011. Torrents of sequence. *Nature Methods* **8**: 44-44.

- Sallon S, Solowey E, Cohen Y, Korchinsky R, Egli M, et al. 2008. Germination, genetics, and growth of an ancient date seed. *Science* **320**: 1464-1464.
- Sams AJ, Hawks J, Keinan A. 2015. The utility of ancient human DNA for improving allele age estimates, with implications for demographic models and tests of natural selection. *Journal of Human Evolution* **79**: 64-72.
- Sanchez-Quinto F and Lalueza-Fox C. 2015. Almost 20 years of Neanderthal palaeogenetics: adaptation, admixture, diversity, demography and extinction. *Philosophical Transactions of the Royal Society B* **370**.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.
- Sapir-Hen L, Meiri M, Finkelstein I. 2015. Iron age pigs: new evidence on their origin and role in forming identity boundaries. *Radiocarbon* **57**: 307-315.
- Sarkissian CD, Allentoft ME, Avila-Arcos MC, et al. 2015. Ancient genomics. *Philosophical Transactions of the Royal Society B* **370**.
- Schaefer NK, Shapiro B, Green RE. 2016. Detecting hybridization using ancient DNA. *Molecular Ecology* **25**: 2398-2412.
- Schroeder H, Avila-Arcos MC, Malaspinas AS, et al. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 3669-3673.
- Scorrano G, Valentini F, Martinez-Labarga C, et al. 2015. Methodological strategies to assess the degree of bone preservation for ancient DNA studies. *Annals of Human Biology* **42**: 10-19.

- Selvin PR. 2000. The renaissance of fluorescence resonance energy transfer. *Nature Structural Biology* **7**: 730-734.
- Shendure J, Porreca GJ, Reppas NB, et al. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728-1732.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**: 1794-1805.
- Slatkin M and Racimo F. 2016. Ancient DNA and human history. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 6380-6387.
- Smith O, Clapham A, Rose P, Liu Y, Wang J, et al. 2014a. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Scientific Reports* **4**: 4003.
- Smith O, Clapham A, Rose P, Liu Y, Wang J, et al. 2014b. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Scientific Reports* **4**: 5559.
- Soni GV and Meller A. 2007. Progress toward ultrafast DNA Sequencing using solid-state nanopores. *Clinical Chemistry* **53**: 1996-2001.
- Sosa C, Vispe E, Nunez C, Baeta M, Casalod Y, et al. 2013. Association Between Ancient Bone Preservation and DNA Yield: A Multidisciplinary Approach. *American Journal of Physical Anthropology* **151**: 102-109.
- Spigelman M, Donoghue HD, Abdeen Z, et al. 2015. Evolutionary changes in the genome of *Mycobacterium tuberculosis* and the human genome from 9000 years BP until modern times. *Tuberculosis (Edinb)* **95**: 145-149.

- Staats M, Erkens RH, van de Vossen B, Wieringa JJ, Kraaijeveld K, et al. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLOS One* **8**.
- Stoddart D, Heron AJ, Mikhailova E, Maglia G and Bayley H. 2009. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 7702-7707.
- Suzuki K, Saso A, Hoshino K, et al. 2014. Paleopathological Evidence and Detection of *Mycobacterium leprae* DNA from Archaeological Skeletal Remains of Nabe-kaburi (Head-Covered with Iron Pots) Burials in Japan. *PLOS One* **9**.
- Tanaka H and Kawai T. 2003. Visualization of detailed structures within DNA. *Surface Science* **539**: 531-536.
- Tanaka H and Kawai T. 2009. Partial sequencing of a single DNA molecule with a scanning tunnelling microscope. *Nature Nanotechnology* **4**: 518-522.
- Tao NJ, Derose JA, Lindsay SM. 1993. Self-assembly of molecular superstructures studied by insitu scanning tunneling microscopy DNA bases on Au(111). *Journal of Physical Chemistry* **97**: 910-919.
- Theves C, Biagini P, Crubezy E. 2014. The rediscovery of smallpox. *Clinical Microbiology and Infection* **20**: 210-218.
- Thompson JF and Steinmann KE. 2010. Single molecule sequencing with a HeliScope genetic analysis system. *Current protocols in molecular biology*: 7-10.
- Thomsen PF and Willerslev E. 2015. Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**: 4-18.

- Thomson VA, Lebrasseur O, Austin JJ, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4826-4831.
- Timp W, Mirsaidov UM, Wang D, Comer J, Aksimentiev A, et al. 2010. Nanopore Sequencing: Electrical Measurements of the Code of Life. *Ieee Transactions on Nanotechnology* **9**: 281-29.
- Villanea FA, Parent CE, Kemp BM. 2016. Reviving Galapagos snails: ancient DNA extraction and amplification from shells of probably extinct endemic land snails. *Journal of Molluscan Studies* **82**: 449-456.
- Wakayama S, Mizutani E, Wakayama T. 2011. Nuclear Transfer Embryonic Stem Cells as a New Tool for Basic Biology. *Stem Cell Biology and Regenerative Medicine*.
- Wakayama S, Ohta H, Hikichi T, Mizutani E, Iwaki T, et al. 2008. Production of healthy cloned mice from bodies frozen at -20 degrees C for 16 years. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 17318-17322.
- Wallace HAC, Marques-Kranc F, Richardson M, Luna-Crespo F, Sharpe JA, et al. 2007. Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**: 197-209.
- Warinner C, Rodrigues JF, Vyas R, et al. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* **46**: 336-344.
- Warinner C, Speller C, Collins MJ. 2015a. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society B* **370**.

- Warinner C, Speller C, Collins MJ, Lewis CM. 2015b. Ancient human microbiomes. *Journal of Human Evolution* **79**: 125-136
- Weyrich LS, Dobney K, Cooper A. 2015. Ancient DNA analysis of dental calculus. *Journal of Human Evolution* **79**: 119-124.
- White RA, III, Blainey PC, Fan HC and Quake SR. 2009. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10**.
- Willerslev E and Cooper A. 2005. Ancient DNA. *Proceedings of the Royal Society B-Biological Sciences* **272**: 3-16.
- Xu C, Dong WP, Shi S, et al. 2015. Accelerating plant DNA barcode reference library construction using herbarium specimens: improved experimental techniques. *Molecular Ecology Resources* **15**: 1366-1374.
- Xu M, Fujita D, Hanagata N. 2009. Perspectives and Challenges of Emerging Single-Molecule DNA Sequencing Technologies. *Small* **5**: 2638-2649.
- Yoshida K, Schuenemann VJ, Cano LM, et al. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**.
- Zhang G, Shoham D, Gilichinsky D, Davydov S, Castello JD et al. 2006. Evidence of influenza A virus RNA in Siberian lake ice. *Journal of Virology* **80**: 12229-12235.
- Ziesemer KA, Mann AE, Sankaranarayanan K, et al. 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* **5**.

## 7.4 SUPPLEMENTARY INFORMATION

**TABLE S7.4.1** | SUMMARY OF THE TYPES OF ADNA STUDIES UNDERTAKEN BETWEEN 2013 AND 2016. Studies marked by ^ indicate that the DNA was extracted from a warm-climate ecosystem.

Topic of investigation	Organism	Substrate	DNA	Sequence	Reference
Population history, conservation	Upland seal ( <i>Arctocephalus fosteri</i> )	Bone	mtDNA	Amplicon	(Salis et al. 2016)
	Eurasian lynx ( <i>Lynx pardinus</i> )	Bone	mtDNA	Amplicon	(Rodriguez-Varela et al. 2016)
	New Zealand little penguins ( <i>Eudyptula novaehollandiae</i> )	Bone	mtDNA	Amplicon	(Grosser et al. 2016)
	Iberian lynx ( <i>Lynx pardinus pardinus</i> )	Bone, Teeth	mtDNA	Amplicon	(Rodriguez-Varela et al. 2015)
	Bettongs ( <i>Bettongia penicillata ogilbyi</i> )	Bone, Skin	mtDNA	Amplicon	(Pacioni et al. 2015; Haouchar 2016)^

	Penguin ( <i>Megadyptes</i> )	Bone	mtDNA	Amplicon	(Rawlence et al. 2015)
Population history, human/archaeology	Pre-Columbian Peruvian ( <i>Homo sapiens</i> )	Bone	mtDNA	Whole genome	(Valverde et al. 2016)
	Red deer ( <i>Cervus elaphus</i> )	Bone	mtDNA	Amplicon	(Stanton et al. 2016)
	Altai Mountains (Hominin)	Bone	mtDNA	Amplicon	(Gubina et al. 2016)
	Altai Mountains (Hominin)	Teeth	mtDNA nuDNA	Amplicon	(Hollard et al. 2014)
	Altai Mountains (Denisovan)	Bone	Whole genome	Shotgun	(Prufer et al. 2014)
	<i>Homo sapiens</i>	Petrous bone	mtDNA nuDNA	Shotgun	(Pinhasi et al. 2015)^
	Atapuerca, Spain ( <i>Homo sapiens</i> )	Teeth	mtDNA	Amplicon	(Gomez-Sanchez et al. 2014)
	Atapuerca, Spain ( <i>Homo sapiens</i> )	Bone, Teeth	mtDNA nuDNA	Shotgun	(Gunther et al. 2015)
	Spain ( <i>Homo sapiens</i> )	Teeth	Whole genome	Shotgun/capture	(Olalde et al. 2015)

Moa ( <i>Euryapteryx</i> spp., <i>Emeus</i> spp., <i>Dinornis</i> sp.)	Eggshell	mtDNA nuDNA	Amplicon	(Jacomb et al. 2014)
Neolithic/Bronze age ( <i>Homo sapiens</i> )	Bone	mtDNA nuDNA	Whole genome	(Allentoft et al. 2015; Cassidy et al. 2016)
Neolithic/Bronze age ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Brandt et al. 2013)
Northern Han Chinese ( <i>Homo sapiens</i> )	Bone, Tooth	mtDNA nuDNA	Amplicon	(Zhao et al. 2015)
Near-Eastern Farmers ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Fernandez et al. 2014)
Korea ( <i>Homo sapiens</i> )	Bone	mtDNA nuDNA	Amplicon	(Jin et al. 2015)
Russia ( <i>Homo sapiens</i> )	Teeth	Whole genome	Shotgun	(Seguin-Orlando et al. 2014)
Di-Qiang populations ( <i>Homo sapiens</i> )	Teeth	mtDNA nuDNA	Amplicon	(Gao et al. 2015)
Shandong Province, China ( <i>Homo sapiens</i> )	Bone	mtDNA nuDNA	Amplicon	(Dong et al. 2015)

Eurasians ( <i>Homo sapiens</i> )	Bone	mtDNA nuDNA	Shotgun	(Mathieson et al. 2015)
Viking age Norwegians ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Krzewinska et al. 2015)
Cantabria, Basque Country ( <i>Homo sapiens</i> )	Teeth	mtDNA	Amplicon	(de-la-Rua et al. 2015)
Poland ( <i>Homo sapiens</i> )	Bone	mtDNA	Amplicon	(Juras et al. 2014)
North Britain ( <i>Homo sapiens</i> )	Bone	Whole genome	Shotgun	(Martiniano et al. 2016)
Sima de los Huesos (Hominin)	Bone, teeth	mtDNA nuDNA	Shotgun	(Meyer M 2014; Meyer et al. 2016)
Hunter-Gatherers from Patagonia ( <i>Homo sapiens</i> )	Bone, teeth	mtDNA nuDNA	Amplicon	(de la Fuente et al. 2015)
Neolithic farmers ( <i>Homo sapiens</i> )	Bone, Teeth	mtDNA	Shotgun, Capture	(Brotherton et al. 2013; Malmstrom et al. 2015)
Jau Dignac et Loirac, France	Teeth	mtDNA nuDNA	Amplicon	(Deguilloux et al. 2014)

<i>(Homo sapiens)</i>					
Tarim Basin, China <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Li et al. 2015)	
Yuan Dynasty, China <i>(Homo sapiens)</i>	Teeth	mtDNA nuDNA	Amplicon	(Cui et al. 2015)	
Neolithic, Copper, Bronze, Iron Age <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Gamba et al. 2014)	
Palaeolithic, Mesolithic Europeans <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Jones et al. 2015)	
Yakuts <i>(Homo sapiens)</i>	Bone	nuDNA	Amplicon	(Keyser et al. 2015)	
Ethiopia <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Gallego Llorente et al. 2015)^	
Iron age Anglo-Saxon <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Schiffels et al. 2016)	
Western Siberia <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Fu et al. 2014)	
Upper Palaeolithic Sibera	Bone	Whole genome	Shotgun	(Raghavan et al. 2014b)	

*(Homo sapiens)*

Syria <i>(Homo sapiens)</i>	Teeth	mtDNA nuDNA	Amplicon	(Witas et al. 2013)
Late Pleistocene Native American <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Rasmussen et al. 2014)
Kennewick man (Native American) <i>(Homo sapiens)</i>	Bone	Whole genome	Shotgun	(Rasmussen et al. 2015a)
Paleo-Inuit <i>(Homo sapiens)</i>	Bone, Teeth, Hair	Whole genome	Shotgun	(Raghavan et al. 2014a)
Human <i>(Homo sapiens)</i>	Bone	mtDNA	Shotgun	(Fu et al. 2013)
Mesolithic North East Europe <i>(Homo sapiens)</i>	Teeth	mtDNA	Shotgun/capture	(Der Sarkissian et al. 2014)
Prehispanic Argentina <i>(Homo sapiens)</i>	Bone, Teeth	mtDNA	Amplicon	(Mendisco et al. 2015)
Neolithic farmers <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Lee et al. 2014)
Egypt	Mummy tissue	mtDNA	Shotgun	(Khairat et al. 2013)^

<i>(Homo sapiens)</i>			nuDNA		
Neolithic Europeans <i>(Homo sapiens)</i>	Bone		mtDNA nuDNA	Shotgun	(Hofmanova et al. 2016)
Europeans <i>(Homo sapiens)</i>	Bone		mtDNA nuDNA	Shotgun/capture	(Haak et al. 2015)
Mesolithic foragers, Neolithic farmers <i>(Homo sapiens)</i>	Teeth		mtDNA	Amplicon	(Bollongino et al. 2013)
Botocudos, Brazil <i>(Homo sapiens)</i>	Teeth		mtDNA	Amplicon	(Goncalves et al. 2013)
Botocudos, Brazil <i>(Homo sapiens)</i>	Bone		Whole genome	Shotgun	(Malaspina et al. 2014)
Prepastoralist Southern African <i>(Homo sapiens)</i>	Bone		mtDNA	Whole genome	(Morris et al. 2014)^
Altamura, Italy <i>(Homo neanderthalensis)</i>	Bone		mtDNA	Amplicon	(Lari et al. 2015)
German farmer <i>(Homo sapiens)</i>	Bone		Whole genome	Shotgun	(Lazaridis et al. 2014)
Norwegian Slaves	Bone		mtDNA	Amplicon	(Naumann et al. 2014)

*(Homo sapiens)*

Romania <i>(Homo sp.)</i>	Bone	mtDNA nuDNA	Shotgun/capture	(Fu et al. 2015)
Early Neolithic farmers <i>(Homo sapiens)</i>	Teeth	mtDNA	Amplicon	(Hervella et al. 2015)
Irish medieval <i>(Homo sapiens)</i>	Bone, Teeth	nuDNA	Amplicon	(Tierney and Bird 2015)
North America <i>(Homo sapiens)</i>	Teeth	Whole genome	Shotgun	(Cui et al. 2013)
Caribbean enslaved Africans <i>(Homo sapiens)</i>	Bone, Teeth	mtDNA nuDNA	Shotgun	(Schroeder et al. 2015)^
Schild cemetery, Illinois <i>(Homo sapiens)</i>	Bone	mtDNA	Amplicon	(Reynolds et al. 2015)
Lesser Antilles <i>(Homo sapiens)</i>	Bone, Teeth	mtDNA		(Mendisco et al. 2015)^
Santa Cruz, Argentina <i>(Homo sapiens)</i>	Bone	mtDNA	Whole genome	(Motti et al. 2015)
Pacific region <i>(Homo sapiens)</i>	<i>Review</i>			(Matisoo-Smith 2015)^

	Neanderthal and Denisovan ( <i>Homo spp.</i> )	Bone	Whole genome	Shotgun/capture	(Kuhlwilm et al. 2016)
	Denisovans ( <i>Homo</i> )	Bone, Teeth	mtDNA nuDNA	Shotgun	(Sawyer et al. 2015)
Population history, domestication	Dromedary ( <i>Camelus dromedarius</i> )	Bone, teeth	mtDNA nuDNA	Amplicon	(Almathen et al. 2016)^
	Wild boar ( <i>Sus scrofa</i> )	Teeth	mtDNA	Amplicon	(Hou et al. 2014b; Evin et al. 2015)
	Pig ( <i>Sus</i> )	Bone, teeth	mtDNA nuDNA	Amplicon	(Krause-Kyora et al. 2013; Ottoni et al. 2013)
	Chicken ( <i>Gallus gallus</i> )	Bone	mtDNA	Amplicon	(Thomson et al. 2014; Xiang et al. 2014)
	Chicken ( <i>Gallus gallus</i> )	Bone	mtDNA nuDNA	Amplicon	(Flink et al. 2014)
	Iron age, medieval, post-medieval sheep ( <i>Ovis aries</i> )	Bone	mtDNA nuDNA	Amplicon	(Niemi et al. 2013)
	Horse	Bone	Whole genome	Shotgun	(Schubert et al. 2014)
	Domestic cattle	Bone	mtDNA	Shotgun/capture	(Zhang et al. 2013)

	Wild aurochs ( <i>Bos primigenius</i> )	Bone	Whole genome	Shotgun	(Park et al. 2015)
	Wolf	Bone	Whole genome	Shotgun	(Skoglund et al. 2015)
	Ancient canids	Bone	Whole mt genome	Shotgun	(Thalmann et al. 2013)
	Dog ( <i>Canis variabilis</i> )	Bone, teeth	mtDNA	Amplicon	(Lee et al. 2015)
Population history, phylogeography, evolutionary processes	Red wolf ( <i>Canis rufus</i> )	Bone/teeth	mtDNA	Amplicon	(Brzeski et al. 2016)
	Macropods ( <i>Protemnodon</i> , <i>Simosthenurus</i> )	Bone	mtDNA	Shotgun	(Llamas et al. 2015)
	Yakutian horses	Bone, Teeth	Whole genome	Shotgun	(Librado et al. 2015)
	Norwegian lemming ( <i>Lemmus lemmus</i> )	Bone	mtDNA	Amplicon	(Lagerholm et al. 2014)
	Lion ( <i>Panthera leo</i> )	Bone, Tissue	mtDNA	Amplicon	(Barnett et al. 2014)
	Alpine chipmunks ( <i>Tamias alpinus</i> )	Museum skins	nuDNA	Shotgun/capture	(Bi et al. 2013)
	Woolly mammoth ( <i>Mammuthus primigenius</i> )	Bone, Tissue	Whole genome	Shotgun	(Palkopoulou et al. 2015)

	Moa ( <i>Dinornithiformes</i> )	Bone	mtDNA nuDNA	Amplicon	(Allentoft 2014)
	Roe deer ( <i>Capreolus careolus</i> )	Bone	mtDNA	Amplicon	(Baker and Hoelzel 2014)
	Red deer ( <i>Cervus elaphus L.</i> )	Bone	mtDNA	Amplicon	(Meiri et al. 2013)
	Passenger pigeon ( <i>Ectopistes migratorius</i> )	Tissue	mtDNA nuDNA	Shotgun	(Hung et al. 2014)
	Honey bees	Museum specimens	mtDNA nuDNA	Shotgun	(Mikheyev et al. 2015)
	Capromyidae	Museum specimens	mtDNA nuDNA	Shotgun/capture	(Fabre et al. 2014)^
	Pike ( <i>Esox Lucius</i> )	Bone	mtDNA	Amplicon	(Wooller et al. 2015)
Phylogeny, time of divergence	Chatham duck ( <i>Pachyanas chathamica</i> )	Bone	mtDNA	Whole genome	(Mitchell <i>et al.</i> 2014)
	Elephant bird ( <i>Aepyornis sp. Mullerornis sp.</i> )	Bone	mtDNA	Whole genome	(Mitchell 2014)^
	Moa	Bone	nuDNA	Shotgun	(Baker 2014)

	(Dinornithiformes)				
	Horse ( <i>Equus</i> spp.)	Bone	Whole genome	Shotgun	(Orlando et al. 2013)
	Horse ( <i>Equus</i> spp.)	Bone, Tissue	mtDNA	Shotgun/capture	(Vilstrup et al. 2013)
	Primates (Guenons)	Museum specimens	mtDNA	Shotgun/capture	(Guschanski et al. 2013)^
	Cave hyaenas ( <i>Crocota crocuta</i> )	Bone	mtDNA	Amplicon	(Sheng G-L 2014)
	<i>Palaeopropithecus ignens</i>	Bone	mtDNA	Shotgun	(Kistler L 2014)^
	<i>Otomyomys phyllotis</i>	Bone	mtDNA	Amplicon	(Gutiérrez-García TA 2014)^
	Woolly rhinoceros ( <i>Coelodonta antiquitatis</i> )	Bone	mtDNA	Amplicon	(Yuan et al. 2014)
Taxonomy	Chatham duck ( <i>Pachyanas chathamica</i> )	Bone	mtDNA	Whole genome	(Mitchell et al. 2014)
	Moa ( <i>Euryapteryx</i> spp.)	Bone	mtDNA	Amplicon	(Huynen 2014)
Palaeodiet,	Laguna Potrok Aike,	Sediment	mtDNA	Amplicon	(Vuillemin et al. 2016)

palaeoenvironment	Argentina				
	Norwegian Lakes <i>Planktothrix</i>	Sediment	Cyanobacterial DNA	Amplicon	(Kyle et al. 2015)
	Vegetation	Peat sediment	chDNA	Amplicon	(Parducci et al. 2015)
	Vascular plants	Sediment			(Epp LS 2015)
	Chum salmon ( <i>Oncorhynchus keta</i> )	Bone	mtDNA	Amplicon	(Halffman et al. 2015)
	Afro-alpine plants	Sediment	chDNA	Amplicon	(Boessenkool et al. 2014)^
	Methane-oxidising bacteria	Sediment	Bacterial DNA	Amplicon	(Belle et al. 2015)
	Cyanobacterial blooms ( <i>Cylindrospermopsis raciborskii</i> )	Sediment	Bacterial DNA	Amplicon	(De La Escalera et al. 2014)
	Photosynthetic plankton	Sediment	nuDNA	Amplicon	(Hou et al. 2014a)
	Foraminifera	Sediment		Amplicon	(Pawlowska et al. 2014)
	Arctic vegetation	Sediment	chDNA	Amplicon	(Willerslev et al. 2014)
	Lake Skartjorna,	Sediment	chDNA	Amplicon	(Alsos et al. 2016)

Svalbard					
Agriculture	Rice ( <i>Oryza sativa</i> )	Seeds	chDNA	Amplicon	(Tanaka et al. 2016)
	Grapevine ( <i>Vitis vinifera</i> )	Carpological remains	chDNA, nuDNA	Amplicon	(Gismondi et al. 2016)
	Catalhoyuk wheat ( <i>Triticum</i> spp.)	Wheat grains	nuDNA	Amplicon	(Bilgic et al. 2016)
	<i>Cyperus</i> sp.		chDNA	Amplicon	(Palla et al. 2013)
	Wheat ( <i>Triticum</i> sp.)	Charred grains	nuDNA	Amplicon	(Fernandez et al. 2013)
	Barley		nuDNA	Kompetitive Allele Specific PCR	(Lister et al. 2013)
	Barley Stripe Mosaic Virus	Barley grain	Whole genome (RNA)	Shotgun	(Smith et al. 2014a)
	Wood ( <i>Populus euphratica</i> )	Wood	chDNA	Amplicon	(Jiao et al. 2015)
	Legumes (Fabaceae)	Seeds	chDNA nuDNA	Shotgun Amplicon	(Mikic 2015)
	Grass weed ( <i>Alopecurus myosuroides</i> )	Museum specimens	nuDNA	Amplicon	(Delye et al. 2013)
	Maize	Cobs	nuDNA	Shotgun	(da Fonseca et al. 2015)

	<i>(Zea mays mays)</i>				
	<i>Phytophthora infestans</i>	Potato and tomato herbarium specimens	Whole genome	Shotgun	(Martin et al. 2013; Yoshida et al. 2013; Martin et al. 2014)
	Wheat <i>(Triticum spp.)</i>	Sediment	Shotgun	Shotgun	(Smith et al. 2015)
	Barley	Grain	Whole genome	Shotgun	(Mascher et al. 2016)
Historical disease and human health	Plague <i>(Yersinia pestis)</i>	Teeth	Whole genome	Shotgun	(Wagner et al. 2014; Rasmussen et al. 2015b; Bos et al. 2016)
	Plague <i>(Yersinia pestis)</i>	Teeth	Bacterial DNA	Amplicon	(Harbeck et al. 2013)
	<i>Mycobacterium leprae</i>	Human bone	Whole genome Bacterial SNPs	Whole genome, Amplicon	(Economou et al. 2013; Schuenemann et al. 2013; Taylor et al. 2013; Mendum et al. 2014; Suzuki et al. 2014; Donoghue et al. 2015; Inskip et al. 2015)
	<i>Clonorchis sinensis</i>	Eggs from a human mummy	mtDNA nuDNA	Amplicon	(Shin et al. 2013)

<i>Mycobacterium tuberculosis</i>	Human bone	Bacterial SNPs	Amplicon	(Masson et al. 2013; Muller et al. 2014b, a; Harkins et al. 2015)
<i>Mycobacterium tuberculosis</i>	Human bone	Whole genome	Shotgun/capture	(Bos et al. 2014)
<i>Mycobacterium tuberculosis</i>	Mummy tissue	Whole genome	Shotgun	(Chan et al. 2013; Kay et al. 2015)
<i>Mycobacterium tuberculosis</i> , Malaria ( <i>Plasmodium falciparum</i> )	Mummy tissue	Bacterial DNA, Protozoa DNA	Amplicon	(Lalremruata et al. 2013)^
<i>Vibrio cholerae</i>	Human tissue	Whole genome	Shotgun/capture	(Devault et al. 2014a)
<i>Helicobacter pylori</i>	Mummy tissue	Whole genome		(Maixner et al. 2016)
Parasite eggs ( <i>Ascaris</i> , <i>Trichuris</i> , <i>Fasciola</i> )	Environmental samples	mtDNA	Amplicon	(Soe et al. 2015)
Koala retrovirus	Museum skins	mtDNA	Amplicon	(Avila-Arcos et al. 2013)
Gut virome	Coprolites	Viral DNA	Shotgun	(Appelt et al. 2014; Rivera-

					Perez et al. 2015)
	Gut microbiome	Coprolites	Bacterial DNA	Amplicon	(Cano et al. 2014)
	Pathogens	<i>Review</i>			(Darling and Donoghue 2014)
	Pathogens	Human bone	Bacterial DNA	Lawrence Livermore Microbial Detection Array	(Devault et al. 2014b)
	<i>Acaris</i>	Coprolites	mtDNA	Amplicon	(Oh et al. 2015)
	Virus	Caribou feces	Viral DNA	Amplicon	(Ng et al. 2014)
	<i>Treponema pallidum pallidum</i>	Bone	Bacterial DNA	Amplicon	(Gaul et al. 2015)
	Oral microbiome	Dental calculus	Bacterial DNA	Shotgun Amplicon	(Adler et al. 2013; De La Fuente et al. 2013; Warinner et al. 2014; Warinner et al. 2015a; Warinner et al. 2015b; Ziesemer et al. 2015)
Novel substrates	Galapagos snails ( <i>Naesiotus</i> )	Mollusk shell	mtDNA	Amplicon	(Villanea et al. 2016)
	Rice ( <i>Oryza sativa</i> )	Seeds	chDNA	Amplicon	(Tanaka et al. 2016)
	Horse, goat	Fur	mtDNA	Amplicon	(Sinding et al. 2015)

Phenotype, adaptation	Sheep ( <i>Ovis</i> )	Parchment	mtDNA	Whole genome	(Teasdale et al. 2015)
	Human ( <i>Homo sapiens</i> )	Bone	nuDNA	Amplicon	(Draus-Barini et al. 2013; Wilde et al. 2014)
	Human ( <i>Homo sapiens</i> )	Teeth	mtDNA nuDNA	Amplicon	(Krutli et al. 2014)
	Human ( <i>Homo sapiens</i> )	Gourd	Whole genome	Shotgun/capture	(Olalde et al. 2014b)
	Mesolithic European ( <i>Homo sapiens</i> )	Tooth	Whole genome	Shotgun	(Olalde et al. 2014a)
	Moa ( <i>Dinornithiformes</i> )	Bone	nuDNA	Shotgun Amplicon	(Huynen et al. 2014)
	Neanderthals ( <i>Homo neanderthalensis</i> )	Bone	nuDNA	Shotgun/capture	(Castellano et al. 2014)
	Woolly mammoth		Whole genome		(Lynch et al. 2015)
	Tasmanian devil ( <i>Sarcophilus harrisii</i> )	Bone, Tissue museum specimens	nuDNA	Amplicon	(Morris et al. 2013)
	Horse	Bone	nuDNA	Shotgun	(Bellone et al. 2013)

	Horse	Bone, Teeth	nuDNA	Amplicon	(Ludwig et al. 2015)
	Mediaeval cattle	Bone	nuDNA	Amplicon	(Svensson et al. 2014)
Epigenetics and RNA	Native American ( <i>Homo sapiens</i> )	Bone	5mC	Bisulfite	(Smith and Bolnick 2013; Smith 2015)
	Paleo-Inuit ( <i>Homo sapiens</i> )	Hair shafts	5mC	Amplification with two different polymerases	(Pedersen et al. 2014)
	Neanderthal and Denisovan ( <i>Homo spp.</i> )	Previous data	C→T	NA	(Gokhman et al. 2014)
	Koala retrovirus	Museum specimen	Viral DNA	Shotgun/capture	(Tsangaras et al. 2014)
	Barley ( <i>Hordeum vulgare</i> )	Barley seeds	5mC	Bisulfite	(Smith et al. 2014b)^

### S7.6.1 SUPPLEMENTARY REFERENCES

- Adler CJ, Dobney K, Weyrich LS, et al. 2013. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics* **45**: 450-455.
- Allentoft ME, Sikora M, Sjogren KG, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* **522**: 167-172.
- Allentoft ME H, Rasmus; Oskam C, L Lorenzen, Eline D, et al. 2014. Extinct New Zealand megafauna were not in decline before human colonization. *Proceedings of the National Academy of Sciences* **111**: 4922-4927.
- Almathen F, Charruau P, Mohandesan E, et al. 2016. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 6707-6712.
- Alsos IG, Sjogren P, Edwards ME, et al. 2016. Sedimentary ancient DNA from Lake Skartjorna, Svalbard: Assessing the resilience of arctic flora to Holocene climate change. *Holocene* **26**: 627-642.
- Appelt S, Fancello L, Le Bailly M, Raoult D, Drancourt M, et al. 2014. Viruses in a 14th-century coprolite. *Applied Environmental Microbiology* **80**: 2648-2655.
- Avila-Arcos MC, Ho SYW, Ishida Y, et al. 2013. One Hundred Twenty Years of Koala Retrovirus Evolution Determined from Museum Skins. *Molecular Biology and Evolution* **30**: 299-304.
- Baker AJ, Haddrath O, McPherson JD, Cloutier A. 2014. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Molecular Biology and Evolution* **31**: 1686-1696.

- Baker KH and Hoelzel AR. 2014. Influence of Holocene environmental change and anthropogenic impact on the diversity and distribution of roe deer. *Heredity* **112**: 607-615.
- Barnett R, Yamaguchi N, Shapiro B, Ho SYW, Barnes I, et al. 2014. Revealing the maternal demographic history of *Panthera leo* using ancient DNA and a spatially explicit genealogical analysis. *BMC Evolutionary Biology* **14**.
- Belle S, Verneaux V, Millet L, Parent C, Magny M. 2015. A case study of the past CH<sub>4</sub> cycle in lakes by the combined use of dual isotopes (carbon and hydrogen) and ancient DNA of methane-oxidizing bacteria: rearing experiment and application to Lake Remoray (eastern France). *Aquatic Ecology* **49**: 279-291.
- Bellone RR, Holl H, Setaluri V, et al. 2013. Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *PLOS One* **8**.
- Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R and Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* **22**: 6018-6032.
- Bilgic H, Hakki EE, Pandey A, Khan MK, Akkaya MS. 2016. Ancient DNA from 8400 Year-Old Catalhoyuk Wheat: Implications for the Origin of Neolithic Agriculture. *Plos One* **11**.
- Boessenkool S, McGlynn G, Epp LS, Taylor D, Pimentel M, et al. 2014. Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. *Conservation Biology* **28**: 446-455.
- Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, et al. 2013. 2000 years of parallel societies in Stone Age Central Europe. *Science* **342**: 479-481.

- Bos KI, Harkins KM, Herbig A, et al. 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**: 494-497.
- Bos KI, Herbig A, Sahl J, et al. 2016. Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* **5**.
- Brandt G, Haak W, Adler CJ, et al. 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**: 257-261.
- Brotherton P, Haak W, Templeton J, et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nature communications* **4**: 1764.
- Brzeski KE, DeBiase MB, Rabon DR, Chamberlain MJ, Taylor SS. 2016. Mitochondrial DNA Variation in Southeastern Pre-Columbian Canids. *Journal of Heredity* **107**: 287-293.
- Cano RJ, Rivera-Perez J, Toranzos GA, Santiago-Rodriguez TM, Narganes-Storde YM, et al. 2014. Paleomicrobiology: Revealing Fecal Microbiomes of Ancient Indigenous Cultures. *PLOS One* **9**.
- Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, et al. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 368-373.
- Castellano S, Parra G, Sanchez-Quinto FA, et al. 2014. Patterns of coding variation in the complete exomes of three Neandertals. *Proceedings of the National Academy of Sciences USA* **111**: 6666-6671.

- Chan JZ, Sergeant MJ, Lee OY, Minnikin DE, Besra GS, et al. 2013. Metagenomic analysis of tuberculosis in a mummy. *The New England journal of medicine* **369**: 289-290.
- Cui Y, Lindo J, Hughes CE, et al. 2013. Ancient DNA analysis of mid-holocene individuals from the Northwest Coast of North America reveals different evolutionary paths for mitogenomes. *PLOS One* **8**: e66948.
- Cui Y, Song L, Wei D, et al. 2015. Identification of kinship and occupant status in Mongolian noble burials of the Yuan Dynasty through a multidisciplinary approach. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **370**.
- da Fonseca RR, Smith BD, Wales N, et al. 2015. The origin and evolution of maize in the Southwestern United States. *Nature plants* **1**: 14003.
- Darling MI and Donoghue HD. 2014. Insights from paleomicrobiology into the indigenous peoples of pre-colonial America - A Review. *Memorias Do Instituto Oswaldo Cruz* **109**: 131-139.
- De La Escalera GM, Antoniades D, Bonilla S, Piccini C. 2014. Application of ancient DNA to the reconstruction of past microbial assemblages and for the detection of toxic cyanobacteria in subtropical freshwater ecosystems. *Molecular Ecology* **23**: 5791-5802.
- De La Fuente C, Flores S, Moraga M. 2013. DNA from human ancient bacteria: a novel source of genetic evidence from archaeological dental calculus. *Archaeometry* **55**.
- de la Fuente C, Galimany J, Kemp BM, Judd K, Reyes O, et al. 2015. Ancient Marine Hunter-Gatherers From Patagonia and Tierra Del Fuego: Diversity and Differentiation Using Uniparentally Inherited Genetic Markers. *American Journal of Physical Anthropology* **158**: 719-729.

- de-la-Rua C, Izagirre N, Alonso S, Hervella M. 2015. Ancient DNA in the Cantabrian fringe populations: A mtDNA study from Prehistory to Late Antiquity. *Quaternary International* **364**: 306-311.
- Deguilloux MF, Pemonge MH, Mendisco F, Thibon D, Cartron I, et al. 2014. Ancient DNA and kinship analysis of human remains deposited in Merovingian necropolis sarcophagi (Jau Dignac et Loirac, France, 7th-8th century AD). *Journal of Archaeological Science* **41**: 399-405.
- Delye C, Deulvot C, Chauvel B. 2013. DNA analysis of herbarium Specimens of the grass weed *Alopecurus myosuroides* reveals herbicide resistance pre-dated herbicides. *PLOS One* **8**: e75117.
- Der Sarkissian C, Brotherton P, Balanovsky O, et al. 2014. Mitochondrial genome sequencing in Mesolithic North East Europe Unearths a new sub-clade within the broadly distributed human haplogroup C1. *PLOS One* **9**: e87612.
- Devault AM, Golding GB, Waglechner N, et al. 2014a. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *The New England Journal of Medicine* **370**: 334-340.
- Devault AM, McLoughlin K, Jaing C, et al. 2014b. Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. *Scientific Reports* **4**.
- Dong Y, Li CX, Luan FS, Li ZG, Li HJ, et al. 2015. Low Mitochondrial DNA Diversity in an Ancient Population from China: Insight into Social Organization at the Fujia Site. *Human Biology* **87**: 71-84.
- Donoghue HD, Michael Taylor G, Marcsik A, et al. 2015. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **31**: 250-256.

- Draus-Barini J, Walsh S, Pospiech E, Kupiec T, Glab H, et al. 2013. Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative genetics* **4**: 3-3.
- Economou C, Kjellstrom A, Liden K, Panagopoulos I. 2013. Ancient-DNA reveals an Asian type of *Mycobacterium leprae* in medieval Scandinavia. *Journal of Archaeological Science* **40**: 465-470.
- Epp LS GG, Boessenkool S, Olsen J, Haile J, Schroder-Nielsen A, et al. 2015. Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews* **117**: 152-163.
- Evin A, Flink LG, Balasescu A, et al. 2015. Unravelling the complexity of domestication: a case study using morphometrics and ancient DNA analyses of archaeological pigs from Romania. *Philosophical Transactions of the Royal Society B* **370**.
- Fabre PH, Vilstrup JT, Raghavan M, Sarkissian CD, Willerslev E, et al. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. *Biology Letters* **10**.
- Fernandez E, Perez-Perez A, Gamba C, Prats E, Cuesta P, et al. 2014. Ancient DNA Analysis of 8000 B.C. Near Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLOS Genetics* **10**.
- Fernandez E, Thaw S, Brown TA, Arroyo-Pardo E, Buxo R, et al. 2013. DNA analysis in charred grains of naked wheat from several archaeological sites in Spain. *Journal of Archaeological Science* **40**: 659-670.
- Flink LG, Allen R, Barnett R, Malmstrom H, Peters J, Eriksson J, Andersson L, Dobney K and Larson G. 2014. Establishing the validity of domestication

- genes using DNA from ancient chickens. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 6184-6189.
- Fu Q, Hajdinjak M, Moldovan OT, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**: 216-219.
- Fu Q, Li H, Moorjani P, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**: 445-449.
- Fu Q, Mittnik A, Johnson PL, et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology* **23**: 553-559.
- Gallego Llorente M, Jones ER, Eriksson A, et al. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* **350**: 820-822.
- Gamba C, Jones ER, Teasdale MD, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature communications* **5**: 5257.
- Gao SZ, Zhang Y, Wei D, Li HJ, Zhao YB, et al. 2015. Ancient DNA Reveals a Migration of the Ancient Di-Qiang Populations Into Xinjiang as Early as the Early Bronze Age. *American Journal of Physical Anthropology* **157**: 71-80.
- Gaul JS, Winter E, Grossschmidt K. 2015. Ancient pathogens in museal dry bone specimens: analysis of paleocytology and aDNA. *Wiener Medizinische Wochenschrift* **165**: 133-139.
- Gismondi A, Di Marco G, Martini F, Sarti L, Crespan M, et al. 2016. Grapevine carpological remains revealed the existence of a Neolithic domesticated *Vitis vinifera* L. specimen containing ancient DNA partially preserved in modern ecotypes. *Journal of Archaeological Science* **69**: 75-84.

- Gokhman D, Lavi E, Prufer K, Fraga MF, Riancho JA, et al. 2014. Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. *Science* **344**: 523-527.
- Gomez-Sanchez D, Olalde I, Pierini F, et al. 2014. Mitochondrial DNA from El Mirador Cave (Atapuerca, Spain) Reveals the Heterogeneity of Chalcolithic Populations. *PLOS One* **9**.
- Goncalves VF, Stenderup J, Rodrigues-Carvalho C, et al. 2013. Identification of Polynesian mtDNA haplogroups in remains of Botocudo Amerindians from Brazil. *Proceedings of the National Academy of Sciences USA* **110**: 6465-6469.
- Grosser S, Rawlence NJ, Anderson CNK, Smith IWG, Scofield RP, et al. 2016. Invader or resident? Ancient-DNA reveals rapid species turnover in New Zealand little penguins. *Proceedings of the Royal Society B-Biological Sciences* **283**.
- Gubina MA, Kulikov IV, Babenko VN, Chikisheva TA, Romashchenko AG, et al. 2016. The dynamics of the composition of mtDNA haplotypes of the ancient population of the Altai Mountains from the early bronze age (3rd millennium BC) to the iron age (2nd-1st centuries BC). *Russian Journal of Genetics* **52**: 93-106.
- Gunther T, Valdiosera C, Malmstrom H, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences USA* **112**: 11917-11922.
- Guschanski K, Krause J, Sawyer S, et al. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systems Biology* **62**: 539-554.
- Gutiérrez-García TA V-DE, Arroyo-Cabrales J, Kuch M, Enk J, King C, et al. 2014. Ancient DNA and the tropics: a rodent's tale. *Biology Letters* **10**.

- Haak W, Lazaridis I, Patterson N, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**: 207-211.
- Halfman CM, Potter BA, McKinney HJ, Finney BP, Rodrigues AT, et al. 2015. Early human use of anadromous salmon in North America at 11,500 y ago. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 12344-12348.
- Haouchar D, Pacioni C, Haile J, McDowell MC, Baynes A, et al. 2016. Ancient DNA reveals complexity in the evolutionary history and taxonomy of the endangered Australian brush-tailed bettongs (Bettongia: Marsupialia: Macropodidae: Potoroinae). *Biodiversity and Conservation*.
- Harbeck M, Seifert L, Hansch S, et al. 2013. Yersinia pestis DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. *PLOS pathogens* **9**: e1003349.
- Harkins KM, Buikstra JE, Campbell T, Bos KI, Johnson ED, et al. 2015. Screening ancient tuberculosis with qPCR: challenges and opportunities. *Philosophical Transactions of the Royal Society B* **370**.
- Hervella M, Rotea M, Izagirre N, et al. 2015. Ancient DNA from South-East Europe Reveals Different Events during Early and Middle Neolithic Influencing the European Genetic Heritage. *PLOS One* **10**.
- Hofmanova Z, Kreutzer S, Hellenthal G, et al. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy Science USA* **113**: 6886-6891.
- Hollard C, Keyser C, Giscard PH, Tsagaan T, Bayarkhuu N, et al. 2014. Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental and ancestry informative markers. *Forensic Science International-Genetics* **12**: 199-207.

- Hou W, Dong H, Li G, et al. 2014a. Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Scientific Reports* **4**.
- Hou XD, Sheng GL, Yin S, Zhu M, Du M, et al. 2014b. DNA analyses of wild boar remains from archaeological sites in Guangxi, China. *Quaternary International* **354**: 147-153.
- Hung CM, Shaner PJJ, Zink RM, Liu WC, Chu TC, et al. 2014. Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 10636-10641.
- Huynen L, Suzuki T, Ogura T, Watanabe Y, Millar CD, et al. 2014. Reconstruction and in vivo analysis of the extinct *tbx5* gene from ancient wingless moa (Aves: Dinornithiformes). *BMC Evolutionary Biology* **14**.
- Huynen LL, DM. 2014. Complex species status for extinct moa (Aves: Dinornithiformes) from the genus *Euryapteryx*. *PLOS One* **9**: e90212.
- Inskip SA, Taylor GM, Zakrzewski SR, et al. 2015. Osteological, biomolecular and geochemical examination of an early anglo-saxon case of lepromatous leprosy. *PLOS One* **10**: e0124282.
- Jacomb C, Holdaway RN, Allentoft ME, Bunce M, Oskam CL, et al. 2014. High-precision dating and ancient DNA profiling of moa (Aves: Dinornithiformes) eggshell documents a complex feature at Wairau Bar and refines the chronology of New Zealand settlement by Polynesians. *Journal of Archaeological Science* **50**: 24-30.
- Jiao LC, Liu XL, Jiang XN, Yin YF. 2015. Extraction and amplification of DNA from aged and archaeological *Populus euphratica* wood for species identification. *Holzforschung* **69**: 925-931.

- Jin HJ, Kim KC, Kim W. 2015. Phylogenetic analysis of two haploid markers of 500-years-old human remains found in a central region of Korea. *Genes & Genomics* **37**: 33-43.
- Jones ER, Gonzalez-Forbes G, Connell S, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature communications* **6**: 8912.
- Juras A, Dabert M, Kushniarevich A, Malmstrom H, Raghavan M, et al. 2014. Ancient DNA Reveals Matrilineal Continuity in Present-Day Poland over the Last Two Millennia. *PLOS One* **9**.
- Kay GL, Sergeant MJ, Zhou Z, et al. 2015. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nature communications* **6**: 6717.
- Keyser C, Hollard C, Gonzalez A, Fausser JL, Rivals E, et al. 2015. The ancient Yakuts: a population genetic enigma. *Philosophical Transactions of the Royal Society B* **370**.
- Khairat R, Ball M, Chang CC, et al. 2013. First insights into the metagenome of Egyptian mummies using next-generation sequencing. *Journal of Applied Genetics* **54**: 309-325.
- Kistler L RA, Godfrey LR, Crowley BE, Hughes CE, Lei R, et al. 2014. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *Journal of Human Evolution* **79**: 45-54.
- Krause-Kyora B, Makarewicz C, Evin A, et al. 2013. Use of domesticated pigs by Mesolithic hunter-gatherers in northwestern Europe. *Nature communications* **4**: 2348.
- Krutli A, Bouwman A, Akgul G, Della Casa P, Ruhli F, et al. 2014. Ancient DNA analysis reveals high frequency of European lactase persistence allele (T-13910) in medieval central Europe. *PLOS One* **9**: e86251.

- Krzewinska M, Bjornstad G, Skoglund P, Olason PI, Bill J, et al. 2015. Mitochondrial DNA variation in the Viking age population of Norway. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **370**.
- Kuhlwilm M, Gronau I, Hubisz MJ, et al. 2016. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**: 429-433.
- Kyle M, Haande S, Sonstebo J, Rohrlack T. 2015. Amplification of DNA in sediment cores to detect historic *Planktothrix* occurrence in three Norwegian lakes. *Journal of Paleolimnology* **53**: 61-72.
- Lagerholm VK, Sandoval-Castellanos E, Ehrich D, et al. 2014. On the origin of the Norwegian lemming. *Molecular Ecology* **23**: 2060-2071.
- Lalremruata A, Ball M, Bianucci R, Welte B, Nerlich AG, et al. 2013. Molecular identification of *falciparum* malaria and human tuberculosis co-infections in mummies from the Fayum depression (Lower Egypt). *PLOS One* **8**: e60307.
- Lari M, Di Vincenzo F, Borsato A, et al. 2015. The Neanderthal in the karst: First dating, morphometric, and paleogenetic data on the fossil skeleton from Altamura (Italy). *Journal of Human Evolution* **82**: 88-94.
- Lazaridis I, Patterson N, Mittnik A, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**: 409-413.
- Lee E, Krause-Kyora B, Rinne C, Schutt R, Harder M, et al. 2014. Ancient DNA insights from the Middle Neolithic in Germany. *Archaeological and Anthropological Sciences* **6**: 199-204.
- Lee EJ, Merriwether DA, Kasparov AK, Nikolskiy PA, Sotnikova MV, et al. 2015. Ancient DNA Analysis of the Oldest Canid Species from the Siberian Arctic and Genetic Contribution to the Domestic Dog. *PLOS One* **10**.

- Li CX, Ning C, Hagedberg E, Li HJ, Zhao YB, et al. 2015. Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genetics* **16**.
- Librado P, Der Sarkissian C, Ermini L, et al. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Science USA* **112**: 6889-6897.
- Lister DL, Jones H, Jones MK, O'Sullivan DM, Cockram J. 2013. Analysis of DNA polymorphism in ancient barley herbarium material: Validation of the KASP SNP genotyping platform. *Taxon* **62**: 779-789.
- Llamas B, Brotherton P, Mitchell KJ, et al. 2015. Late Pleistocene Australian Marsupial DNA Clarifies the Affinities of Extinct Megafaunal Kangaroos and Wallabies. *Molecular Biology and Evolution* **32**: 574-584.
- Ludwig A, Reissmann M, Benecke N, et al. 2015. Twenty-five thousand years of fluctuating selection on leopard complex spotting and congenital night blindness in horses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **370**: 20130386.
- Lynch VJ, Bedoya-Reina OC, Ratan A, Sulak M, Drautz-Moses DI, et al. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell reports* **12**: 217-228.
- Maixner F, Krause-Kyora B, Turaev D, et al. 2016. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* **351**: 162-165.
- Malaspina AS, Lao O, Schroeder H, et al. 2014. Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. *Current Biology* **24**: 1035-1037.

- Malmstrom H, Linderholm A, Skoglund P, et al. 2015. Ancient mitochondrial DNA from the northern fringe of the Neolithic farming expansion in Europe sheds light on the dispersion process. *Philosophical Transactions of the Royal Society B* **370**.
- Martin MD, Cappellini E, Samaniego JA, et al. 2013. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature communications* **4**: 2172.
- Martin MD, Ho SY, Wales N, Ristaino JB, Gilbert MT. 2014. Persistence of the mitochondrial lineage responsible for the Irish potato famine in extant new world phytophthora infestans. *Molecular Biology and Evolution* **31**: 1414-1420.
- Martiniano R, Caffell A, Holst M, et al. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature communications* **7**: 10326.
- Mascher M, Schuenemann VJ, Davidovich U, Marom N, et al. 2016. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication of barley. *Nature Genetics* **48**: 1089-1093.
- Masson M, Molnar E, Donoghue HD, Besra GS, Minnikin DE, et al. 2013. Osteological and biomolecular evidence of a 7000-year-old case of hypertrophic pulmonary osteopathy secondary to tuberculosis from neolithic hungary. *PLOS One* **8**: e78252.
- Mathieson I, Lazaridis I, Rohland N, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**.
- Matisoo-Smith E. 2015. Ancient DNA and the human settlement of the Pacific: A review. *Journal of Human Evolution* **79**: 93-104.

- Meiri M, Lister AM, Higham TF, Stewart JR, Straus LG, et al. 2013. Late-glacial recolonization and phylogeography of European red deer (*Cervus elaphus L.*). *Molecular Ecology* **22**: 4711-4722.
- Mendisco F, Pemonge MH, Leblay E, Romon T, Richard G, et al. 2015. Where are the Caribs? Ancient DNA from ceramic period human remains in the Lesser Antilles. *Philosophical Transactions of the Royal Society B* **370**.
- Mendum TA, Schuenemann VJ, Roffey S, et al. 2014. Mycobacterium leprae genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. *BMC Genomics* **15**: 270.
- Meyer M, Arsuaga JL, de Filippo C, et al. 2016. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**: 504-507.
- Meyer M FQ, Aximu-Petri A, Glocke I, et al. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* **505**.
- Mikheyev AS, Tin MM, Arora J, Seeley TD. 2015. Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nature communications* **6**: 7991.
- Mikic AM. 2015. The First Attested Extraction of Ancient DNA in Legumes (Fabaceae). *Frontiers in Plant Science* **6**.
- Mitchell KJ, Llamas B, Soubrier J, Rawlence NJ, Worthy TH, et al. 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898-900.
- Morris AG, Heinze A, Chan EKF, Smith AB, Hayes VM. 2014. First Ancient Mitochondrial Human Genome from a Prepastoralist Southern African. *Genome Biology and Evolution* **6**: 2647-2653.

- Morris K, Austin JJ, Belov K. 2013. Low major histocompatibility complex diversity in the Tasmanian devil predates European settlement and may explain susceptibility to disease epidemics. *Biology Letters* **9**.
- Motti JMB, Hagelberg E, Lindo J, Malhi RS, Bravi CM, et al. 2015. First complete mitochondrial genome sequence from human skeletal remains of the coast of Santa Cruz, Argentina. *Magallania* **43**: 119-131.
- Muller R, Roberts CA, Brown TA. 2014a. Biomolecular identification of ancient Mycobacterium tuberculosis complex DNA in human remains from Britain and continental Europe. *American Journal of Physical Anthropology* **153**: 178-189.
- Muller R, Roberts CA, Brown TA. 2014b. Genotyping of ancient Mycobacterium tuberculosis strains reveals historic genetic diversity. *Proceedings of the Royal Society B-Biological Sciences* **281**.
- Naumann E, Krzewinska M, Gotherstrom A, Eriksson G. 2014. Slaves as burial gifts in Viking Age Norway? Evidence from stable isotope and ancient DNA analyses. *Journal of Archaeological Science* **41**: 533-540.
- Ng TFF, Chen LF, Zhou YC, et al. 2014. Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 16842-16847.
- Niemi M, Blauer A, Iso-Touru T, Nystrom V, Harjula J, et al. 2013. Mitochondrial DNA and Y-chromosomal diversity in ancient populations of domestic sheep (*Ovis aries*) in Finland: comparison with contemporary sheep breeds. *Genetics, selection, evolution* **45**.
- Oh CS, Seo M, Hong JH, Chai JY, Oh SW, et al. 2015. Ancient Mitochondrial DNA Analyses of Ascaris Eggs Discovered in Coprolites from Joseon Tomb. *Korean Journal of Parasitology* **53**: 237-242.

- Olalde I, Allentoft ME, Sanchez-Quinto F, et al. 2014a. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**: 225-228.
- Olalde I, Sanchez-Quinto F, Datta D, et al. 2014b. Genomic analysis of the blood attributed to Louis XVI (1754-1793), king of France. *Scientific Reports* **4**.
- Olalde I, Schroeder H, Sandoval-Velasco M, et al. 2015. A Common Genetic Origin for Early Farmers from Mediterranean Cardial and Central European LBK Cultures. *Molecular Biology and Evolution* **32**: 3132-3142.
- Orlando L, Ginolhac A, Zhang G, et al. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**: 74-78.
- Otonari C, Flink LG, Evin A, et al. 2013. Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics. *Molecular Biology and Evolution* **30**: 824-832.
- Pacioni C, Hunt H, Allentoft ME, Vaughan TG, Wayne AF, et al. 2015. Genetic diversity loss in a biodiversity hotspot: ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Molecular Ecology* **24**: 5813-5828.
- Palkopoulou E, Mallick S, Skoglund P, et al. 2015. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Current Biology* **25**: 1395-1400.
- Palla F, Billeci N, Spallino RE, Raimondo FM. 2013. Molecular approach for the characterization of ancient/degraded *Cyperus* sp specimens. *Science and Technology for the Conservation of Cultural Heritage*.
- Parducci L, Valiranta M, Salonen JS, Ronkainen T, Matetovici I, et al. 2015. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA.

*Philosophical transactions of the Royal Society of London Series B, Biological sciences* **370**.

Park SD, Magee DA, McGettigan PA, et al. 2015. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biology* **16**: 234.

Pawlowska J, Lejzerowicz F, Esling P, Szczucinski W, Zajaczkowski M et al. 2014. Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. *Geobiology* **12**: 277-288.

Pedersen JS, Valen E, Velazquez AM, et al. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Research* **24**: 454-466.

Pinhasi R, Fernandes D, Sirak K, et al. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLOS One* **10**.

Prufer K, Racimo F, Patterson N, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**: 43-49.

Raghavan M, DeGiorgio M, Albrechtsen A, et al. 2014a. The genetic prehistory of the New World Arctic. *Science* **345**.

Raghavan M, Skoglund P, Graf KE, et al. 2014b. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**: 87-91.

Rasmussen M, Anzick SL, Waters MR, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**: 225-229.

Rasmussen M, Sikora M, Albrechtsen A, et al. 2015a. The ancestry and affiliations of Kennewick Man. *Nature* **523**: 455-458.

- Rasmussen S, Allentoft ME, Nielsen K, et al. 2015b. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**: 571-582.
- Rawlence NJ, Perry GLW, Smith IWG, Scofield RP, Tennyson AJD, et al. 2015. Radiocarbon-dating and ancient DNA reveal rapid replacement of extinct prehistoric penguins. *Quaternary Science Reviews* **112**: 59-65.
- Reynolds AW, Raff JA, Bolnick DA, Cook DC, Kaestle FA. 2015. Ancient DNA From the Schild Site in Illinois: Implications for the Mississippian Transition in the Lower Illinois River Valley. *American Journal of Physical Anthropology* **156**: 434-448.
- Rivera-Perez JI, Cano RJ, Narganes-Storde Y, Chanlatte-Baik L, Toranzos GA. 2015. Retroviral DNA Sequences as a Means for Determining Ancient Diets. *PLOS One* **10**.
- Rodriguez-Varela R, Garcia N, Nores C, Alvarez-Lao D, Barnett R, et al. 2016. Ancient DNA reveals past existence of Eurasian lynx in Spain. *Journal of Zoology* **298**: 94-102.
- Rodriguez-Varela R, Tagliacozzo A, Urena I, Garcia N, Cregut-Bonnoure E, et al. 2015. Ancient DNA evidence of Iberian lynx palaeoendemism. *Quaternary Science Reviews* **112**: 172-180.
- Salis AT, Easton LJ, Robertson BC, Gemmell N, Smith IWG, et al. 2016. Myth or relict: Does ancient DNA detect the enigmatic Upland seal? *Molecular Phylogenetics and Evolution* **97**: 101-106.
- Sawyer S, Renaud G, Viola B, et al. 2015. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 15696-15700.

- Schiffels S, Haak W, Paajanen P, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature communications* **7**: 10408.
- Schroeder H, Avila-Arcos MC, Malaspinas AS, et al. 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 3669-3673.
- Schubert M, Jonsson H, Chang D, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proceedings of the National Academy of Science of the United States of America* **111**: 5661-5669.
- Schuenemann VJ, Singh P, Mendum TA, et al. 2013. Genome-Wide Comparison of Medieval and Modern Mycobacterium leprae. *Science* **341**: 179-183.
- Seguin-Orlando A, Korneliussen TS, Sikora M, et al. 2014. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**: 1113-1118.
- Sheng GL SJ, Liu JY et al. 2014. Pleistocene Chinese cave hyenas and the recent Eurasian history of the spotted hyena, *Crocuta crocuta*. *Molecular Ecology* **23**.
- Shin DH, Oh CS, Lee HJ, Chai JY, Lee SJ, et al. 2013. Ancient DNA analysis on Clonorchis sinensis eggs remained in samples from medieval Korean mummy. *Journal of Archaeological Science* **40**: 211-216.
- Sinding MHS, Arneborg J, Nyegaard G, Gilbert MTP. 2015. Ancient DNA unravels the truth behind the controversial GUS Greenlandic Norse fur samples: the bison was a horse, and the muskox and bears were goats. *Journal of Archaeological Science* **53**: 297-303.
- Skoglund P, Ersmark E, Palkopoulou E, Dalen L. 2015. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology* **25**: 1515-1519.

- Smith O, Clapham A, Rose P, Liu Y, Wang J, et al. 2014a. A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Scientific Reports* **4**: 4003.
- Smith O, Clapham AJ, Rose P, Liu Y, Wang J, et al. 2014b. Genomic methylation patterns in archaeological barley show de-methylation as a time-dependent diagenetic process. *Scientific Reports* **4**: 5559.
- Smith O, Momber G, Bates R, Garwood P, Fitch S, et al. 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* **347**: 998-1001.
- Smith RW, Monroe C, Bolnick DA. 2015. Detection of cytosine methylation in ancient DNA from five native american populations using bisulfite sequencing. *PLOS One* **10**: e0125344.
- Smith RWA and Bolnick DA. 2013. Analysis of cytosine methylation in Native American ancient DNA. *American Journal of Physical Anthropology* **150**: 258-258.
- Soe MJ, Nejsum P, Fredensborg BL, Kapel CMO. 2015. DNA typing of ancient parasite eggs from environmental samples identifies human and animal worm infections in viking-age settlement. *Journal of Parasitology* **101**: 57-63.
- Stanton DWG, Mulville JA, Bruford MW. 2016. Colonization of the Scottish islands via long-distance Neolithic transport of red deer (*Cervus elaphus*). *Proceedings of the Royal Society B-Biological Sciences* **283**.
- Suzuki K, Saso A, Hoshino K, et al. 2014. Paleopathological Evidence and Detection of Mycobacterium leprae DNA from Archaeological Skeletal Remains of Nabe-kaburi (Head-Covered with Iron Pots) Burials in Japan. *PLOS One* **9**.

- Svensson EM, Hasler S, Nussbaumer M, Rehazek A, Omrak A et al. 2014. Mediaeval cattle from Bern (Switzerland): An archaeozoological, genetic and historical Approach. *Schweizer Archiv Fur Tierheilkunde* **156**: 17-26.
- Tanaka K, Kamijo N, Tabuchi H, Hanamori K, Matsuda R, et al. 2016. Morphological and molecular genetics of ancient remains and modern rice (*Oryza sativa*) confirm diversity in ancient Japan. *Genetic Resources and Crop Evolution* **63**: 447-464.
- Taylor GM, Tucker K, Butler R, et al. 2013. Detection and strain typing of ancient *Mycobacterium leprae* from a medieval leprosy hospital. *PLOS One* **8**: e62406.
- Teasdale MD, van Doorn NL, Fiddymment S, Webb CC, O'Connor T, et al. 2015. Paging through history: parchment as a reservoir of ancient DNA for next generation sequencing. *Philosophical Transactions of the Royal Society B* **370**.
- Thalmann O, Shapiro B, Cui P, et al. 2013. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science* **342**: 871-874.
- Thomson VA, Lebrasseur O, Austin JJ, et al. 2014. Using ancient DNA to study the origins and dispersal of ancestral Polynesian chickens across the Pacific. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4826-4831.
- Tierney SN and Bird JM. 2015. Molecular sex identification of juvenile skeletal remains from an Irish medieval population using ancient DNA analysis. *Journal of Archaeological Science* **62**: 27-38.
- Tsangaras K, Siracusa MC, Nikolaidis N, Ishida Y, Cui P, et al. 2014. Hybridization capture reveals evolution and conservation across the entire Koala retrovirus genome. *PLOS One* **9**: e95633.

- Valverde G, Romero MIB, Espinoza IF, Cooper A, Fehren-Schmitz L, et al. 2016. Ancient DNA Analysis Suggests Negligible Impact of the Wari Empire Expansion in Peru's Central Coast during the Middle Horizon. *PLOS One* **11**.
- Villanea FA, Parent CE, Kemp BM. 2016. Reviving Galapagos snails: ancient DNA extraction and amplification from shells of probably extinct endemic land snails. *Journal of Molluscan Studies* **82**: 449-456.
- Vilstrup JT, Seguin-Orlando A, Stiller M, et al. 2013. Mitochondrial phylogenomics of modern and ancient equids. *PLOS One* **8**: e55950.
- Vuillemin A, Ariztegui D, Leavitt PR, Bunting L, Team PS. 2016. Recording of climate and diagenesis through sedimentary DNA and fossil pigments at Laguna Potrok Aike, Argentina. *Biogeosciences* **13**: 2475-2492.
- Wagner DM, Klunk J, Harbeck M, et al. 2014. *Yersinia pestis* and the plague of Justinian 541-543 AD: a genomic analysis. *The Lancet Infectious diseases* **14**: 319-326.
- Warinner C, Rodrigues JF, Vyas R, et al. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* **46**: 336-344.
- Warinner C, Speller C, Collins MJ. 2015a. A new era in palaeomicrobiology: prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society B* **370**.
- Warinner C, Speller C, Collins MJ, Lewis CM. 2015b. Ancient human microbiomes. *Journal of Human Evolution* **79**: 125-136.
- Wilde S, Timpson A, Kirsanow K, et al. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proceedings of the National Academy of Science USA* **111**: 4832-4837.

- Willerslev E, Davison J, Moora M, et al. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**: 47-51.
- Witas HW, Tomczyk J, Jedrychowska-Danska K, Chaubey G, Ploszaj T. 2013. mtDNA from the early Bronze Age to the Roman period suggests a genetic link between the Indian subcontinent and Mesopotamian cradle of civilization. *PLOS One* **8**.
- Wooller MJ, Gaglioti B, Fulton TL, Lopez A, Shapiro B. 2015. Post-glacial dispersal patterns of Northern pike inferred from an 8800 year old pike (*Esox cf. lucius*) skull from interior Alaska. *Quaternary Science Reviews* **120**: 118-125.
- Xiang H, Gao JQ, Yu BQ, et al. 2014. Early Holocene chicken domestication in northern China. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 17564-17569.
- Yoshida K, Schuenemann VJ, Cano LM, et al. 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *Elife* **2**.
- Yuan JX, Sheng GL, Hou XD, Shuang XY, Yi J, et al. 2014. Ancient DNA sequences from *Coelodonta antiquitatis* in China reveal its divergence and phylogeny. *Science China-Earth Sciences* **57**: 388-396.
- Zhang H, Paijmans JL, Chang F, et al. 2013. Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nature communications* **4**: 2755.
- Zhao YB, Zhang Y, Zhang QC, Li HJ, Cui YQ, et al. 2015. Ancient DNA Reveals That the Genetic Structure of the Northern Han Chinese Was Shaped Prior to 3,000 Years Ago. *PLOS One* **10**.
- Ziesemer KA, Mann AE, Sankaranarayanan K, et al. 2015. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification. *Scientific Reports* **5**.

— **APPENDIX I** —

---

**RESEARCH OUTPUT**

---

## I.I PUBLICATIONS

---

**Grealy A**, McDowell MC, Scofield P, Murray DC, Fusco DA, Haile J, Prideaux GJ, Bunce M. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. *Quaternary Science Reviews* **128**: 37-47.

**Grealy A**, Macken A, Allentoft ME, Rawlence NJ, Reed E, Bunce M. 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *Journal of Quaternary Science* **31**: 33-45.

**Grealy A**, Douglass K, Haile J, Bruwer C, Gough C, Bunce M. 2016. Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar. *Journal of Archaeological Science* **75**: 82-88.

**Grealy A**, Phillips M, Miller G, Gilbert MTP, Rouillard J, Lambert D, Bunce M, Haile J. 2017. Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Molecular Phylogenetics and Evolution* **109**: 151-163.

## I.II AWARDS *and* CONFERENCE PRESENTATIONS

---

- 2016** | *Eggshell palaeogenomics and the evolution of birds*  
International Society for Molecular Biology and Evolution conference
- 2016** | *Ancient DNA reveals past subsistence practices in southwest Madagascar*  
International Society of Africanist and Archaeologists conference
- 2015** | *Bones to biodiversity: ancient DNA preservation at the Naracoorte Caves, SA*  
Local Combined Biological Sciences Meeting
- 2014** | *Palaeognath biogeography: the elephant (bird) in the room*  
Local Combined Biological Sciences Meeting (Edith Cowan University Student Talk Prize)
- 2013** | *Morphology versus molecules: ¿por qué no los dos?*  
Conference on Australian Vertebrate Evolution, Palaeontology and Systematics (Student Poster Prize)
- Murdoch University Student Poster Day (Omics Australasia Poster Prize)

## I.III MEDIA ENGAGEMENT

---

- 2016** | Radio interview with ABC Radio Mt Gambier regarding our aDNA results from the Naracoorte Caves
- 2016** | *Ancient fossil DNA found at Naracoorte Caves*  
Naracoorte Herald  
[www.naracoorteherald.com.au/story/3777368/exciting-find-at-caves/](http://www.naracoorteherald.com.au/story/3777368/exciting-find-at-caves/)
- 2016** | *Old DNA reveals fishing habits*  
Research Highlights, *Nature* **539**: 333

**2016** | *What do you call a school of bones?*

Kasulis, K. *The Boston Globe*, December 10, 2016

<https://www.bostonglobe.com/ideas/2016/12/10/brainiac/MSMhznzjhJNW0GB0dxyNCyK/story.html>



# A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages



Alicia C. Grealy<sup>a, b</sup>, Matthew C. McDowell<sup>c</sup>, Paul Scofield<sup>d</sup>, Dáithí C. Murray<sup>a, b</sup>,  
Diana A. Fusco<sup>c</sup>, James Haile<sup>a, b</sup>, Gavin J. Prideaux<sup>c</sup>, Michael Bunce<sup>a, b, \*</sup>

<sup>a</sup> Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Kent St, Bentley, WA 6102, Australia

<sup>b</sup> Ancient DNA Laboratory, School of Veterinary and Life Sciences, Murdoch University, South St, Murdoch, WA 6150, Australia

<sup>c</sup> School of Biological Sciences, Flinders University, Bedford Park, SA 5042, Australia

<sup>d</sup> Canterbury Museum, Christchurch 8013, New Zealand

## ARTICLE INFO

### Article history:

Received 22 April 2015

Received in revised form

6 September 2015

Accepted 11 September 2015

Available online xxx

### Keywords:

aDNA

ancient DNA

Archaeology

Biodiversity

Bulk bone

Experimental error

Fossil

Metabarcoding

Next-generation sequencing

Palaeontology

## ABSTRACT

When pooled for extraction as a bulk sample, the DNA within morphologically unidentifiable fossil bones can, using next-generation sequencing, yield valuable taxonomic data. This method has been proposed as a means to rapidly and cost-effectively assess general ancient DNA preservation at a site, and to investigate temporal and spatial changes in biodiversity; however, several caveats have yet to be considered. We critically evaluated the bulk bone metabarcoding (BBM) method in terms of its: (i) repeatability, by quantifying sampling and technical variance through a nested experimental design containing sub-samples and replicates at several stages; (ii) accuracy, by comparing morphological and molecular family-level identifications; and (iii) overall utility, by applying the approach to two independent Holocene fossil deposits, Bat Cave (Kangaroo Island, Australia) and Finsch's Folly (Canterbury, New Zealand). For both sites, bone and bone powder sub-sampling were found to contribute significantly to variance in molecularly identified family assemblage, while the contribution of library preparation and sequencing was almost negligible. Nevertheless, total variance was small. Sampling over 80% fewer bones than was required to morphologically identify the taxonomic assemblages, we found that the families identified molecularly are a subset of the families identified morphologically and, for the most part, represent the most abundant families in the fossil record. In addition, we detected a range of extinct, extant and endangered taxa, including some that are rare in the fossil record. Given the relatively low sampling effort of the BBM approach compared with morphological approaches, these results suggest that BBM is largely consistent, accurate, sensitive, and therefore widely applicable. Furthermore, we assessed the overall benefits and caveats of the method, and suggest a workflow for palaeontologists, archaeologists, and geneticists that will help mitigate these caveats. Our results show that DNA analysis of bulk bone samples can be a universally useful tool for studying past biodiversity, when integrated with existing morphology-based approaches. Despite several limitations that remain, the BBM method offers a cost-effective and efficient way of studying fossil assemblages, offering complementary insights into evolution, extinction, and conservation.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

For over a century, the study of fossils has played a major role in understanding prehistoric life and evolutionary processes. In

particular, morphological analyses of fossils can reveal species that existed in the past, help elucidate the evolutionary relationships of extinct and extant species (e.g., Donoghue et al., 1989; Deméré et al., 2005; Manos et al., 2007), and assist the development of palaeoenvironment reconstructions that provide insights into the evolutionary and ecological impacts of environmental changes (e.g., Rodríguez-Aranda and Calvo, 1998; Zhang et al., 2008). However, such traditional methods have limitations. For instance, taxonomic assignments of fossils have been necessarily reliant on morphological distinctions, making the identification of

\* Corresponding author. Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Kent St, Bentley, WA 6102, Australia.

E-mail address: [michael.bunce@curtin.edu.au](mailto:michael.bunce@curtin.edu.au) (M. Bunce).

fragmented or taxonomically-mixed fossil material challenging, if not impossible. This limitation can be partially overcome in some Late Quaternary contexts with the application of ancient DNA (aDNA) techniques. Over the past two decades, aDNA has proved to be a useful complement to the morphological study of fossils, and is rapidly growing in popularity, accessibility, and applicability. In combination with next-generation sequencing (NGS), aDNA has been used to test phylogenetic relationships, and timing of speciation and extinction trajectories (e.g., Krause et al., 2010), resolve taxonomy (e.g., Rohland et al., 2010), reconstruct palaeoenvironments (e.g., Willerslev et al., 2003), and measure historic genetic diversity (e.g., Larson et al., 2002; Allentoft et al., 2014).

Despite the utility of aDNA analysis, unidentifiable bone fragments that are retrieved from palaeontological and archaeological excavations are often too numerous and small to justify the expense of aDNA analysis. However, if such bones are pooled for aDNA extraction as one bulk sample, the pool may be sequenced cost-effectively to yield valuable systematic data useful for assessing past biodiversity over time and space (Murray et al., 2013). In addition, bulk bone samples may be useful for evaluating general aDNA preservation at a site, without requiring the destruction of complete or precious fossil specimens (Murray et al., 2013). The bulk bone method employs a metabarcoding approach (Taberlet et al., 2012), which involves: (1) simultaneous extraction of aDNA from multiple unidentifiable fragments of bone; (2) amplification of short, 'diagnostic' regions of mitochondrial genes by polymerase chain reaction (PCR); and (3) sequencing (via NGS) of these amplicons to identify the species present by comparison with a genetic database of known species (e.g., GenBank; Altschul et al., 1990; Benson et al., 2006). Metabarcoding has been used to evaluate both present and past biodiversity (Epp et al., 2012) through the analysis of environmental samples such as sediments (e.g., Jørgensen et al., 2011; Andersen et al., 2012; Pedersen et al., 2013; Epp et al., 2015; Pansu et al., 2015), seawater (e.g., Minamoto et al., 2012; Thomsen et al., 2012), coprolites (e.g., Hofreiter et al., 2003), and middens (e.g., Murray et al., 2012), and has even been able to detect taxa that were considered extinct based on the macrofossil record (e.g., Haile et al., 2009; Haouchar et al., 2014). Using a metabarcoding approach to generate biodiversity data has the potential to significantly reduce workload and costs compared with a morphological approach that can be labour intensive, or require large amounts of taxonomic expertise and time investment (Ji et al., 2013). When combined with the use of indexing (Binladen et al., 2007; Meyer et al., 2007; Kircher et al., 2012) (where DNA from each bulk sample is 'tagged' with a few unique bases), multiple DNA samples can be combined with equimolarity and sequenced in parallel (i.e., 'sample multiplexing') on an NGS platform, increasing throughput and further reducing cost and time.

Although the bulk bone metabarcoding (BBM) method has been implemented in several recent studies (e.g., Murray et al., 2013; Haouchar et al., 2014), several caveats and biases of this method have yet to be addressed. Firstly, the amount of variance attributable to experimental error in the BBM method, as well as other environmental metabarcoding methods (Andersen et al., 2012; Pedersen et al., 2013; Porter et al., 2013), has not been measured. As such, it is unknown at what step, and to what extent, efforts need to be concentrated to minimise experimental error (Earp et al., 2011), and an optimal experimental protocol has not been developed. This is important if we wish to confidently compare how biodiversity has changed across time or space, in order to reliably determine what has driven those changes (Wooley et al., 2010). The "ability of the researcher to obtain a statistically significant result" (Kitchen et al., 2010) is influenced by the treatment effect, and repeatability (or precision), which is affected by biological variability and technical noise. For the BBM method, the

treatment effect is the variance in biodiversity *between* samples that arises from differences between palaeontological collection sites (space) or stratigraphic layers (time); biological variability refers to the differences in biodiversity *within* samples resulting from sub-sampling effort and differential DNA preservation in the fossils; and technical noise is the variability in biodiversity introduced by the experimental protocol itself (including sub-sampling bone powder for DNA extraction, human error, random contamination, stochastic variations in quantitative PCR and amplification biases, aDNA damage, PCR and sequencing errors, and amplicon pooling during the creation of NGS libraries). In order to isolate the treatment effect from the background (Kitchen et al., 2010), we can quantify the contribution that each of these factors makes to the total variance in the data (the 'experimental error') through a careful experimental design containing multiple sub-samples, biological and technical replicates, stringent laboratory protocols, and the use of multiple blank controls at each step (Kuehl, 2000; Macgregor, 2007; Kitchen et al., 2010).

Secondly, most metabarcoding studies of environmental samples have found discrepancies between estimates of biodiversity obtained from DNA metabarcoding methods and those obtained from traditional biodiversity sampling methods (Ji et al., 2013) because some species identified morphologically were not identified via DNA methods, and vice versa (Hajibabaei et al., 2011; Murray et al., 2013). These discrepancies arise from differences in the biomass and behaviour of animals (Andersen et al., 2012), as well as sampling effort, differential preservation of both fossils and the aDNA within them, technical 'noise' (such as amplification bias, PCR and sequencing error; Fonseca et al., 2012), and deficiencies in reference genetic databases, such as GenBank (Pedersen et al., 2014). It is likely that BBM studies would be affected by similar biases (Murray et al., 2013); however, the extent to which bulk-bone taxonomic identifications reflect those arising from the fossil record has yet to be examined.

In this paper we critically evaluate the BBM approach in terms of its repeatability, accuracy and overall utility. Repeatability was assessed by estimating the contribution to variance made by experimental error using a nested experimental design containing pooled bone sub-samples (biological variability), bone powder (extraction) sub-samples, and library preparation and sequencing run replicates (technical variability)—this allows us to determine where sampling effort and replication need to be concentrated in order to reduce variance in the detection of families and operational taxonomic units (OTUs). Accuracy was assessed by comparing the family assemblages derived from morphological identification of fossil collections with those derived from a subset subjected to BBM analysis. Finally, overall utility was assessed by applying the approach to two independent Holocene fossil deposits, Bat Cave (BC; Kangaroo Island, Australia) and Finsch's Folly (FF; Canterbury, New Zealand). These methods enabled us to gauge the strengths, limitations, and biases of the BBM approach in order to assess how it complements traditional palaeontological methods.

## 2. Materials and methods

### 2.1. Study systems and sample collection

#### 2.1.1. Bat Cave

Located in the Kelly Hill Caves Conservation Park, south-west Kangaroo Island, South Australia (SI Fig. 1a), Bat Cave (BC; 35° 59' S, 136° 54' E; Cave Exploration Group of South Australia no. 5K65) consists of a single chamber with a rock-pile entrance (SI 1). The taphonomic characteristics of the BC assemblage (maximum species body mass, presence of invertebrate remains, degree of

digestive erosion, and inclusion of predatory species in the accumulation) are consistent with accumulation by the predatory activity of Boobook owls (*Ninox novaeseelandiae*; Fusco, 2014). Gloves were worn to collect fossils from shallow unstratified, calcareous sandy sediments found among rocks on the entrance slope. Samples were dry sieved using 1.5 mm mesh to concentrate bones. Diagnostic elements were removed for identification and analysis using a small brush and forceps. Morphologically non-diagnostic bones were stored in an airtight bag at  $-20^{\circ}\text{C}$  until needed for molecular analyses.

### 2.1.2. Finsch's Folly

FF (44 2' S, 170 43.5' E) is a shallow (<8 m) pitfall cave located in an isolated outcrop of limestone in south-western Canterbury (SI Fig. 1b). Fossils were collected from muddy, unstratified, calcareous sediments that had accumulated at the bottom of the pitfall and also from sediments that have been concreted to the walls during a previous period of infill. Samples were wet sieved using 1 mm mesh to concentrate bones. Diagnostic elements were removed for identification and analysis. The remaining morphologically non-diagnostic bones were dried and stored in airtight bags at  $-20^{\circ}\text{C}$  until needed for molecular analyses.

### 2.2. Dating

One bone sample from BC (the only bone suitable for radiocarbon dating, a left humerus of *N. novaeseelandiae*, Wk-36239), and three samples from FF were Accelerator Mass Spectrometry (AMS) radiocarbon dated by Waikato Radiocarbon Dating Laboratory. The conventional ages of the samples were calibrated with OxCal v4.1.7 (Bronk Ramsey, 2010) using the SHCal13 Southern Hemisphere calibration curve (Hogg et al., 2013). The *N. novaeseelandiae* bone yielded a conventional age of  $2862 \pm 25$  BP and calibrated age of 2750–3000 cal BP (Fusco, 2014). While it is likely that this bone does not represent the age of all the bones in the BC accumulation, the presence of introduced species in the assemblage bolsters our confidence that it accumulated in the late Holocene, which provides adequate temporal resolution for comparison with FF. The samples from FF yielded conventional ages of  $1344 \pm 25$  BP,  $1646 \pm 25$  BP, and  $1645 \pm 24$  BP, and calibrated ages of 1150–1300 cal BP, 1400–1600 cal BP and 1400–1600 cal BP, respectively.

### 2.3. Morphological assessment of fossil assemblage

Diagnostic bones from both sites (whole and part skulls, maxillae, dentaries and/or teeth, and post-cranial bones) were identified using published descriptions and comparative specimens (c.f. Fusco, 2014). All specimens were identified to the lowest taxonomic level possible (usually species). The minimum number of individuals (MNI) was determined by counting the number of the most common diagnostic element of each species in each assemblage. MNI was converted to relative abundance (Ri%), an expression of the MNI of a given species as a proportion of the total MNI for that collection (McDowell et al., 2012; Fusco, 2014).

### 2.4. Experimental design

In order to partition the treatment effect from the technical 'noise' (Kitchen et al., 2010), we need to either control extraneous factors such that they do not contribute to variance, or quantify the contribution that each of these factors makes to the total variance in the data (Kuehl, 2000). To partition the components of experimental variance, differences in family biodiversity between and within sub-samples and replicates at different levels of the

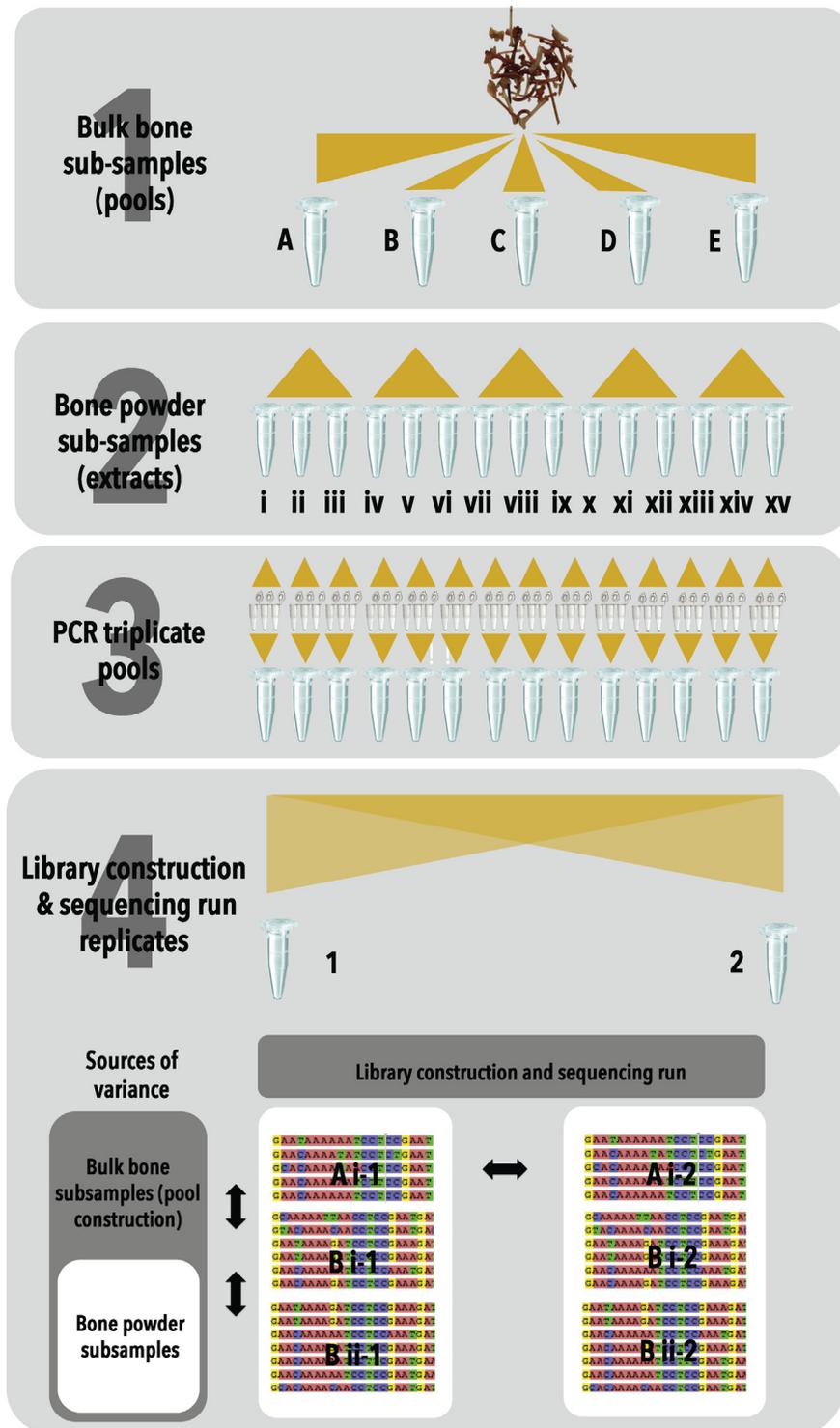
experimental protocol was compared using a nested experimental design (Fig. 1; SI II). Differences in taxonomic assemblages between replicate sequencing runs within an extract allows us to determine the amount of technical variance that can be attributed to library preparation and sequencing run; comparing differences between extraction replicates within a bulk bone sub-sample allows us to determine the amount of variance that can be attributed to subsampling bulk bone powder during the extraction process; and comparing differences between bulk bone sub-samples (randomly drawn) allows us to determine the amount of biological variability within the greater bulk bone sample. Other aspects of the experimental protocol that could contribute to variance were controlled: all bulk bone powder sub-samples were extracted together such that the extraction process does not contribute to variance, and PCR amplification was replicated, but replicates were then combined—in this way the contribution of the PCR process to variance was accounted for albeit not quantified.

### 2.5. DNA extraction of bulk bones

The sieved and sorted bulk bone was sub-sampled, then ground into powder in a designated clean facility at Murdoch University, WA, Australia (SI Fig. II). All preparation surfaces and tools were cleaned with a solution of 10% bleach followed by 70% ethanol, and personal protective equipment (SI III) was worn, in order to minimise contamination of samples with exogenous DNA. DNA from three, 100 mg aliquots of bone powder from each 50-bone sub-sample were extracted by incubating the powder in 1.5 ml of digest buffer containing final concentrations of 0.1% Triton-X-100, 0.02 M Tris-HCl, 1 mg/ml Proteinase K powder, 0.01 M DTT in 0.5 M EDTA for 24 h at  $55^{\circ}\text{C}$ . The supernatant was collected, concentrated to a volume of 50  $\mu\text{l}$  in a 30,000 MWCO Vivaspin-500 column, and purified using a modified QIAGEN protocol (SI III), eluting in 50  $\mu\text{l}$  EB buffer (QIAGEN, cat. No. 19066). One extraction control was included per 10 extractions. Extractions and subsequent qPCR reactions were prepared in a separate ultra-clean facility to avoid contamination.

### 2.6. Next-generation sequencing of DNA amplicons

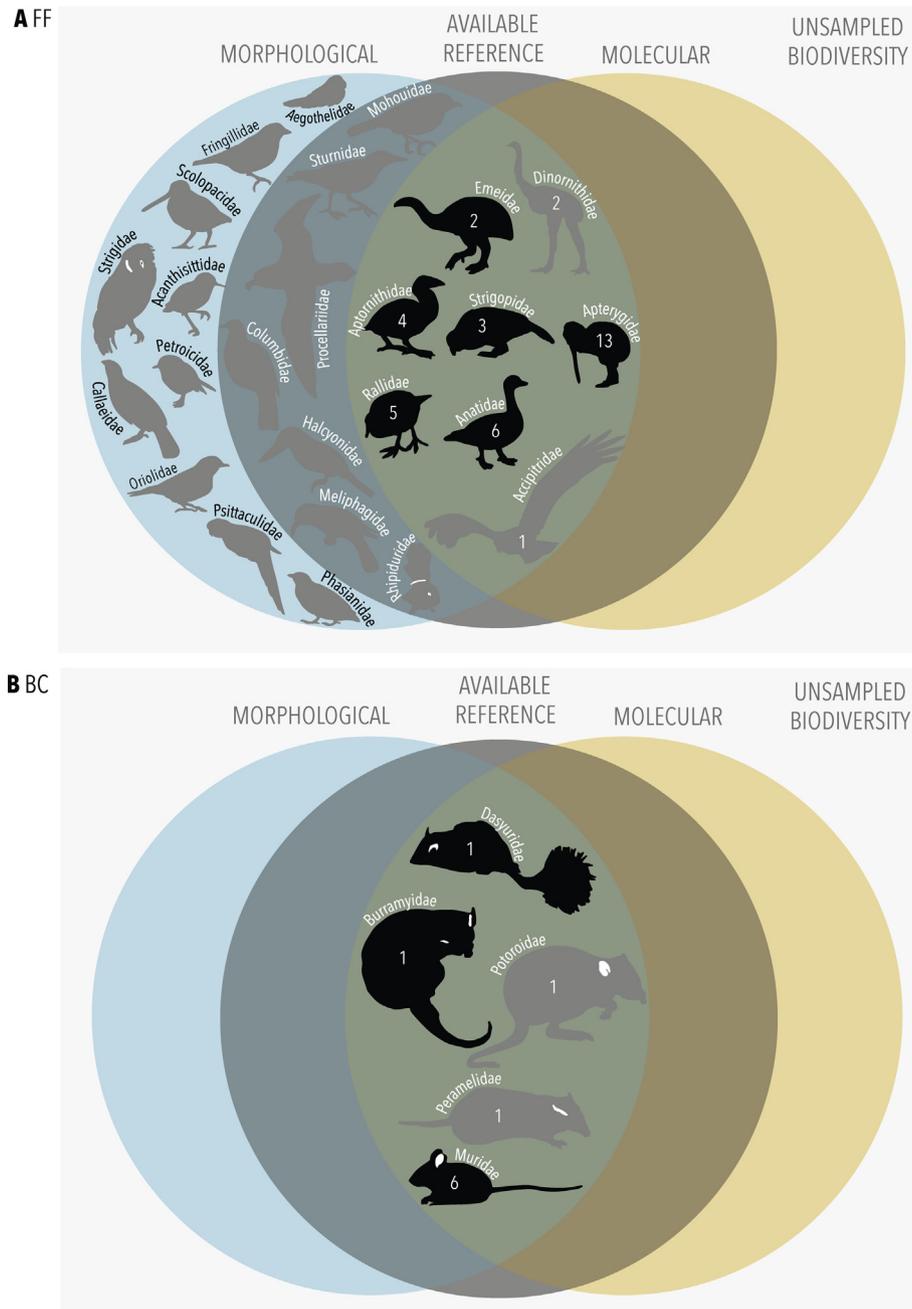
Following the protocol outlined in Murray et al. (2013), extracts were diluted and screened for inhibition and amplification efficiency via qPCR (SI IV). Dilutions exhibiting the least inhibition and greatest amplification efficiency were then amplified in triplicate via qPCR using indexed primers specific for a diagnostic barcoding region of the 16S rRNA gene in mammals for the BC samples (131 bp or 91 bp *sans* primer), and 12S rRNA gene (274 bp or 232 bp *sans* primer) in birds for the FF samples (SI IV, SI Table 1). These meta-barcodes were chosen according to the predominant class of organism identified morphologically at each site (aves at FF, mammals at BC) but without prior knowledge of the species within, and are preferred over standard barcoding genes such as CO1 or cytb due to their universality, the taxonomic resolution of the amplified insert, and small size (Deagle et al., 2014; Pedersen et al., 2014; Thomsen and Willerslev, 2015). 25  $\mu\text{l}$  qPCR reactions consisted of final concentrations of 0.4 mg/ml Bovine Serum Albumin, 1X ABI GeneAmp PCR Buffer, 2 mM  $\text{MgCl}_2$ , 0.4  $\mu\text{M}$  forward and reverse primers, 0.25 mM dNTPs, as well as 0.25  $\mu\text{l}$  of AmpliTaqGold DNA polymerase and 0.6  $\mu\text{l}$  of 1:2000 SYBR green, in HPLC-grade water (SI IV). Thermocycling conditions included heat denaturation at  $95^{\circ}\text{C}$  for 5 min, followed by 50 cycles of  $95^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 45 s, and a  $+1^{\circ}\text{C}$  melt curve at  $95^{\circ}\text{C}$  for 15 s,  $60^{\circ}\text{C}$  for 1 min,  $95^{\circ}\text{C}$  for 15 s, with a final  $72^{\circ}\text{C}$  extension for 10 min (SI IV). Triplicates were combined and purified using an Agencourt AMPure XP PCR purification kit according the manufacturer's



**Fig. 1.** Nested experimental design with replication at 3 stages; bulk bone sub-sampling or pool construction (experimental replicates), bone powder sub-sampling (extraction), and library construction and sequencing run (technical replicates). 5 different bone-powder pools were constructed by randomly drawing 50 bones from the same bulk bone sample. 3 extractions were performed on each pool by sub-sampling 100 mg of bone powder from each pool. For each site, a sequencing library was constructed containing a blend of the amplicons generated from each extract, and this library was sequenced twice. Sources of error can be partitioned by examining the differences within and between sequencing replicates (e.g., sequencing run 1 and 2), extraction replicates (e.g., Bi and Bii), and bone powder pools (e.g., A and B).

instructions, and samples were pooled in approximately equimolar concentrations twice (SI V) to create two sequencing libraries per primer set (Fig. 1). Extraction controls, three negative controls (water), and a positive control were included for each qPCR. All post-PCR methods were performed in a physically separated

laboratory in keeping with standard aDNA practice (Willerslev and Cooper, 2005; Shapiro and Hofreiter, 2012). The absolute concentration of sequencing libraries was quantified via qPCR as per Murray et al. (2012) (SI I) to determine how much to add to the sequencing reaction. Emulsion PCR and enrichment was performed



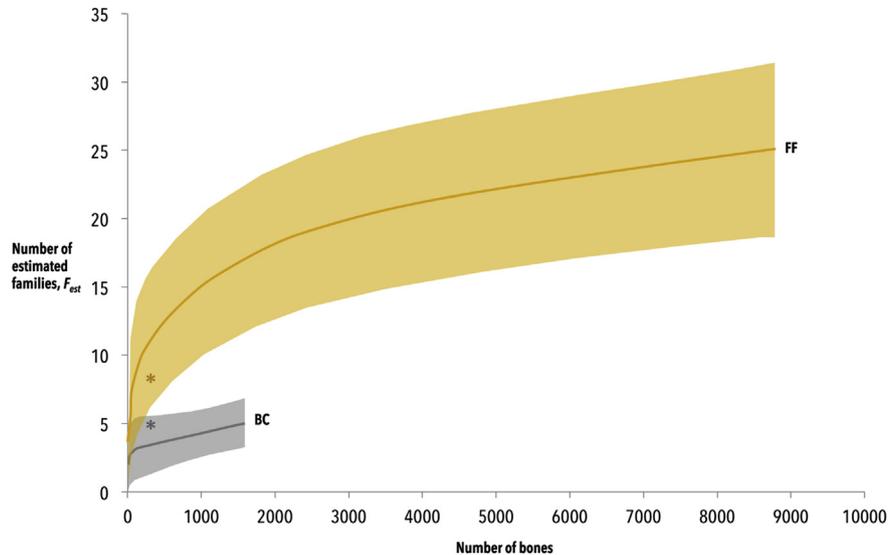
**Fig. 2.** Venn diagram showing the families detected morphologically (blue) and molecularly (yellow) for (A) FF, and (B) BC. The grey circle represents families where the meta-barcoding region was present in GenBank for the species identified morphologically. Taxa shaded black had an abundance >1% of the fossil assemblage, and taxa shaded grey had an abundance <1% of the fossil assemblage, and would therefore not be expected to be detected molecularly. The numbers inside the silhouettes represent the number of molecular OTUs per family. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by following the manufacturer's instructions for the Ion PGM Template OT2 200 kit (for both BC libraries), and the Ion PGM Template OT2 400 kit (for both FF libraries). Each library was sequenced separately using Ion PGM Sequencing 200 v2 kits and 400 kits as per the manufacturer's instructions.

### 2.7. Sequence identification and bioinformatics analysis

After sequencing, the FastQ file was downloaded and imported into *Geneious 7.0.6* (<http://www.geneious.com>, Kearse et al., 2012), sorted by barcode and primer sequence (with only exact matches accepted), and trimmed of all primers using methods in Murray

et al., 2013) with minor changes (SI VI). Quality control, chimera filtering, and abundance filtering were performed using several packages within *Galaxy* ([usegalaxy.org](http://usegalaxy.org); Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010) and *USEARCH* v. 6.1 (Edgar, 2010; Edgar et al., 2011) as implemented in *QIIME* v. 1.8.0 ([qiime.org](http://qiime.org); Caporaso et al., 2010b) using methods in Murray et al. (2013) with minor changes (SI VI); namely, reads with an average quality score less than 25, as well as chimeric reads, and low-abundant reads (<0.1% of the total number of reads) were discarded. Taxonomic identification was achieved by using *YABI* ([cgg.murdoch.edu.au/yabi](http://cgg.murdoch.edu.au/yabi); Hunter et al., 2012) to compare the sequences to NCBI's GenBank (Benson et al., 2006) nucleotide



**Fig. 3.** Rarefaction curves showing how many families would have been detected if fewer bones had been sampled morphologically for both FF (yellow) and BC (grey). Shading represents 95% confidence intervals (*EstimateS*; Colwell, 2013). Asterisks represent how many families were detected through BBM of 250 bones from FF (yellow) and BC (grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reference database via BLASTn (Altschul et al., 1990; no low complexity filter, gap penalties existence of 5 and extension 2 (default), e-value  $<1e-10$ , word size 7). BLAST results were imported into MEGAN 4.70.4 ([ab.inf.uni-tuebingen.de/data/software/megan4](http://ab.inf.uni-tuebingen.de/data/software/megan4); Huson et al., 2007) for taxonomic assignment (SI VI): families were considered present if the at least 95% of the query aligned to a known member of the family with a similarity of 90% or more. In order to estimate overall genetic diversity at the 3% level within each family, sequences within each identified family were then subjected to taxonomy-independent (OTU; operational taxonomic unit) analyses by clustering sequences at 97% identity using USEARCH v. 6.1 (Edgar, 2010; Edgar et al., 2011) as implemented in QIIME v. 1.8.0 ([qiime.org](http://qiime.org); Caporaso et al., 2010b) (SI VI). 97% identity was used to account for sequencing errors.

## 2.8. Statistical analysis

We restricted the scope of our analyses to quantifying the variance around the detection (presence/absence) of families and OTUs, and comparing their detection between morphological and molecular datasets generated from our study sites. As morphological and molecular techniques suffer different biases (Fig. 4; SI IX), fossil abundance does not correlate with read abundance (Bohmann et al., 2014; SI VIII.I, SI Table III). Therefore, it is not valid to compare abundance between morphological and molecular datasets. Similarly, compared to morphological analysis, it is difficult to make highly credible species- and genus-level molecular identifications. Therefore the morphological and molecular datasets are comparable only at the family level.

A nested non-parametric multivariate analysis of variance (nested NPMANOVA or PERMANOVA) was used (<https://www.stat.auckland.ac.nz/~mja/Programs.htm>; Anderson, 2001) to analyse the variance in the detection of families and OTUs from sequence data attributable to biological sub-samples and experimental replicates (SI VII, SI Table II). Sørensen's index of similarity (Sørensen, 1948) was used to compare the diversity of families (presence/absence) between morphological and molecular data sets for both sites (calculated by hand; SI VII). We also calculated the percentage of families we would have expected to identify that were actually identified through BBM: these were families that represent  $>1\%$  of

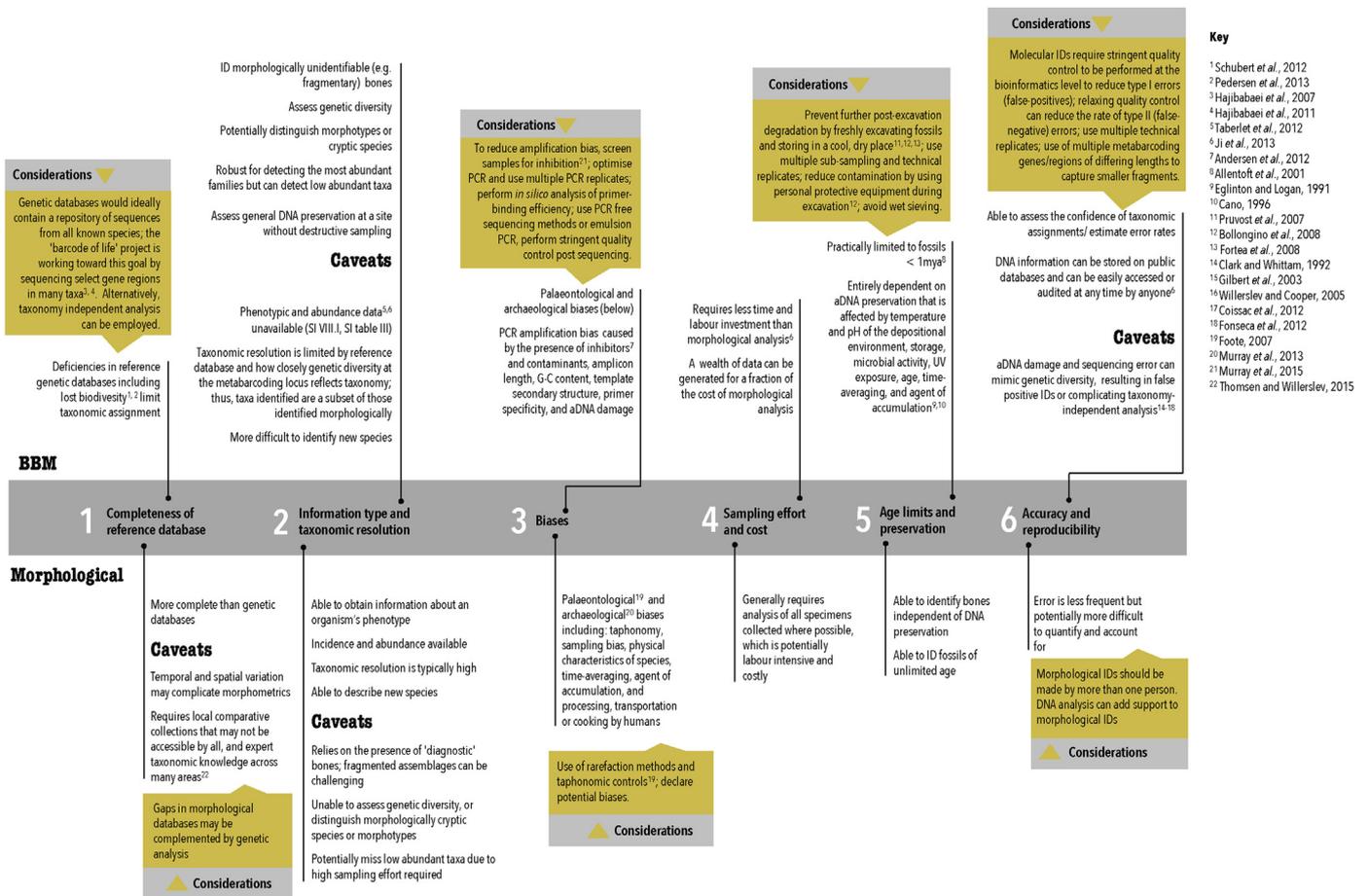
the fossil record (i.e., had a 95% chance of one representative bone being sampled in 250), that have at least one of the species identified within that family represented in GenBank for the metabarcoding gene used. In addition, *EstimateS* v. 9.1.0 ([purl.oclc.org/estimateS](http://purl.oclc.org/estimateS); Colwell, 2013) was used to generate rarefaction curves in order to examine how differences in sampling depth between morphological and molecular samples influenced the number of families that could be identified.

## 3. Results and discussion

5 sub-samples of 50 bones were extracted three times, independently amplified for one of two mitochondrial gene regions, and sequenced twice to yield, on average, a total of 27,722 and 72,639 sequence reads per run, with an average of 1848 and 4842 sequence reads (SI Table IV) per sample for FF and BC, respectively. 8 and 5 families were identified molecularly, with on average 4 and 2 OTUs per family, for FF and BC, respectively (Table 1; SI Table V). These sequence data were used to investigate the repeatability, accuracy, and overall utility of the BBM method.

### 3.1. Repeatability of the BBM method

We found that technical errors in the BBM method contribute similarly to the variance between widely different samples at both the family level and OTU level. Library construction and sequencing run contribute little to the variance at the family level, for both sites (8.1% FF, 2.5% BC; Table 1). Therefore, if family-level resolution is required, library construction and sequencing run may not need to be replicated, particularly on less error-prone sequencing platforms: this occurs despite the differences in sequencing depth between run replicates (SI Table IV), indicating that the amount of coverage has little impact on repeatability as we capture the same taxonomic assemblages with a small amount of coverage as we do with high coverage (Caporaso et al., 2010a). However, library construction and sequencing run contributed more to the variance at the OTU-level for both sites (20.2% FF, 24.3% BC; Table 1). The higher variance observed at the OTU-level compared with variance at the family-level may simply reflect how nested taxonomic levels (e.g., species or OTUs) are inherently more diverse than the level they are



**Fig. 4.** A comparison of the strengths and limitations of both morphological and molecular taxonomic identification of fossils. Detailed discussion of these caveats can be found in the [Supplementary Information \(SI IX\)](#).

nested within (e.g., families). Consequently, we will more consistently sample families than the taxa within them (Footo and Miller, 2007), resulting in a lower variance at the family level. In addition, OTU diversity is also notoriously sensitive to aDNA damage, PCR and sequencing errors, as well as sequencing depth, both of which can unpredictably inflate diversity (Bragg et al., 2013; Murray et al., 2015). Therefore, sequencing run replication and coverage become extremely important, albeit costly, at the OTU level as it contributes to about a quarter of the variance in OTU diversity. The implementation of stringent quality control protocols during sequence data filtering, as well as sub-sampling (i.e., rarefying) sequences to account for differences in coverage may also help eliminate variance in OTU diversity (Murray et al., 2013, 2015). These results also suggest that high diversity samples will require significantly more replication at all levels in order to consistently recover that diversity.

Although library preparation and sequencing run appears to make similar contributions to variance, each sample may require specific optimisation to determine the best sampling strategy to account for the variability within bulk bone samples and extract sub-samples: for FF, the greatest contribution to variance was from extract sub-sample nested within bulk bone sub-sample pool (58.9% for family, 61.0% for OTU; Table 1; SI VIII.II, SI Fig. III), while for BC, most of the variability is introduced when sub-sampling the bulk bone sample to form pools (98.7% for family, 75.8% for OTU; Table 1; SI VIII.II, SI Fig. III). Similarly, pools contributed less to the variance for FF (33.0% for family, 18.4% for OTU; Table 1; SI VIII.II, SI

Fig. III), while extract sub-sample contributed less to the variance at BC (2.5% for family, <0.1% for OTU; Table 1; SI VIII.II, SI Fig. III). Thus, sub-sampling needs to be concentrated at different levels in different samples, and this is likely to be largely dictated by the diversity of the sample. Finally, total variance within FF and BC at the family-level is small and is about the same (standard deviation = 0.36 Jaccard dissimilarity; Table 1); however, to test whether the overall repeatability, or reproducibility, of the BBM method is truly universal would require a survey of many more sites.

Nevertheless, sub-sampling and replication at various steps in the bulk bone method will help reduce error variance and increase the power to detect significant temporal and spatial changes in biodiversity. That is, multiple sub-sampling and replication is essential to be able to distinguish whether the differences we observe in biodiversity between samples are real, or simply a result of insufficient sampling, aDNA damage, and technical error (SI VIII.V). However, the cost-efficacy of the BBM method would be lost by sampling beyond what is necessary. Knowledge of the contributions to variance quantified by a pilot study can help determine how much sampling and replication would be required to obtain enough statistical power to detect a predetermined effect size significantly (if one exists). Resource allocation can be further optimised by factoring in cost per replicate and total budget into sample size considerations (Kitchen et al., 2010).

**Table 1**

Statistics table, including: nested PERMANOVA (4999 permutations, Jaccard dissimilarity or distance) showing the contributions to variance in the molecular family assemblages (presence/absence; bold), and the OTU assemblages (presence/absence) of FF and BC (asterisks indicate significance or  $\alpha < 0.05$ ); Standard Deviation (St.Dev., Jaccard dissimilarity); Sørensen's index of similarity between morphological and molecular family assemblages; the percentage of expected (> 1% of the fossil record with reference sequence present in GenBank) and unexpected (< 1% of the fossil record or with no reference sequence present in GenBank) families detected molecularly; the number of families detected morphologically had only 250 bones been sampled.

Site	Source of variance	DF	Contribution to variance (%)	St.Dev (Jaccard dissimilarity)	Sørensen's Index of similarity	Expected families detected molecularly (%)	Unexpected families detected molecularly (%)	Families detected morphologically if 250 bones were sampled ( $\pm 95\%$ CI)
Finsch's Folly	Pool	4	<b>33.0*</b> 18.4*	0.36	0.48	100 (6 out of 6)	10.5 (2 out of 19)	10.40 $\pm$ 4.76 (23–61%)
	Extract (Pool)	10	<b>58.9*</b> 61.0*					
	Library construction and sequencing run (residual)	15	<b>8.1</b> 20.6*					
Bat Cave	Pool	4	<b>95.0*</b> 72.0*	0.36	1.00	100 (3 out of 3)	100 (2 out of 2)	3.31 $\pm$ 1.15 (43–90%)
	Extract (Pool)	10	<b>2.5*</b> <0.1					
	Library construction and sequencing run (residual)	15	<b>2.5</b> 28.0					

### 3.2. Accuracy of the BBM method

In order to assess how accurately the taxa in the bulk bone assemblage that were identified genetically reflect the taxa in the fossil assemblage that were identified morphologically, we compared family biodiversity from FF and BC detected by the two methods. We found that the families identified via the BBM method are largely consistent with those identified through morphological analysis of the fossil record (Fig. 2; SI Table V). For both sites, the families identified molecularly are a subset of the families identified morphologically: for FF, 25 avian families were identified through morphology, with 8 of these (32%) also identified through BBM, while for BC, 5 mammalian families were identified morphologically with all (100%) also identified through BBM. This is likely to be a consequence of sampling bias and sampling effort, DNA preservation, and the presence (or absence) of a reference sequence in GenBank (Fig. 4). Several studies on plant family diversity, whether assessed using pollen and macrofossils, or sedimentary ancient DNA (*sedDNA*) and metabarcoding, reported similar results (Jørgensen et al., 2011; Pedersen et al., 2013), but also suggest *sedDNA* tends to reveal different and less diversity than these alternative methods (Boessenkool et al., 2013; Parducci et al., 2014). In this way the BBM method may be more reflective of the fossil record than *sedDNA*, but this then raises the question: why weren't all the families found the fossil record detected by BBM?

Assuming that the abundance of each family in bulk bone sample is similar to the morphological sample, we would not expect rare families (families that make up less than 1% of the total number of fossils) to have been represented in a sub-sample total of 250 bone fragments. When we look at the percentage of families we would have expected to sample, we detected 100% of expected families in FF and BC (Table 1, Fig. 2). As such, the BBM method is robust for detecting the most abundant families. Indeed, for both sites, the most abundant families (Anatidae and Muridae, from FF and BC, respectively) are the ones most consistently recovered. On the other hand, we detected families that were not expected to be detected due to their low abundance in the fossil record (Table 1); 10.5% of unexpected families were detected at FF (i.e., Dinornithidae and Accipitridae; SI Table V), and 100% of unexpected families detected at BC (i.e., Peramelidae, Potoroidae). This shows that the method is potentially sensitive enough to detect at least some rare taxa, but that this ability is sporadic. Furthermore, the diversity of the sample will influence the number of rare taxa present and the frequency of the rarest taxa, which may explain why half the families are shared between the morphological and molecular assemblages for FF (Sørensen index of similarity = 0.48; Table 1), while the morphological and molecular family assemblages are identical for BC (Sørensen index of similarity = 1; Table 1): the number of rare families in the more diverse FF (that also have a GenBank reference) is 9 with the abundance of the rarest families (5) being 0.01%, whereas the number of rare families (that also have a GenBank reference) in the less diverse BC is 2, with the abundance of the rarest families (2) being 0.06% (SI Table V). This may also explain why a smaller percentage of unexpected families were detected at FF compared with BC.

In reality, the abundance of each taxon in a bulk bone sample may not be the same as the morphological sample, and the extent of this difference will differ between sites. For instance, bulk bone samples are sub-samples of all the fossils collected that are in turn a sub-sample of the total fossils—as such, rare taxa in the fossil record can become rarer still in bulk bone sample, and with each successive sub-sampling the likelihood of rare species being lost increases. In addition, rare diagnostic bones may be removed from the bulk bone collection prior to DNA analysis, further biasing the sample towards abundant taxa (SI IX.1). Alternatively, some bones

may be broken into several fragments, increasing their chance of being sampled, while others may not. On top of this, both DNA preservation that differs between bones, PCR bias (Fig. 4), and deficiencies in reference databases (SI IX) will further affect the ability of taxa to be detected genetically, regardless of abundance. For example, post-mortem DNA degradation results in highly fragmented sequences that may not be long enough for both primers of the pair to bind and amplify, resulting in taxa with poor preservation not being detected. Again, this may explain why some rare families were detected molecularly and others were not, especially in FF where the amplified region was considerably longer than that of BC. Amplifying shorter metabarcoding regions, using multiple metabarcoding primers, using species-specific primers and increasing the sampling effort may also help mitigate some of the biases in order to recover more taxa, more consistently (SI XI). However, there remains a trade-off between amplicon length and taxonomic discrimination, and the use of multiple primer sets and cost. Furthermore, although increasing the sampling effort will improve the repeatability and may result in the detection of more families (especially samples of high diversity), this will plateau as a consequence of the principle of diminishing returns; the point of sampling 'saturation' is best determined experimentally.

In comparison to morphological methods, the same amount of sampling or less is required to detect the same number of families using bulk bone analysis. Rarefaction can be used to estimate the number of families that would have been identified morphologically had fewer bones (250) been examined morphologically (Fig. 3; Foote and Miller, 2007; Wooley et al., 2010). This was compared with the number of families detected by DNA analysis of the same number of bones. For FF,  $10.40 \pm 4.76$  (95% CI) families (23–61%) would have been detected if 250 bones were sampled as opposed to 25 families from a sample of 8771 bones—this is not significantly different from the percentage of families that were detected using the bulk bone method at the same level of sampling (32%, or 8 out of 25 families; Fig. 2A). For BC,  $3.31 \pm 1.15$  (95% CI) families would have been detected if 250 bones were sampled as opposed to 5 families from 1600 bones (43–90%); in comparison, we detected all of the families from BC via DNA analysis of 250 bulk bones (100%, or 5 out of 5 families; Fig. 2B). In this way, the BBM method can generate as much data as morphological approaches with the same amount of sampling, potentially in a fraction of the time.

### 3.3. Overall utility and further applications of the BBM method

Although morphological identifications can typically achieve higher taxonomic resolution (species) than bulk bone metabarcoding can (family/genus level) at present (especially given the incomplete nature of the reference genetic databases (SI IX.IV); Fig. 4), we were able to confidently make some species-level assignments. Here, we define a species-level match as >98% similarity to a reference sequence across 100% of the amplicon length, with all other species within the genus present in GenBank, and equal similarity to no other species. For FF, we detected the extinct species *Dinornis robustus* (South Island giant moa), and the highly endangered species *Strigops habroptilus* (Kākāpō), while for BC, we detected the vulnerable and locally extirpated species, *Phascogale tapoatafa* (Brush-tailed phascogale; SI Table V). This shows that, like morphological methods and environmental DNA (eDNA) metabarcoding methods, the BBM approach has the potential for monitoring changes in biodiversity over time, which may be important for the management of already critically endangered species, or for identifying species that may have future risk of extinction (Boessenkool et al., 2013; Bohmann et al., 2014; Pansu et al., 2015). Conservation can potentially also be aided by

estimates of genetic diversity within species (McDowell, 2014) that can be obtained using BBM. For example, although there is only one species of Kākāpō within the genus *Strigops*, three distinct genetic variants (i.e., OTUs) (that cannot be accounted for by aDNA damage or error) at the locus studied were detected (SI Table V), and it is possible that this indicates the presence of higher historical genetic diversity, or even several unknown sub-species of *Strigops*. However, because metabarcodes are designed to minimise intra-specific variability, high-resolution estimates of genetic diversity within species (such as haplotype diversity) need to be obtained through the amplification of several specific molecular markers. Nevertheless, it would not be possible to glean this type of genetic information using morphological approaches, and shows how the BBM method, like eDNA metabarcoding methods, could also potentially help detect taxa that have not been formerly described in the fossil record (Krause et al., 2010; Murray et al., 2013; Parducci et al., 2014). This could assist in more accurately determining the timing extinction events (Haile et al., 2009; Pedersen et al., 2013; Willerslev et al., 2014) or species' former distributions. Similarly, for BC, we detected a taxon with 98% similarity to several species of bandicoot belonging to the genus *Isoodon*; however, this genus was not identified in the morphological analyses, which may indicate additional diversity within Peramelidae. Alternatively, the *Isoodon* taxon may represent the indeterminate bandicoot species identified morphologically but assigned to the genus *Perameles* (SI Table V). This interpretation could be supported by the fact that several *Perameles* voucher specimens have references in GenBank but were not the closest match, and only one OTU was identified for this family. In this way, genetic analysis may add weight to or challenge morphological identifications, particularly if specimens are morphologically ambiguous (e.g., *Macropus fuliginosus* and *M. giganteus*, or *Pseudochirus peregrinus* and *P. occidentalis*), cryptic species (using high resolution molecular markers), or tiny post-cranial bones that are typically difficult to distinguish at the species level (e.g., fish, reptiles, amphibians). Finally, the BBM method (as well as the use of species-specific primers on bulk bone samples) may offer a more sensitive tool for detection of invasive species (Bohmann et al., 2014), which may improve estimates of the timing of major historical events that have impacted biodiversity, such as human arrival, or the presence or absence of land bridges (e.g., land bridges connecting Kangaroo Island to the Australian mainland) (Pansu et al., 2015).

## 4. Conclusion

As with morphological methods of identifying fossil bones, BBM has intrinsic biases that make it difficult to identify every taxon present in the fossil assemblage (Fig. 4; SI IX). The biases and limitations of the BBM method need to be considered when designing experiments that seek to apply it (Fig. 4; SI X). Weighing up the advantages and disadvantages of the BBM method for the taxonomic identification of fossils relative to traditional morphological taxonomic identification, we propose some practical recommendations for palaeontologists, archaeologists, and geneticists who are eager to apply the BBM method (Fig. 4; SI X and SI XI). Essentially, it is important to collect and sub-sample bulk bone material following a predetermined, statistically-optimised experimental design with clearly-defined questions. Many limitations of the method can be overcome through greater sub-sampling and replication at the most error-prone stages in order to quantify and minimise the collective contribution of these biases to variance in the results. Therefore, it is advisable that, when possible, a small pilot study with a hierarchical experimental design be employed to experimentally determine to what extent, and where, such effort should be concentrated. Ultimately, the way the method is applied

will depend on the desired level of resolution and the type of questions being addressed. There also remains room for development in shotgun sequencing (Taberlet et al., 2012; Smith et al., 2015; SI IX.V, SI XII.II), targeted (including enrichment capture or the use of species-specific primers) (Shokralla et al., 2012; Taberlet et al., 2012; Thomsen and Willerslev, 2015; SI XII.III), and PCR-free approaches (Ji et al., 2013), as well as sample preparation protocols (SI XII.I). Although genetic analysis cannot currently achieve the taxonomic resolution achievable using morphological analysis, this will improve with access to more comprehensive DNA reference material in the future. Metabarcoded samples also allow many diverse taxa to be identified simultaneously, and in this way, BBM can be rapid, and less reliant on regional taxonomic expertise than morphological analysis. For these reasons, the BBM method holds great potential as cost-effective technique for identifying taxonomic assemblages, estimating genetic diversity, and assessing general aDNA preservation, at a range of sites from bones that were once considered disposable. Furthermore, the BBM has applications for the detection of cryptic or morphologically undiscoverable taxa, historical faunal turnover, and population fragmentation, which may provide a powerful tool for species conservation (McDowell, 2014). Although both morphological and molecular methods have limitations, a combination of these approaches gives greater insight into past biodiversity, as the limitations of one are complemented by the other. We believe this study will help bridge the gap between disciplines, and hope that future development of the aDNA methods continues to integrate palaeontological, ecological, and genetic principles, in order to overcome the limitations that remain.

#### Author contributions

AG designed experiments and wrote the manuscript, with methods written by AG, MM, DF, and PS. All authors contributed to editing of the manuscript. AG, DM, MM, MB assisted with analysis. Samples were collected by MM, DM, PS, and GP. Morphological identification was carried out by MM, DF, PS.

#### Author declarations

All necessary permits were obtained for the described study, which complied with all relevant regulations.

The authors declare no competing interests.

#### Acknowledgements

Work was funded by Australian Research Council grants (DP120104435). MB and GP were supported in this research by an ARC Future Fellowships FT0991741 and FT130101728. We thank the Pawsey supercomputer facility for assistance in data analysis. PS wishes to thank K. Davis and Y.P. Montelle for help with FF collection and analysis, and J. Woods for facilitating AMS dating of FF.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.quascirev.2015.09.014>.

Sequencing data has been deposited on the online data repository DataDryad, and is available at: <http://dx.doi.org/10.5061/dryad.gt377>.

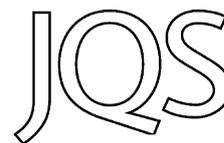
#### References

Allentoft, M.E., Heller, R., Oskam, C.L., Lorenzen, E.D., Hale, M.L., Gilbert, T.M.P., Jacomb, C., Holdaway, R.N., Bunce, M., 2014. Extinct New Zealand megafauna were not in decline before human colonization. *Proc. Natl. Acad. Sci.* 111,

- 4922–4927.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Andersen, K., Bird, K.L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K.H., Orlando, L., Gilbert, M.T.P., Willerslev, E., 2012. Meta-barcoding of ‘dirt’ DNA from soil reflects vertebrate biodiversity. *Mol. Ecol.* 21, 1966–1979.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2006. GenBank. *Nucleic Acids Res.* 34, D16–D20.
- Binladen, J., Gilbert, M.T.P., Bollback, J.P., Panitz, F., Bendixen, C., Nielsen, R., Willerslev, E., 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. In: Ausubel, Frederick M., et al. (Eds.), *Current Protocols in Molecular Biology*, 0 19, Unit 19.1021.
- Boessenkool, S., McGlynn, G., Epp, L.S., Taylor, D., Pimentel, M., Gizaw, A., Nemomissa, S., Brochmann, C., Popp, M., 2013. Use of ancient sedimentary DNA as a novel conservation tool for high-altitude tropical biodiversity. *Conserv. Biol.* 28, 446–455.
- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Ecol. Evol.* 29, 358–367.
- Bragg, L.M.S., Glenn, Butler, Margaret K., Hugenholtz, Philip, Tyson, Gene W., 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 9.
- Bronk Ramsey, C., 2010. *OxCal*, v. 4.1.7.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2010a. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* 108, 4516–4522.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.
- Colwell, R.K., 2013. EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples. Version 9. Available: [purl.oclc.org/estimates](http://purl.oclc.org/estimates).
- Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P., 2014. DNA metabarcoding and the cytochrome oxidase subunit 1 marker: not a perfect match. *Biol. Lett.* <http://dx.doi.org/10.1098/rsbl.2014.0562>.
- Deméré, T.A., Berta, A., McGowen, M.R., 2005. The taxonomic and evolutionary history of fossil and modern balaenopteroid mysticetes. *J. Mammalian Evol.* 12, 99–143.
- Donoghue, M.J., Doyle, J.A., Gauthier, J., Kluge, A.G., Rowe, T., 1989. The importance of fossils in phylogeny reconstruction. *Annu. Rev. Ecol. Syst.* 20, 431–460.
- Earp, M.A., Rahmani, M., Chew, K., Brooks-Wilson, A., 2011. Estimates of array and pool-construction variance for planning efficient DNA-pooling genome wide association studies. *BMC Med. Genom.* 4.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461.
- Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200.
- Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erséus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A., Stenøien, H.K., Hassel, K., Kausrud, H., Yoccoz, N.G., Bräthen, K., Willerslev, E., Taberlet, P., Coissac, E., Brochmann, C., 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21, 1821–1833.
- Epp, L.S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schröder-Nielsen, A., Ludikova, A., Hassel, K., Stenøien, H.K., Funder, S., Willerslev, E., Kjaer, K., Brochmann, C., 2015. Lake sediment multi-taxon DNA from north Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quat. Sci. Rev.* 117, 152–163.
- Fonseca, V.G., Nichols, B., Lallias, D., Quince, C., Carvalho, G.R., Power, D.M., Creer, S., 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Res.* 40.
- Foot, M., Miller, A.I., 2007. *Principles of Palaeontology*, third ed. W. H. Freeman and Company, New York City, NY.
- Fusco, D.A., 2014. The Biodiversity of the Holocene Terrestrial Mammal Faunas of the Murray Mallee. Biological Sciences (Honours) Unpublished Honours thesis. Flinders University, South Australia.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Eltniski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A., 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455.
- Goetsch, J., Nekrutenko, A., Taylor, J., The Galaxy Team, 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86.
- Haile, J., Froese, D.G., MacPhee, R.D.E., Roberts, R.G., Arnold, L.J., Reyes, A.V., Rasmussen, M., Nielsen, R., Brook, B.W., Robinson, S., Demuro, M., Gilbert, M.T.P., Munch, K., Austin, J.J., Cooper, A., Barnes, I., Möller, P., Willerslev, E., 2009. Ancient DNA reveals late survival of mammoth and horse in interior Alaska.

- Proc. Natl. Acad. Sci. U. S. A. 106, 22352–22357.
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C., Baird, D.J., 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river Benthos. *PLoS One* 6.
- Haouchar, D., Haile, J., McDowell, M.C., Murray, D.C., White, N.E., Allcock, R.J.N., Phillips, M.J., Prideaux, G.J., Bunce, M., 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–Holocene cave deposit on Kangaroo Island, South Australia. *Quat. Sci. Rev.* 84, 56–64.
- Hofreiter, M., Betancourt, J.L., Sbriller, A.P., Markgraf, V., McDonald, H.G., 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Curá, Neuquén Province, southwest Argentina. *Quat. Res.* 59, 364–378.
- Hogg, A.G., Hua, Q., Blackwell, P.G., Niu, M., Buck, C.E., Guilderson, T.P., Heaton, T.J., Palmer, J.G., Reimer, P.J., Reimer, R.W., Turney, C.S.M., Zimmerman, S.R.H., 2013. SHCAL13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* 55, 1–15.
- Hunter, A.A., Macgregor, A.B., Szabo, T.O., Wellington, C.A., Bellgard, M.I., 2012. Yabi: an online research environment for grid, high performance and cloud computing. *Source Code Biol. Med.* 7, 1.
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386.
- Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C., Yu, D.W., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* 16, 1245–1257.
- Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N.H., Andersen, K., Orlando, L., Gilbert, M.T.P., Willerslev, E., 2011. A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* 21, 1989–2003.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thieler, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic Acids Res.* 40.
- Kitchen, R.R., Kubista, M., Tichopad, A., 2010. Statistical aspects of quantitative real-time PCR experiment design. *Methods* 50, 231–236.
- Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., Pääbo, S., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, 894–897.
- Kuehl, R.O., 2000. *Design of Experiments: Statistical Principles of Research Design and Analysis*, second ed. Brooks/Cole Publishing Company, Pacific Grove, CA.
- Larson, S., Jameson, R., Etnier, M., Fleming, M., Bentzen, P., 2002. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Mol. Ecol.* 11, 1899–1903.
- Macgregor, S., 2007. Most pooling variation in array-based DNA pooling is attributable to array error rather than pool construction error. *Eur. J. Hum. Genet.* 15, 501–504.
- Manos, P.S., Soltis, P.S., Soltis, D.E., Manchester, S.R., Oh, S., Bell, C.D., Dilcher, D.L., Stone, D.E., 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Syst. Biol.* 56, 412–430.
- McDowell, M.C., Baynes, A., Medlin, G., Prideaux, G., 2012. The impact of European colonization on the late-Holocene non-volant mammals of Yorke Peninsula, South Australia. *Holocene* 22, 1441–1450.
- McDowell, M., 2014. Holocene vertebrate fossils aid the management and restoration of Australian ecosystems. *Ecol. Manag. Restor.* 15, 58–63.
- Meyer, M., Stenzel, U., Myles, S., Prüfer, K., Hofreiter, M., 2007. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res.* 35.
- Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N., Kawabata, Z.i., 2012. Surveillance of fish species composition using environmental DNA. *Limnology* 13, 193–197.
- Murray, D.C., Pearson, S.G., Fullagar, R., Chase, B.M., Houston, J., Atchison, J., White, N.E., Bellgard, M.I., Clarke, E., Macphail, M., Gilbert, M.T.P., Haile, J., Bunce, M., 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quat. Sci. Rev.* 58, 135–145.
- Murray, D.C., Haile, J., Dortch, J., White, N.E., Haouchar, D., Bellgard, M.I., Allcock, R.J., Prideaux, G.J., Bunce, M., 2013. Scrapheap Challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Sci. Rep.* 3.
- Murray, D.C., Coghlan, M.L., Bunce, M., 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLoS One* 10, e0124671.
- Pansu, J., Giguet-Coxev, C., Ficetola, G.F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., Choler, P., 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Mol. Ecol.* 24, 1485–1498.
- Parducci, L., Väiliranta, Salonen, J.S., Ronkainen, T., Matetovici, I., Fontana, S.L., Eskola, T., Sarala, P., Suyama, Y., 2014. Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. *Philos. Trans. R. Soc. B* 370, 20130382.
- Pedersen, M.W., Ginolhac, A., Orlando, L., Olsen, J., Andersen, K., Holm, J., Funder, S., Willerslev, E., Kjær, K.H., 2013. A comparative study of ancient environmental DNA to pollen and macrofossils from lake sediments reveals taxonomic overlap and additional plant taxa. *Quat. Sci. Rev.* 75, 161–168.
- Pedersen, M.W., Overballe-Petersen, S., Ermini, L., Sarkissian, C.D., Haile, J., Hellstrom, M., Spens, J., Thomsen, P.F., Bohmann, K., Cappellini, E., Schnell, I.B., Wales, N.A., Carøe, C., Campos, P.F., Schmidt, A.M.Z., Gilbert, M.T.P., Hansen, A.J., Orlando, L., Willerslev, E., 2014. Ancient and modern environmental DNA. *Philos. Trans. R. Soc. B* 370, 20130383.
- Porter, T.M., Golding, G.B., King, C., Froese, D., Zazula, G., Poinar, H.N., 2013. Amplicon pyrosequencing late Pleistocene permafrost: the removal of putative contaminant sequences and small-scale reproducibility. *Mol. Ecol. Resour.* 5, 798–810.
- Rodríguez-Aranda, J.P., Calvo, J.P., 1998. Trace fossils and rhizoliths as a tool for sedimentological and palaeoenvironmental analysis of ancient continental evaporite successions. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 140, 383–399.
- Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R.E., Georgiadis, N.J., Roca, A.L., Hofreiter, M., 2010. Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* 8, 12.
- Shapiro, B., Hofreiter, M., 2012. *Ancient DNA: Methods and Protocols*. Humana Press (Springer Science), New York.
- Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. *Mol. Ecol.* 21, 1794–1805.
- Smith, O., Mombere, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V., Allaby, R.G., 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* 347, 998–1001.
- Sørensen, T., 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *K. Dan. Vidensk. Selsk.* 5, 1–34.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050.
- Thomsen, P.F., Kielgast, J., Iversen, L.L., Möller, P.R., Rasmussen, M., Willerslev, E., 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One* 7.
- Thomsen, P.F., Willerslev, E., 2015. Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18.
- Willerslev, E., Hansen, A.J., Binladen, J., Brand, T.B., Gilbert, M.T.P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D.A., Cooper, A., 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300, 791–795.
- Willerslev, E., Cooper, A., 2005. Ancient DNA. *Proc. R. Soc. B Biol. Sci.* 272, 3–16.
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S., Pearman, P.B., Cheddadi, R., Murray, D., Bräthen, K.A., Yoccoz, N., Binney, H., Cruaud, C., Wincker, P., Goslar, T., Alsos, I.G., Bellemain, E., Brysting, A.K., Elven, R., Sønstebo, J.H., Murton, J., Sher, A., Rasmussen, M., Rønn, R., Mourier, T., Cooper, A., Austin, J., Möller, P., Froese, D., Zazula, G., Pompanon, F., Rioux, D., Niderkorn, V., Tikhonov, A., Savvinov, G., Roberts, R.G., MacPhee, R.D.E., Gilbert, M.T.P., Kjær, K.H., Orlando, L., Brochmann, C., Taberlet, P., 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506, 47–51.
- Wooley, J.C., Godzik, A., Friedberg, I., 2010. A primer on metagenomics. *PLoS Comput. Biol.* 6.
- Zhang, X., Shi, G.R., Gong, Y., 2008. Middle Jurassic trace fossils from the ridang formation in Sajia county, South Tibet, and their palaeoenvironmental significance. *Facies* 54, 45–60.

# An assessment of ancient DNA preservation in Holocene–Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia



ALICIA GREALY,<sup>1</sup> AMY MACKEN,<sup>2</sup> MORTEN E. ALLENTOFT,<sup>3</sup> NICOLAS J. RAWLENCE,<sup>4</sup> ELIZABETH REED<sup>2,5</sup> and MICHAEL BUNCE<sup>1\*</sup>

<sup>1</sup>Trace and Environmental DNA(TrEnD) Laboratory, Department of the Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

<sup>2</sup>School of Biological Sciences, Flinders University, Bedford Park SA 5042, Australia

<sup>3</sup>Centre for GeoGenetics, Natural History Museum, University of Copenhagen, 1350, Denmark

<sup>4</sup>Allan Wilson Centre, Department of Zoology, University of Otago, Dunedin 9016, New Zealand

<sup>5</sup>School of Physical Sciences, The University of Adelaide, Adelaide SA 5000, Australia

Received 9 September 2015; Revised 30 October 2015; Accepted 19 November 2015

**ABSTRACT:** Although there is a long history of research into the fossil deposits of the Naracoorte Caves (South Australia), ancient DNA (aDNA) has not been integrated into any palaeontological study from this World Heritage site. Here, we provide the first evidence of aDNA preservation in Holocene- and Pleistocene-aged fossil bone from a deposit inside Robertson Cave. Using a combination of metabarcoding and shotgun next-generation sequencing approaches, we demonstrate that aDNA from diverse taxa can be retrieved from bulk bone as old as 18 600 cal a BP. However, the DNA is highly degraded and contains a lower relative proportion of endogenous sequences in bone older than 8400 cal a BP. Furthermore, modelling of DNA degradation suggests that the decay rate is rapid, and predicts a very low probability of obtaining informative aDNA sequences from extinct megafaunal bones from Naracoorte (ca. 50 000 cal a BP). We also provide new information regarding the past faunal biodiversity of Robertson Cave, including families that have not been formerly described in the fossil record from here before. Collectively, these data demonstrate the potential for future aDNA studies to be conducted on material from Naracoorte, which will aid in the understanding of faunal turnover in southern Australia.  
Copyright © 2016 John Wiley & Sons, Ltd.

**KEYWORDS:** ancient DNA; bulk bone metabarcoding; DNA preservation; Naracoorte Caves; palaeontology.

## Introduction

Vertebrate fossil assemblages are an important source of information about the timing and extent of past biodiversity change. The Naracoorte Caves in south-eastern South Australia are world renowned for their preservation of vertebrate fossils, which includes mammals, birds, reptiles and frogs. Some deposits across the site date from approximately 1000 to 500 000 years in age, while others still actively trap animals today (Prideaux *et al.*, 2007; Macken and Reed, 2013). The natural and scientific value of the caves is reflected in their World Heritage listing, as well as in ongoing palaeontological research that has contributed to our knowledge about taxonomy, the past biodiversity of southern Australia, and the effects of climate and associated habitat changes on the distribution of both extinct and living species (e.g. Prideaux *et al.*, 2007; Macken *et al.*, 2012; Macken and Reed, 2013; Macken and Reed, 2014). Research into the fossil deposits within the Naracoorte Caves World Heritage Area (NCWHA) has involved examination of the geochemical and physical properties of fossil bones themselves, as well as additional materials that preserve information about past ecosystems (i.e. faunal diet, local climatic conditions, etc.), such as sediments and cave formations (e.g. Ayliffe *et al.*, 1998; Forbes and Bestland, 2007; Macken, 2009). Newly discovered organic plant macro- and microfossils preserved in the Naracoorte caves (e.g. seeds, leaves, pollen and phytoliths) also provide direct evidence about past habitats and how these have changed over time in the south-east region of South Australia (Darrénougué *et al.*, 2009; Reed,

2012), and also suggest the possibility of biomolecule preservation at Naracoorte.

Ancient DNA (aDNA) isolated from fossil assemblages represents another powerful source of information about past biodiversity. aDNA refers to degraded, fragmented and chemically modified DNA that exists in trace amounts in material such as bones (Kuhn *et al.*, 2010), leaves (Jaenicke-Despres *et al.*, 2003), feathers (Rawlence *et al.*, 2009), eggshell (Oskam *et al.*, 2010), coprolites (Wood *et al.*, 2012) and sediment (Willerslev *et al.*, 2003; Haile, 2012; Rawlence *et al.*, 2014). aDNA can be used, for example, to estimate the effective population size and past genetic diversity of animal communities that is not available from the morphological study of fossil bones (Hadly *et al.*, 2004; O'Keefe *et al.*, 2009). Studies using aDNA have also addressed a wide range of research questions regarding taxonomy (e.g. Bunce *et al.*, 2003), phylogeny (e.g. Mitchell *et al.*, 2014), palaeodiet (e.g. Hofreiter *et al.*, 2003), palaeoclimate (e.g. Jørgensen *et al.*, 2012), climate change (e.g. Hadly *et al.*, 2004; Magyari *et al.*, 2011), population dynamics (e.g. Leonard *et al.*, 2002; Bunce *et al.*, 2009; Allentoft *et al.*, 2014) and interspecies relationships (e.g. Wood *et al.*, 2013a, 2013b). Furthermore, the use of aDNA has had implications for the conservation and management of endangered animal populations (Shepherd and Lambert, 2008). With advances in DNA sequencing techniques, aDNA is being extracted from increasingly older fossil material in increasingly poor preservation conditions (Orlando *et al.*, 2011; Murray *et al.*, 2012), including the relatively warm, moist microclimates of many Australian cave systems (Murray *et al.*, 2013; Haouchar *et al.*, 2014; Llamas *et al.*, 2014; Grealy *et al.*, 2015).

Despite sustained palaeontological interest in the Naracoorte Caves (e.g. Wells *et al.*, 1984; Prideaux *et al.*, 2007;

\*Correspondence to: M. Bunce, as above  
E-mail: michael.bunce@curtin.edu.au

Macken *et al.*, 2012), aDNA has not been integrated into any palaeontological study from this locality because past attempts to extract aDNA from fossil materials were unsuccessful. To address this deficit, we provide the first systematic study to explore the preservation of aDNA in fossil bones from the Naracoorte Caves and its potential as a tool for future palaeontological research. Our study focused on a fossil record spanning the last ca. 18 500 years from the Robertson Cave Entrance Chamber (RCEC), located within the NCWHA. RCEC preserves the youngest dated *in situ* fossil remains for the Naracoorte Caves, which offers the best opportunity to retrieve aDNA and assess its preservation. It contains diverse small mammal fauna (body mass <2.5 kg) of 37 species identified from approximately 2000 specimens (Macken and Reed, 2013); however, large mammal, reptile, frog, bird and plant fossils have also been recovered from the site. Following examples from other vertebrate fossil localities within Australia (Murray *et al.*, 2013; Haouchar *et al.*, 2014; Grealy *et al.*, 2015), we applied a bulk bone metabarcoding (BBM) method to test for the presence aDNA in the vertebrate fossil record from six layers of the RCEC deposit. The BBM method has been demonstrated to be an accurate, rapid, cost-effective way of assessing aDNA preservation and investigating past biodiversity, that does not require the destruction of whole fossil specimens (Murray *et al.*, 2013). This method makes use of non-diagnostic bone fragments that are numerous in fossil deposits such as RCEC. By extracting, amplifying, next-generation sequencing (NGS) and comparing mitochondrial DNA from many morphologically unidentifiable bone fragments in parallel, it is possible to identify the taxa present in the assemblage (Murray *et al.*, 2013). We also aimed to assess the quality (degradation and contamination) and limits of aDNA preservation at RCEC using a ‘shotgun’ sequencing approach, where total genomic DNA from three layers (youngest, middle-aged and oldest) was sequenced and used to model DNA decay (Allentoft *et al.*, 2012). This information allowed us to assess the site’s potential for addressing more sophisticated palaeoecological questions relating to population ecology, genetic diversity, extinction processes and genetic responses to climate change.

## Materials and methods

### Study site and dating

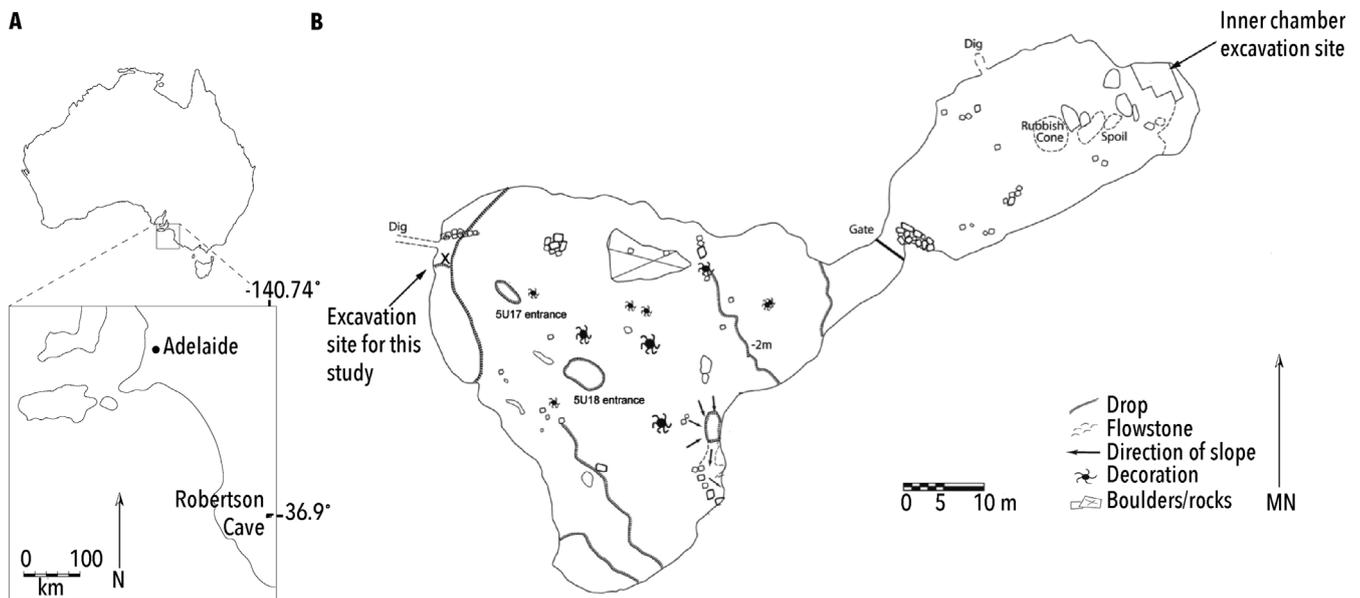
Robertson Cave (37°5.789’S, 140°50.125’E) is located south-east of Naracoorte, in South Australia (refer to Macken and Reed (2013) for details; Fig. 1). The site was excavated to a depth of 83 cm below datum, covering six distinct sedimentary strata (Fig. 2) that were radiocarbon dated to between 906 and 18 629 cal a BP (Fig. 2; Appendix S1; Table S1). Charcoal for accelerator mass spectrometry (AMS) <sup>14</sup>C dating was selected from each of the sedimentary layers, with multiple samples (spits) collected from some layers. AMS <sup>14</sup>C dating was completed by the ANU Radiocarbon Dating Centre (Canberra, Australia) using the techniques described by Fallon *et al.* (2010). Dates were calibrated using the ShCal13 calibration curve to two SD (95.4%; Hogg *et al.* 2013) using OxCal (Bronk Ramsey, 2009).

### Sample collection and bulk bone sampling

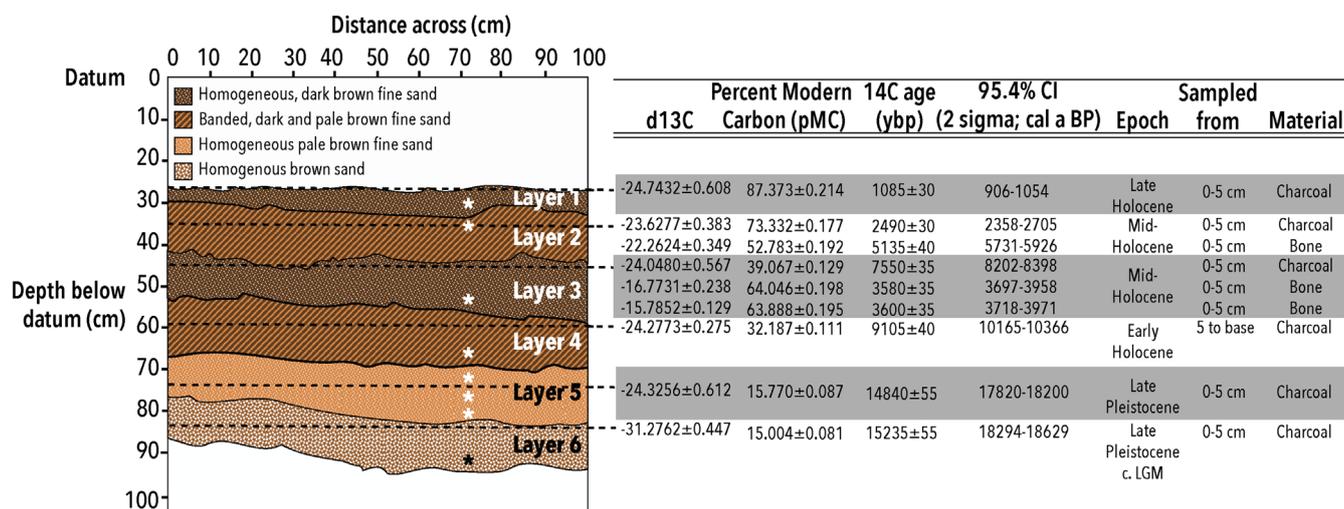
Vertebrate fossil material was sorted from wet-screened samples excavated from RCEC. All fossil material for analysis was collected from excavations conducted in 2011 (as above). The samples were excavated following standard palaeoecological methods with sediments and associated fossil material collected from individual sedimentary layers. Respirators were worn during the excavation and facemasks were worn during sorting and wet sieving. Other DNA contamination reduction methodologies (Allentoft, 2013) were not employed during the excavation of this material. An assortment of 105 bone fragments were randomly collected from screened material from each of Layers 1–6 and placed into labelled vials.

### Sample preparation

In a designated ultra-clean aDNA facility (TRACE) at Curtin University, WA, Australia, bulk bones from each layer were divided into three sub-samples of 35 bones each (average mass 28 mg, range 5–250 mg), following standard aDNA practice (Fig. S1; Appendix S2.1). For each sub-sample,



**Figure 1.** (A) Location of the Naracoorte Caves World Heritage Area in south-east South Australia, indicating the relative location of Robertson Cave (adapted from Macken and Reed, 2013). (B) Map of Robertson Cave Entrance Chamber (5U17), Naracoorte (surveyed by L. Reed and M. Kouklina in 2004, drawn by D. Grindley, adapted from CEGSA Map No. 4014, September 2007).



**Figure 2.** The stratigraphy and dating for Layers 1–6 of the excavation. Dotted lines indicate where samples were taken for dating analysis. Asterisks indicate approximately where bulk bone samples were taken (5-cm spits).

approximately 20 mg of each bone was placed into a clean stainless steel grinding pot and ground into a fine powder using a planetary ball mill (Retsch PM200; 200 r.p.m. for 3 min), which was then transferred to a clean 15-mL tube.

#### DNA extraction and total DNA quantification

DNA from two 100-mg aliquots of bone powder from each sub-sample was extracted (i.e. 36 extracts in total) using the method described by Dabney *et al.* (2013) with minor changes (Appendix S2.2). Two blank (DNA-free) extraction controls were included per 10 extractions. This DNA was then used for both amplicon and shotgun sequencing (Fig. S1). All extractions and downstream quantitative PCRs (qPCRs) were prepared in a physically isolated, pre-PCR ultra-clean environment following standard aDNA practice (Willerslev and Cooper, 2005; Knapp *et al.*, 2012; Shapiro and Hofreiter, 2012). One additional extraction per layer was performed and used for subsequent quantification via spectrophotometry, gel electrophoresis and relative qPCR (Fig. S1). The total DNA concentration was quantified using a Nanodrop spectrophotometer (Thermo Scientific), following the manufacturer's instructions (Appendix S2.3). The, 5 µL of genomic DNA from each layer was run on a 3% agarose gel electrophoresis (Appendix S2.4) to visualize the relative concentrations and total fragment lengths of the genomic DNA, cognizant of the fact that much of the visualized DNA was likely to be of microbial origin. qPCR was performed to determine the efficiency of amplification and assess the relative quantities of template DNA extracted from each layer (Appendix S2.5.1).

#### Amplicon sequencing

Following the suggestions of Murray *et al.* (2013, 2015), the 36 extracts (above) were diluted and screened for inhibition using qPCR (Applied Biosystems) (Appendix S2.5.2). Dilutions exhibiting the least inhibition were then amplified via qPCR (in triplicate) using uniquely indexed Illumina fusion primers specific for a small, diagnostic barcoding region of the mitochondrial 12S rRNA gene in mammals and birds (12SAO, a 150-bp amplicon or 103 bp without primers; Cooper, 1994; Table S2), as well as an additional region specific for marsupials (12SMarsMini, an 85-bp amplicon or 41 bp without primers; D. Haouchar *et al.* (unpublished); Table S2). The purpose of targeting two different sized

amplicons from different barcoding regions is two-fold: it allows us to gauge the fragment length of amplifiable target DNA, as well as to independently verify taxonomic IDs, potentially resolve taxonomic ambiguities or identify more taxa than could be identified by using only one region (Grealy *et al.*, 2015). Extraction controls and three template-free PCR controls were included, and unique indexes were used to eliminate the possibility of sequence contamination arising from previously amplified DNA. Details of the qPCR reaction and thermocycling conditions used are described in Appendix S2.5.2. Amplicons were purified using an Agencourt AMPure XP PCR purification kit with minor changes (Beckman-Coulter; Appendix S2.6), run on a 2% agarose gel electrophoresis (Appendix S2.7), and pooled in approximately equimolar concentrations to create sequencing libraries (Murray *et al.*, 2013). The absolute concentration of the sequencing library was quantified via qPCR (Murray *et al.*, 2012; Appendix S2.8; Table S3) to determine how much to add to the sequencing reaction. Unidirectional sequencing was performed on Illumina's MiSeq platform by following the manufacturer's instructions for the MiSeq 300 V2 Nano kit using a custom sequencing primer (Appendix S2.9).

#### Shotgun sequencing

Shotgun sequencing allows us to more accurately estimate the proportion of the total DNA that is endogenous vs. exogenous (non-target), as well as visualize the fragment length distribution of endogenous aDNA (Allentoft *et al.*, 2012; Heintzman *et al.*, 2014), estimate the DNA decay rate (Allentoft *et al.*, 2012) and examine damage (cytosine deamination) patterns (Briggs *et al.*, 2007). Shotgun libraries were prepared on extracts from Layers 1, 3 and 6, a DNA-free control, and a positive oligonucleotide control by following the protocol outlined by Gansauge and Meyer (2013) with minor modifications (Appendix S2.10; Table S4). After amplifying libraries with uniquely indexed fusion primers (Appendix S2.11), each library was purified using an Agencourt AMPure XP PCR purification kit (Beckman-Coulter) following the manufacturer's instructions with minor modifications (Appendix S2.6). The absolute concentration of each shotgun library was quantified via qPCR using a synthetic standard of known molarity as per Murray *et al.* (2012; Appendix S2.8; Table S3) and the libraries were then pooled in approximately equimolar concentrations to create

the final sequencing library. The absolute concentration of the sequencing library was quantified via qPCR as above to determine how much to add to the sequencing reaction (Appendix S2.8). Unidirectional sequencing was performed on Illumina's MiSeq platform by following the manufacturer's instructions for the MiSeq 150 V3 kit, with minor modifications (Appendix S2.12).

### Bioinformatics and data analysis

After sequencing, raw FastQ files were downloaded and imported into Geneious v. 7.1.4 ([www.geneious.com](http://www.geneious.com); Kearse *et al.*, 2012) for index separation and trimming (Murray *et al.*, 2013). Only reads flanked either side by sequences matching indexes and primers with 100% identity were accepted. Quality control was performed using the web-based workflow Galaxy ([usegalaxy.org](http://usegalaxy.org); Giardine *et al.*, 2005; Blankenberg *et al.*, 2010; Goecks *et al.*, 2010), and chimera filtering and abundance filtering were performed using USEARCH v. 6.1 (Edgar, 2010; Edgar *et al.*, 2011) in QIIME v. 1.8.0 (Caporaso *et al.*, 2010), as per Grealy *et al.* (2015): reads were required to have an average quality score above 25 (i.e.  $\geq Q25$ ), 100% of bases over Q10, 98% of bases over Q15, and 90% of bases over Q20 to be accepted; chimeric reads and low-abundance reads (comprising  $<0.1\%$  of the total number of reads) were discarded. Abundance filtering was not performed on shotgun data, as PCR amplification was stopped during the linear phase, resulting in few clonal copies. For both amplicon and shotgun datasets, sequences present in the extraction or PCR controls (typically microbial or human in origin) were mapped to each dataset and subtracted from them to obtain the final dataset. Taxonomic identification of amplicons was achieved by comparing unique sequences to NCBI's GenBank (Benson *et al.*, 2006) nucleotide reference database via BLASTn (-F No, -e 0.01, -m Pairwise, -G 5, -E 2, -v 20, -b 20, -W 7, -reward 1) (Altschul *et al.*, 1990), with searches executed in YABI ([cgc.murdoch.edu.au/yabi](http://cgc.murdoch.edu.au/yabi); Hunter *et al.*, 2012). BLAST results were imported into MEGAN v. 4.70.4 ([ab.inf.uni-tuebingen.de/data/software/megan4](http://ab.inf.uni-tuebingen.de/data/software/megan4); Huson *et al.*, 2007) to evaluate taxonomic assignments. The LCA parameters used in MEGAN were: min support of 1, min score of 35, top per cent of 10, min complexity 0.44, win score 0. Identifications were based on the percentage sequence similarity of the query to the reference across 100% of the query, in the BLAST hits of the terminal nodes of the MEGAN tree. Current and historical taxon distributions described refer to records housed in the online database Atlas of Living Australia ([www.ala.org.au](http://www.ala.org.au), accessed 30 October 2015).

### Modelling of aDNA preservation and decay over time

Each shotgun dataset (not dereplicated) was rarefied in MEGAN v. 4.70.4 (Huson *et al.*, 2007) to account for differences in sequencing depth between the different libraries: 60 000 sequences (approximately half the smallest dataset) were sufficient to capture almost all the diversity, and as such, were randomly sub-sampled from each layer. Fragment-length distributions of endogenous aDNA from the rarefied dataset (defined as the unique marsupial shotgun sequences obtained from mapping sequences against GenBank's non-redundant nucleotide database using BLASTn as above) were constructed for Layers 1, 3 and 6. The summary statistics for each distribution (number of sequences, mean fragment length, standard deviation, mode fragment length and maximum length retrieved) were calculated in Geneious v. 7.1.4 (Kearse *et al.*, 2012). Under

an exponential decay model, fragment length should be inversely proportional to the log of the copy number, where the probability of a bond in the DNA backbone being broken,  $\lambda$ , is equal to the exponential coefficient (Allentoft *et al.*, 2012). An exponential relationship was modelled using the declining part of the distribution, excluding biases in the distribution tails (Allentoft *et al.*, 2012), and the exponential coefficient ( $\lambda$ ) and fit ( $R^2$ ) of the relationship was calculated in Microsoft Excel. The per nucleotide fragmentation rate per year ( $k$ , per site per year) was calculated for the average, minimum and maximum possible calibrated ages of layers with  $>500$  marsupial sequences (i.e. Layers 1 and 3), as  $\lambda$  divided by age (cal a BP) (Allentoft *et al.*, 2012).

The estimations of  $k$  for Layers 1 and three were used to predict a number of relevant measures (Allentoft *et al.*, 2012): the average fragment length of the DNA in the extract (bp) ( $1/\lambda$ ); the number of years until the DNA is completely degraded (i.e. average fragment length = 1 bp) ( $1/k$ ); the decay constant ( $k_{30} = 1 - e^{-k \cdot 30}$ ) and the molecular half-life (years) of the smallest informative fragment size of 30 bp [ $(\ln(2)/k_{30})$ , corresponding to the number of years it would take 50% of 30-bp fragments to be gone, or, an average of 0.5 strand breaks per 30 bp]; and the expected proportion of broken bonds after 50 000 years ( $\lambda_{50,000} = k \cdot 50,000$ ). Next, the expected proportion of 30-bp fragments left after time  $t$  was modelled by plotting  $e^{-k_{30} \cdot t}$ , and the proportion of 30-bp fragments left after 50 000 years was calculated using  $e^{-k_{30} \cdot 50,000}$ . Finally, the probability of a fragment of size  $L$  bp surviving after 50 000 years was modelled by plotting  $e^{-\lambda_{50,000} \cdot L}$ , and the probability of a 30-bp fragment surviving 50 000 years was calculated using  $e^{-\lambda_{50,000} \cdot 30}$ .

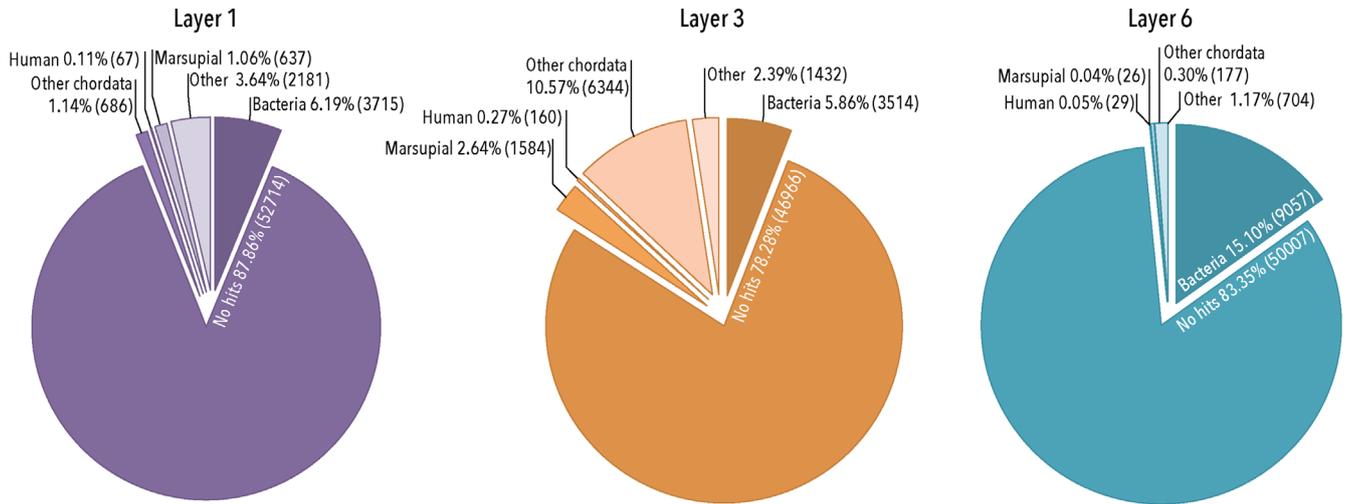
### Base composition analysis

The GC content of the unique marsupial shotgun sequences for Layers 1, 3 and six was calculated in Geneious v. 7.1.4 (Kearse *et al.*, 2012) as the proportion of all the bases in the dataset that were either a G or a C nucleotide. The T-content of the first and last base of these same sequences was also calculated as the proportion of first-base nucleotides in the dataset that were a T, or the last-base nucleotides in the dataset that were a T.

## Results and discussion

### DNA preservation at RCEC

Fossils preserved in caves typically exhibit less degraded aDNA in comparison with fossils from other non-frozen environments, as caves usually offer a cool, dry environment and tend to buffer temperature (Bollongino *et al.*, 2008; Elsner *et al.*, 2015; Gutiérrez-García *et al.*, 2014; Olalde *et al.*, 2014). However, aDNA retrieval from caves within the NCWHA has never been published, despite the fact that they house a rich palaeontological record, displaying exceptional preservation of fossils dating as far back as 500 000 years. Here, NGS metabarcoding of two short mitochondrial gene regions amplified from DNA extracted from bulk bones present across six stratigraphic layers spanning approximately 18 000 cal a BP yielded a total of 411 717 sequences (Table S5). NGS of shotgun libraries from Layers 1, 3 and six yielded 648 739, 117 428 and 240 293 sequences, respectively; sub-sampling 60 000 sequences from each layer yielded between 16 and 746 unique marsupial sequences per layer. Examination of the base composition of these reads (Appendix S3.2; Table S6; Fig. S2) shows that there is a bias towards thymine bases at both the 5' and the 3' ends



**Figure 3.** Pie charts for Layers 1 (purple), 3 (orange) and 6 (blue) showing the percentage of shotgun sequence reads out of 60 000 sub-sampled sequences that map to bacterial, human, marsupial, other chordate and other organisms, and nothing by BLASTn to GenBank's non-redundant nucleotide database.

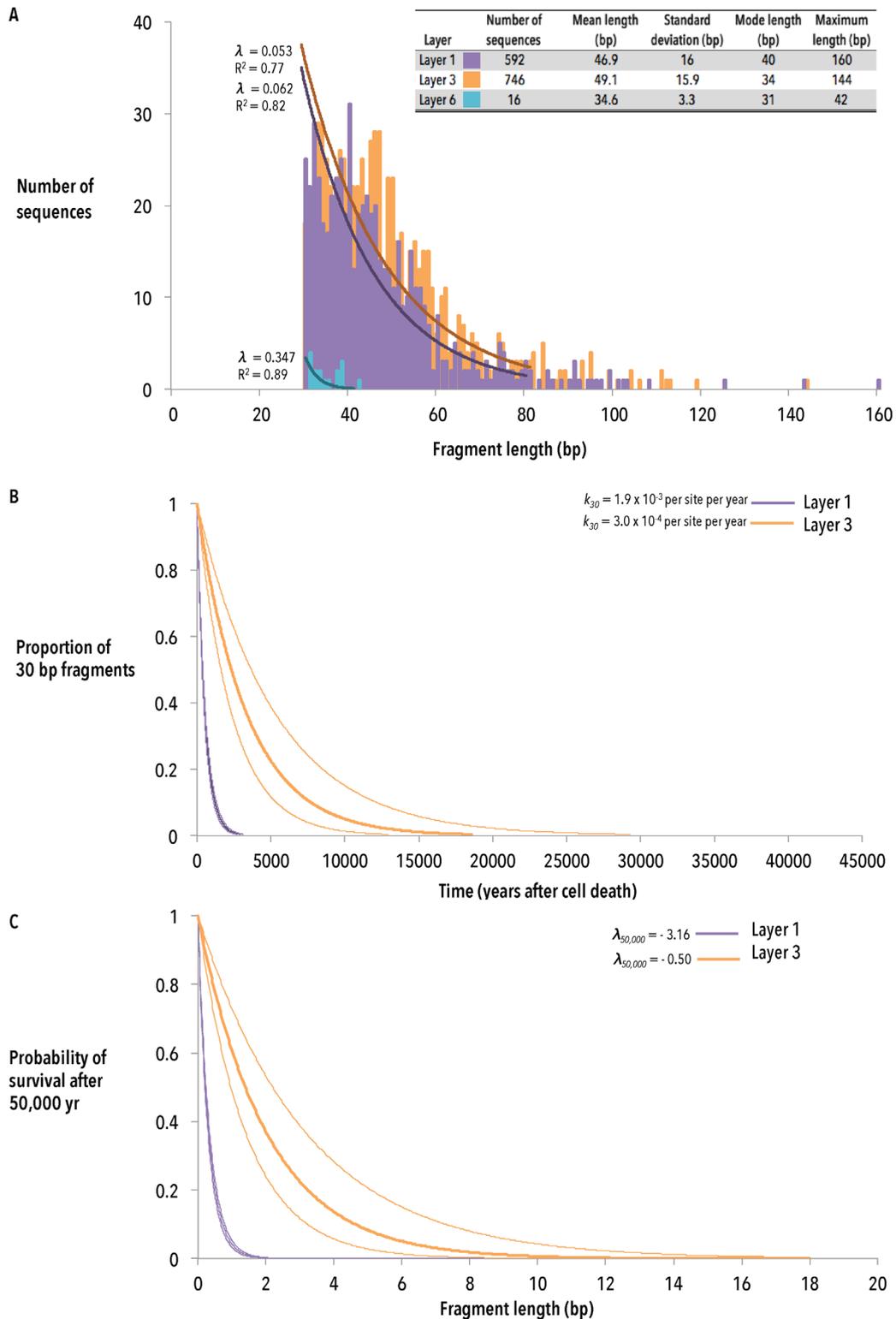
(>35–75%), which is to be expected in aDNA (Gansauge and Meyer, 2013) as DNA hydrolysis and oxidation can result in the deamination of cytosine to uracil (sequenced as a T) (Hoss *et al.*, 1996; Hofreiter *et al.*, 2001; Gilbert *et al.*, 2003). Although we cannot be certain that this bias is caused by cytosine deamination, the composition of thymine in these positions is higher than would be expected from undamaged DNA (roughly 25%). The bias towards thymine bases at the ends of fragments is also higher in Layer 6 than in Layers 1 and 3 (Table 2). This type of damage pattern is highly suggestive that the DNA obtained is genuinely ancient in nature.

While we were able to recover aDNA from all six layers, the aDNA retrieved from layers older than ca. 8000 years (early Holocene) was considerably scarcer, more degraded and more contaminated than that of the younger layers (note that the overlapping of ages for Layers 2 and three suggests that post-depositional disturbance may have occurred, disrupting the temporal integrity of these layers; however, for the purposes of assessing aDNA preservation, and in the context of the whole site, Layers 2 and three provide a mid-Holocene sample for comparison against the late Holocene record of Layer 1, the early Holocene record of Layer 4, and the late Pleistocene record of Layers 5 and 6). Firstly, although much of the genomic DNA is likely to be of bacterial origin, the total amount of genomic DNA recovered from Layers 1–3 is similarly higher (on average  $30.75 \text{ ng } \mu\text{L}^{-1}$ ) and more pure (average  $260/280 = 1.66$ , where the  $260/280$  ratio of pure DNA = 1.8) than that of Layers 4–6 ( $25.57 \text{ ng } \mu\text{L}^{-1}$ ,  $260/280 = 1.56$ ) as measured by spectrophotometry (Table S7). Visualization of genomic DNA by agarose gel electrophoresis (Table S7) shows brighter smears of similar intensity for Layers 1–3 compared with Layers 4–6, also indicating that while the DNA from all layers is highly degraded, DNA from Layers 1–3 is more concentrated than Layers 4–6. These results are also supported by the finding that the relative amount of amplifiable DNA (expressed by the 'fold difference') for both primer sets is higher in extracts from Layers 1–3 compared with 4–6 as determined by qPCR (Table 1; Table S8); for 12SAO, Layers 1–3 yielded similar (the same order-of magnitude) amounts of template DNA, about 24–60 times as much as Layers 4–6, and similarly, Layers 1–3 contain over 100-fold more 12SMarsMini template DNA than Layers 4–6, which themselves contain very low

quantities (i.e. higher  $C_T$  values) of template DNA (Table 1). Moreover, the poor efficiency (<95%; Table S9; Figs S3 and S4) and the stochasticity (i.e. the lack of consistency between the  $C_T$  values of duplicate PCRs – Table S8 – and sporadic amplification with both primer sets in extraction replicates – Table S10) of successful target DNA amplification in Layers 4–6 suggests that they are more inhibited and contain considerably fewer template DNA molecules than Layers 1–3 (Table 1). Finally, Layers 1 and three contain a lower proportion of bacterial DNA (5.86 and 6.19%, respectively) and a higher proportion of endogenous (marsupial) DNA (1.06 and 12.64%, respectively) compared with Layer 6 (15.10% bacteria, 0.04% marsupial) as determined by shotgun sequencing (Fig. 3). That is, Layer 6 contains about 27 times less endogenous DNA than Layer 1 and about 316 times less than Layer 3. Thus, these findings suggest that the quantity of endogenous aDNA is greater in Layers 1–3 compared with Layers 4–6.

In addition, the shorter mean and maximum fragment length of aDNA in Layers 4–6 compared with Layers 1–3, gauged by both amplicon and shotgun sequencing, indicate that endogenous DNA from Layers 4–6 is more degraded than Layers 1–3. Although some endogenous mtDNA was amplified with both primer sets in extracts from Layers 4–6, most taxa were only amplified with the short 12SMarsMini primer set (Table 3). While this suggests that there is some endogenous mtDNA that is at least 150 bp in length present in these layers, most of it is expected to be much shorter than this and would be in trace amounts that approach the limits of detection. In contrast, many of the same taxa were detected in Layers 1–3 with both primer sets, and therefore these extracts have a greater proportion of mtDNA that is at least 150 bp in length compared with Layers 4–6. These results are consistent with those of the shotgun sequencing of DNA from Layers 1, 3 and 6 (Fig. 4A), which was used to determine the distribution of fragment lengths of informative (>30 bp) endogenous (marsupial) DNA: the mean fragment length of sequences from Layers 1 and three was between 45 and 50 bp, with the maximum fragment retrievable being above 140 bp in length, whereas for Layer 6, the mean fragment length was around 35 bp, with the maximum fragment retrievable being 42 bp (Fig. 4A).

Although these results show that DNA preservation appears largely to be decreasing through time (as would be expected



**Figure 4.** (A) Fragment length distributions of unique marsupial shotgun reads for Layers 1 (purple), 3 (orange) and 6 (blue) showing number of sequences (out of 60 000 sub-sampled sequences) for each read length (bp) that mapped to marsupial sequences by BLASTn to GenBank’s non-redundant nucleotide database. The table embedded within the histogram shows summary statistics for the distribution. The slope ( $\lambda$ ) and fit ( $R^2$ ) of the exponential relationship modelled is also shown. (B) The proportion of unique marsupial 30-bp sequences expected to be remaining given a certain time (years after cell death), as predicted by the average (thick line), minimum and maximum (thin lines) decay constants calculated from both Layer 1 (purple) and Layer 3 (orange). (C) The probability of survival of 50 000 years after cell death for various fragment lengths (bp), as predicted by the average (thick line), minimum and maximum (thin lines) decay constants calculated from both Layer 1 (purple) and Layer 3 (orange).

by a rate model of DNA fragmentation), we find that the molecules recovered from Layer 3 appear to be less degraded than Layer 1, despite being approximately five times older. Using the observed fragment length distributions of marsupial shotgun sequences from Layers 1 and 3 (Fig. 4A) to model the

rate of DNA degradation by depurination (the generation of abasic sites, resulting in strand breaks) over time (Allentoft *et al.*, 2012), the per nucleotide fragmentation rate,  $k$ , was found to be approximately six times faster for Layer 1 than Layer 3:  $k_{av}$  for Layer 1 was estimated to be  $6.33 \times 10^{-5}$  per

**Table 1.** Replicate and average qPCR cycle-threshold ( $C_T$ ) for both 12SAO and 12SMarsMini primer sets for Layers 1–6; the fold difference in the quantity of amplifiable DNA is calculated relative to Layer 6.

	12SAO						12SMarsMini					
	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
$C_T$ value replicate 1	27.48	26.55	26.77	32.73	29.86	30.87	38.21	29.69	30.15	45.35	n/a	n/a
$C_T$ value replicate 2	27.56	26.01	27.58	30.33	30.83	33.44	33.72	31.56	29.77	n/a	n/a	44.11
Average $C_T$ value	27.52	26.28	27.18	31.53	30.35	32.16	35.97	30.63	29.96	45.35	n/a	44.11
Fold difference in the quantity of amplifiable DNA (relative to Layer 6)	24.85	58.69	31.56	1.54	3.51	1.00	283.07	11 465.41	18 179.19	0.42	n/a	1.00

site per year, while for Layer 3,  $k_{av}$  was estimated to be  $9.95 \times 10^{-6}$  per site per year (Table 2). This translates to an estimated true average fragment length in the DNA extract of 16 bp for Layer 1, and 19 bp for Layer 3. Layer 3 also contained more unique marsupial DNA than Layer 1 (Fig. 3). This supports the finding that Layer 3 also had lower average  $C_T$  values in the qPCR assays than Layer 1 for both primer sets (Table 1), indicating the presence of more template copies of DNA, despite amplification being less efficient in Layer 3 (Table S9). Note that although the decay rate of mtDNA is slower than that of nuDNA (Allentoft *et al.*, 2012), the proportion of marsupial DNA that mapped to annotated mitochondrial reference genomes was only between one and 2% – too little to quantify the mtDNA decay rate. However, this may be possible to determine with increased sequencing depth in the future.

These results suggest that DNA preservation at RCEC is influenced by more than age alone (Haynes *et al.*, 2002; Hansen *et al.*, 2006). The discrepancy between the age and the observed DNA preservation ‘likely derives from differences in taphonomy or bone diagenesis’, including microbial digestion, bone thickness (Allentoft *et al.*, 2012), and other conditions within the cave, such as oxygenation, moisture, radiation (Campos *et al.*, 2012; Sawyer *et al.*, 2012), pH, salinity, redox potential and temperature (Smith *et al.*, 2003; Allentoft *et al.*, 2012). Sediment type (Rawlence *et al.*, 2014) and character, such as ‘grain size, mineralogical composition, organic matter load’ (Corinaldesi *et al.*, 2008), the presence of humics and humates (Allentoft *et al.*, 2012), movement, compaction and lithification (Eglinton and Logan, 1991) can all further influence aDNA preservation. In addition, the physical aspects of caves such as entrance type and size, depth in limestone, surface topography and relative depth of passage all influence how materials are accumulated, as well as the frequency, volume and direction of water flows into caves, and the stability or otherwise of within-cave environmental conditions (Reed, 2003, 2009). Depurination may also not be the only mechanism of DNA damage in caves within the NCWHA: DNA can be damaged by other mechanisms, such as crosslinking, and blocking and miscoding lesions (Hoss *et al.*, 1996; Hofreiter *et al.*, 2001; Gilbert *et al.*, 2003). Although these mechanisms can play a greater role in DNA damage than depurination (Hansen *et al.*, 2006), they have not been taken into account in this model of the DNA decay. Together, these factors are expected to play a large role in the likelihood of aDNA preservation, which may explain why aDNA has not been recovered from any other cave in the NCWHA before now. This lack of sole dependence of DNA degradation on age could explain why mtDNA preservation appeared to be largely consistent up to 8000 cal a BP after which it ‘drops off’, but remains consistently low

for the next 10 000 years, or why DNA decay in Layer 3 has proceeded more slowly than in Layer 1. It is possible that something about the environment ca. 8000 years ago may have been more favourable for DNA survival. For example, specimens from Layer 3 may have been more rapidly buried than Layer 1. Another possibility is that Layer 3 has been disturbed less than Layer 1, which may have been compacted and contaminated by recent human and animal activity in the cave.

The decay rates for Layers 1 and 3 were used to estimate the molecular half-life of a 30-bp DNA fragment (Table 2) and the likelihood of such a minimally informative DNA sequence surviving through time. Such models can be useful to predict the probability of recovering informative DNA from megafaunal species (e.g. sthenurine kangaroos) in the 40 000–50 000 cal a BP period over which they became extinct (Llamas *et al.*, 2014). Assuming that the decay rate of DNA in older layers is comparable to those in Layers 1 or 3, modelling (Fig. 4B,C) suggests that after 50 000 years, the proportion of unique marsupial 30-bp sequences in an extract, as well as the probability of a 30-bp fragment surviving 50 000 years, is expected to be between  $10^{-42}$  and  $10^{-7}$  (Table 2). A ‘best case scenario’ would predict that 1 in 10 million 30-bp fragments would still be available today, which would be equivalent to about 10 random 30-bp DNA fragments surviving per cell, assuming a genome size of about 2.9 Gbp (the size of the Tammar wallaby *Macropus eugenii* genome; Renfree *et al.*, 2011). Even if only a few hundred cells survive in a bone sample, it cannot be excluded that a small amount of authentic DNA could be obtained from much older fossils. We have also shown that the DNA fragmentation at RCEC is not entirely dependent on age, meaning that the DNA loss in megafaunal bones from the older layers may have been slower than the younger layers. Improvements in DNA technology such as target enrichment through hybridization capture and higher sequencing coverage may improve the chances of retrieving low-copy number DNA (Llamas *et al.*, 2014), particularly mtDNA. Post-excavation conditions will also influence the retrieval of aDNA. For instance, it has been observed that recently excavated, dry-sieved and untouched fossil bones contain more endogenous DNA than bones that have been stored and excavated under standard procedures (Pruvost *et al.*, 2007). Other factors including exposure to UV radiation, thermal stress, human contact and wet sieving all reduce the amount of DNA that can be obtained (Bollongino *et al.*, 2008). Because the bones sampled here were handled and washed (albeit out of necessity), it is highly likely that DNA retrieval was adversely affected, which could explain the stochasticity of amplification (and contamination) observed in the older layers. This may also have influenced estimates of the decay rate. As

**Table 2.** Empirical decay constant ( $k_{av}$ , per site per year) for Layers 1, 3 and 6, modelled using  $\lambda$  as determined from the exponential part of the fragment length distributions (Fig. 4) and the average, maximum and minimum estimates of bone age (Fig. 2) (where  $k = \lambda/\text{age}$ ). The decay constant was not calculated for Layer 6 because the sample size (i.e. number of fragments in the distribution) was too small to accurately estimate  $k$ . The G+C content (% of bases either G or C) and T content bias (% of bases T) at the 5' and 3' ends of the unique marsupial sequences were also calculated for Layers 1, 3 and 6 based on the shotgun data.

Parameter	Layer 1	Layer 3	Layer 6
Av. empirical decay rate, $k_{av}$ (per site per year) (min-max)	$6.33 \times 10^{-5}$ (5.88 $\times 10^{-5}$ –6.84 $\times 10^{-5}$ )	$9.95 \times 10^{-6}$ (6.31 $\times 10^{-6}$ –1.43 $\times 10^{-5}$ )	n/a
Av. fragment length of extract (bp)	16	19	n/a
Av. fragment length of extract after 50 000 years (bp) (min-max)	0.31 (0.29–0.34)	2 (1.40–3.17)	n/a
Av. number of years until av. fragment length is 1 bp (min-max)	15 806 (14 619–17 007)	100 452 (69 930–158 479)	n/a
Av. $k_{50}$ (per site per year), 30 bp (min-max)	$1.9 \times 10^{-3}$ (1.80 $\times 10^{-3}$ –2.0 $\times 10^{-3}$ )	$3.0 \times 10^{-4}$ (1.9 $\times 10^{-4}$ –4.3 $\times 10^{-4}$ )	n/a
Av. half-life (years), 30 bp (min-max)	366 (393–338)	2321 (3622–1616)	n/a
Av. proportion of 30-bp fragments left after 50 000 years (min-max)	$6.7 \times 10^{-42}$ (3.1 $\times 10^{-45}$ –5.4 $\times 10^{-39}$ )	$3.3 \times 10^{-7}$ (4.98 $\times 10^{-10}$ –7.8 $\times 10^{-5}$ )	n/a
Av. probability of a 30-bp fragment surviving 50 000 years (min-max)	$6.1 \times 10^{-42}$ (2.8 $\times 10^{-45}$ –5.0 $\times 10^{-39}$ )	$3.3 \times 10^{-7}$ (4.8 $\times 10^{-10}$ –7.6 $\times 10^{-5}$ )	n/a
G+C content (%)	42.5	43.1	43.8
T content first base (%)	40.2	43.6	37.5
T content last base (%)	63.0	63.0	75.0

such, we would expect DNA retrieval to be even better from fossils that have been freshly excavated under sterile conditions. While there remains the potential for aDNA to be retrieved from even older layers within RCEC, it will always be difficult to predict the cumulative effects of the many factors that could influence DNA preservation and the probability of capturing megafaunal DNA from Naracoorte; ultimately, it needs to be empirically tested.

Nevertheless, the aDNA preservation in bone at RCEC is largely comparable to that of other cave systems around Australia that have yielded aDNA. For instance, the sediments found in the Kelly Hill Caves complex on Kangaroo Island, South Australia, are made up of primarily silty sand and are largely similar to the sediments of the NCWHA caves, including RCEC (McDowell *et al.*, 2013). Mammal aDNA between 115 and 175 bp was amplified from bones preserved at this site that were ranged from ca. 6800 to over 20 000 cal a BP (Haouchar *et al.*, 2014). However, several layers between 6800 and 10 000 cal a BP, and over 20 000 cal a BP, did not yield mammal DNA (Haouchar *et al.*, 2014) and the mammalian diversity identified through aDNA was much lower than this study (see below) – about four families across 15 layers. This may indicate that fragment length sizes were smaller than could be amplified with the primer sets used, i.e., shorter than 100 bp – similar to fragment lengths observed in this study. In another study, aDNA was isolated from archaeological bulk bone samples up to 46 000 years old from south-west Australia (Devil's Lair and Tunnel Cave) (Murray *et al.*, 2013). Using two mammalian-specific primer sets to amplify 150 bp, eight mammalian families were detected spanning ca. 4000–22 000 and 10 000–46 000 cal a BP at Tunnel Cave and Devil's lair, respectively (Murray *et al.*, 2013). At these sites, aDNA preservation is clearly exceptional; in comparison, much shorter fragments were retrieved from much younger samples at RCEC. More recently, 37–121 unique mtDNA sequences averaging 37–40 bp were recovered from 40 000–50 000 cal a BP megafaunal bone from high altitude caves in Mt Cripps, Tasmania (Llamas *et al.*, 2014); however, no DNA was able to be amplified through PCR, and libraries were enriched for endogenous DNA. While there were too few endogenous sequences to estimate  $k$  for Layer 6, we still managed to retrieve mtDNA and nuDNA from bones over 18 000 years old without enrichment. This further supports the notion that enrichment may be necessary to capture megafaunal DNA from Naracoorte, but also that the likelihood of success could be even smaller given the warmer climate of Naracoorte compared with Mt Cripps. To date, no study has conducted a thorough investigation into the factors that have the greatest impact on aDNA preservation in Australian cave systems. Thus, it would be highly informative to compare aDNA preservation with a variety of cave conditions to identify the factors that are most conducive to aDNA preservation; this will assist in predicting those sites with the likelihood of best aDNA preservation.

### Biodiversity

DNA attributable to a variety of taxa was amplified and sequenced with high coverage from all six layers of RCEC using two primer sets covering mitochondrial loci of differing lengths (Table 3). Approximately 300 000 and 100 000 sequence reads were generated in total for the 12SAO and 12SMarsMini primer sets (Appendix S3.5; Table S10), yielding 103 and 41 bp of informative sequence, respectively. We were able to identify a range of marsupial and placental mammal, bird and amphibian families from each layer

**Table 3.** Summary of the taxa identified from RCEC for Layers 1–6 by comparison of sequence information for the metabarcoding primer sets 12SAO (regular typeface) and 12SMarsMini (bold typeface) to GenBank’s non-redundant nucleotide database using BLASTn.

	Family 90–95% similarity	Genus 95–98% similarity	Species 98–100% similarity	Common name	Layer						
					1	2	3	4	5	6	
Marsupial#^	Acrobatidae#	<i>Acrobates#</i>	<i>pygmaeus#</i>				‡ ‡ ‡		‡ ‡ ‡	‡ ‡ ‡	
	Vombatidae^	<i>Vombatus^</i>	<i>ursinus^</i>	Common wombat	‡/‡ ‡/‡ ‡/‡			‡		‡	
	Petauridae#	<i>Petaurus#</i>	<i>breviceps#</i>	Sugar glider			‡ ‡ †				
	Dasyuridae#	<i>Dasyurus#</i>		Quoll	‡	‡	‡	‡/‡	‡		
	Burramyidae#^	<i>Cercartetus#</i>	<i>nanus#</i>	Eastern pygmy possum			‡ ‡ ‡			‡ ‡ ‡	
	Potoroidea#	<i>Bettongia#</i>		Bettong	‡	‡	‡/‡	‡/‡			
	Peramelidae#	<i>Isoodon#</i>		Short nosed bandicoots	‡/‡ †	‡	‡/‡	‡/‡	‡	‡/‡	
	Macropodidae^	<i>Lagorchestes*</i> <i>Onychogalea*</i>	<i>unguifera*</i>	Northern nail tail wallaby	‡/‡	‡/‡	‡/‡	‡/‡	‡/‡	‡	
		<i>Macropus</i>	<i>agilis</i>	Agile wallaby	‡	‡/‡		‡	‡	‡	
			<i>fuliginosus</i>	Western grey kangaroo	†				†		
			<i>rufogriseus</i>	Red-necked wallaby		†					
	Phalangeridae#^	<i>Trichosurus#</i>	<i>vulpecula#</i>	Common brushtail possum	‡/‡ ‡/‡ †/†	‡/‡ ‡/‡ †/†	‡/‡ ‡/‡ †/†	‡/‡ ‡/‡ †/†	‡/‡ ‡/‡ †/†	‡/‡ ‡/‡ †/†	
	Pseudocheiridae#^	<i>Pseudocheirus#</i>	<i>peregrinus#</i>	Common ringtail possum	‡	‡	‡/‡	‡	‡	‡	
Placental	Miniopteridae^	<i>Miniopterus</i>	<i>fuliginosus</i>	Eastern bent-winged bat	‡ ‡ Δ	‡ ‡ Δ	‡ ‡ Δ		‡ ‡ Δ		
	Muridae^	<i>Rattus</i>			‡	‡	‡	‡	‡	‡	
Avian^	Rhipiduridae	<i>Rhipidura</i>	<i>fuliginosa</i>	New Zealand fantail			‡ † Δ				

continued

Table 3. (Continued)

Family 90–95% similarity	Genus 95–98% similarity	Species 98–100% similarity	Common name	Layer						
				1	2	3	4	5	6	
Columbidae	<i>Phaps</i>	<i>chalcoptera</i>	Common bronzewing	‡ ‡ †	‡ ‡ †		‡ ‡ †			
Cracticidae/Artamidae	<i>Cracticus</i>	<i>tibicen</i>	Australian magpie			‡ ‡ †				
Amphibian	Hylidae	<i>Litoria</i>	Australasian tree frogs	‡ ‡	‡ ‡	‡ ‡		‡ ‡		
Myobatrachidae	<i>Limnodynastes</i>		Australian swamp frogs		‡ ‡	‡ ‡	‡ ‡	‡ ‡		

#The taxon was identified morphologically from RCEC in Macken and Reed (2013).

^The taxon was identified morphologically in single-source bones from the same accumulation (unpublished). Shaded squares indicate that the taxon was detected in at least two or more replicates (either sub-sample replicates or extraction replicates).

Sequence similarity between query and reference cut-offs (with 100% query coverage) for each taxonomic level include: 90–95% for family, 95–98% for genus and 98–100% for species (note that for the 12SMarsMini primer set, species are only considered when the similarity to the reference is 100%).

Taxonomic assignments are classified as: ‡highly credible, being within the cut-off percentage, all other members present in GenBank, and is known to region.

†credible, being within the cut-off percentage, but where other members of the taxon are equally possible (equal similarity to multiple members or other members not present in GenBank); Δuncertain, being within the cut-off percentage, but not all genus members present in GenBank and is not known from region; and §inferred, not within the cut-off but no other genus members described, and is known from the region.

\*Possible presence of an extinct genus known from the region, but not represented in GenBank.

spanning the ages of ca. 950–18 500 cal a BP, including all the small mammal families that were identified morphologically by Macken and Reed (2013). In addition, we identified several large-bodied marsupial families that have not been previously described from the RCEC fossil record (Vombatidae, Macropodidae), as well as families that can be difficult to distinguish morphologically (e.g. amphibian taxa, such as *Limnodynastes*, *Litoria*). This adds weight to the validity of the both the morphological and the molecular taxonomic identifications, and is yet another example of how aDNA analysis of bulk bone is sensitive enough to rapidly capture a large portion of the biodiversity (Murray *et al.*, 2013; Haouchar *et al.*, 2014), including some hitherto undescribed biodiversity. Of the taxa detected, highly credible native animal families (90–95% sequence similarity between query and reference, a representative of all other members present in GenBank, and is known in the region) include Vombatidae (wombats), Petauridae (striped and gliding possums), Dasyuridae (carnivorous/insectivorous marsupials), Burramyidae (pygmy possums), Potoroidae (potoroos), Peramelidae (bandicoots), Macropodidae (kangaroos and wallabies), Phalangeridae (possums), Pseudocheiridae (ring-tailed possums), Minopteridae (long-winged bats), Muridae (rodents), Rhipiduridae (fantails), Columbidae (pigeons), Cracticidae (magpies), Hylidae (Australasian tree frogs) and Myobatrachidae (Australian ground frogs) (Table 3). Highly credible species assignments could not be made for most of the native taxa due to the incompleteness of reference databases, the possibility that sequences may derive from extinct (or extirpated) species, and that DNA damage, amplification error and sequencing error can confound evolutionary change leading to false identifications (Grealy *et al.*, 2015). This is particularly true for the very small 12SMarsMini amplicons

because, depending on the locus, a few base changes could result in the misidentification of species; however, we estimate that a mutation rate of 3% (well above the error rate of Illumina sequencing platforms) would result in a genus misidentification only 0.1% of the time, with family misidentifications being even rarer (Appendix S3.6; Tables S11 and S12). Misidentifications at the genus and family levels are further reduced by stringent quality and abundance filtering, high coverage, independent confirmation with another primer set, and replication with multiple extractions. Where species identifications were able to be confidently made, the taxa either were not detected using multiple primer sets (e.g. *Cercartetus nanus*; Table 3) or were not detected in more than one extract within a layer (e.g. *Pseudocheirus peregrinus*; Table 3), which casts some doubt on their authenticity due to lack of repeatability. Nevertheless, the presence of regionally extinct genera such as *Lagorchestes* supports the authenticity of the aDNA obtained, as it is extremely unlikely that these sequences could have originated from modern contamination. Non-native taxa are likely to be contaminants (e.g. *Homo sapiens*, *Ovis*; Appendix S3.5; Table S10) as they are very well described in GenBank and there is no evidence of their presence at Naracoorte between 950 and 18 500 cal a BP. Improvements in bulk-bone approaches, such as enrichment through the use of phylogenetically broad baits in hybridization capture, may potentially improve the resolution of the molecular taxonomic identifications, remove contaminating sequences and recover a larger proportion of the biodiversity, allowing us to better characterize biodiversity change through time.

The results presented here suggest that aDNA analysis of fossils at Naracoorte can be applied to investigate a range of palaeontological questions regarding adaptation of animal

taxa to environmental change, as well as population shifts in space and time. For example, it is hypothesized from fossil evidence that *Petaurus norfolcensis* underwent a range expansion into the Naracoorte region during the late Pleistocene; however, modern genetic analysis of South Australia populations remains inconclusive (Macken and Reed, 2013). Targeted aDNA approaches (for instance, use of species-specific primers, whole genome sequencing or hybridization capture) using single source bone from fossil specimens of *P. norfolcensis* have the potential to clarify this. In addition, the identification of extinct taxa can help investigate extinction processes: the genus *Lagorchestes* was identified with high confidence, but the only member of the genus known from South Australia is the extinct species *Lagorchestes leporides*, which does not have any representative sequences in GenBank. Similarly, the genus *Onychogalaea* identified may represent the extinct species *O. lunata* or the locally extirpated species *O. fraenata*. Information about such species can be used to investigate the timing, mechanisms and causes of past biodiversity change, which is an important step towards better conservation and management of endangered extant relatives (Bohmann *et al.*, 2014; Pansu *et al.*, 2015). Thus, when integrated with palaeontological and geological data, aDNA from Naracoorte has the potential to offer new insights into historical faunal turnover in South Australia. Such approaches may also be able to resolve ambiguous species IDs, such as *Pseudomys novaehollandiae*, which is difficult to distinguish morphologically from *P. apodemoides* when dentaries are absent (Macken and Reed, 2013).

## Conclusion

The World Heritage listed Naracoorte Caves, South Australia, are of significant educational and environmental importance due to their extensive palaeontological fossil deposits. These deposits play an integral role in generating an appreciation for scientific research and cultivating awareness of environmental change and human impact in Australia. While many in-depth morphological studies of these fossils have been carried out, aDNA research at these caves has been limited despite its potential to help answer a variety of palaeoecological questions. Here, the successful purification of aDNA from Holocene and late Pleistocene bulk bone samples from RCEC within the NCWHA, coupled with NGS technology, enabled an assessment of aDNA preservation at this site that was found to be comparable to other Australian cave sites. For the first time, it was possible to identify a range of marsupial and placental mammal, bird and amphibian families and genera with high confidence using aDNA as old as ca. 18 500 cal a BP, some of which are extinct, or may be difficult to distinguish morphologically. While we were able to recover aDNA from layers dating back as far as the Last Glacial Maximum, there remains a small possibility that aDNA may be retrieved from even older specimens within RCEC; however, future improvements in technology may improve this likelihood. Thus, our results suggest that further aDNA research at Naracoorte is warranted, especially as aDNA preservation in other organic materials from the caves, such as plants (including leaves, nuts, seeds and cones), pollen and sediments has yet to be assessed. This study shows that aDNA can be a valuable and complementary tool for understanding the past, and will add to the continuing impact of the NCWHA to palaeontological research in Australia.

**Acknowledgments.** All necessary permits were obtained for the described study, which complied with all relevant regulations. We acknowledge funding for this project via Friends of Naracoorte Caves

(E.R.), Sir Mark Mitchell Research Foundation (A.M.), and Caring for our Country grant 0C11-00487. M.B. was supported in this research by an Australian Research Council (ARC) future fellowships FT0991741 and DP120104435. Computing resources at the University of Copenhagen as well as the Pawsey Centre, WA, were used to perform BLAST searches. We thank S. Bourne for assistance to E.R. with the 2011 excavations of RCEC. The authors thank D. Haouchar for permitting the use of the 12SMarsMini primers. The authors declare no competing interests. E.R. excavated and collected materials from RCEC, provided dates, stratigraphy and faunal data from morphological analyses. A.M. provided faunal data from morphological analyses and palaeoecological input. A.G. designed, performed and analysed experiments with input from M.A., M.B. and N.R. A.G. wrote the manuscript with contributions and edits from all co-authors.

**Abbreviations.** aDNA, ancient DNA; AMS, accelerator mass spectrometry; BBM, bulk bone metabarcoding; NCWHA, Naracoorte Caves World Heritage Area; NGS, next-generation sequencing; RCEC, Robertson Cave Entrance Chamber.

## Supplementary information

All supplementary data related to this article can be found at in the file 'Supplementary Information' published with the online version of the article. Sequencing data are deposited on the online data repository Data Dryad, and are available at: <http://dx.doi.org/10.5061/dryad.vf345>.

Appendix S1. Radiocarbon dating.

Appendix S2. Supplementary methods.

Appendix S3. Supplementary results.

Appendix S4. Supplementary references.

## References

- Allentoft ME. 2013. Recovering samples for ancient DNA research – guidelines for the field archaeologist. *Antiquity* **87**: 338.
- Allentoft ME, Collins M, Harker D *et al.* 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences* **279**: 4724–4733 [DOI: 10.1098/rspb.2012.1745].
- Allentoft ME, Heller R, Oskam CL *et al.* 2014. Extinct New Zealand megafauna were not in decline before human colonization. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4922–4927 [DOI: 10.1073/pnas.1314972111].
- Altschul SF, Gish W, Miller W *et al.* 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410 [DOI: 10.1016/S0022-2836(05)80360-2].
- Ayliffe LK, Marianelli PC, Moriarty KC *et al.* 1998. 500 ka precipitation record from southeastern Australia: evidence for interglacial relative aridity. *Geology* **26**: 147–150 [DOI: 10.1130/0091-7613(1998)026<0147:KPRFSA>2.3.CO;2].
- Benson DA, Karsch-Mizrachi I, Lipman DJ, *et al.* 2006. Gen-Bank. *Nucleic Acids Research* **34**: D16eD20.
- Blankenberg D, Von Kuster G, Coraro N *et al.* 2010. Galaxy: a web-based genome analysis tool for experimentalists. In *Current Protocols in Molecular Biology* **Unit 19.10**: 11–21.
- Bohmann K, Evans A, Gilbert MTP *et al.* 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution* **29**: 358–367 [DOI: 10.1016/j.tree.2014.04.003].
- Bollongino R, Tresset A, Vigne JD. 2008. Environment and excavation: pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol* **7**: 91–98 [DOI: 10.1016/j.crpv.2008.02.002].
- Briggs AW, Stenzel U, Johnson PLF *et al.* 2007. Patterns of damage in genomic DNA sequences from a Neanderthal. *Proceedings of the National Academy of Sciences* **104**: 14616–14621 [DOI: 10.1073/pnas.0704665104].
- Bronk Ramsey C. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**: 337–360.
- Bunce M, Worthy TH, Ford T *et al.* 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* **425**: 172–175 [DOI: 10.1038/nature01871].

- Bunce M, Worthy TH, Phillips MJ *et al.* 2009. The evolutionary history of the extinct ratite moa and New Zealand Neogene paleogeography. *Proceedings of the National Academy of Sciences* **106**: 20646–20651 [DOI: 10.1073/pnas.0906660106].
- Campos PF, Craig OE, Turner-Walker G *et al.* 2012. DNA in ancient bone – where is it located and how should we extract it? *Annals of Anatomy-Anatomischer Anzeiger* **194**: 7–16.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336 [DOI: 10.1038/nmeth.f.303].
- Cooper A. 1994. Ancient DNA sequences reveal unsuspected phylogenetic relationships within New Zealand wrens (Acanthisittidae). *Experientia* **50**: 558–563 [DOI: 10.1007/BF01921725].
- Corinaldesi C, Beolchini F, Dell'Anno A. 2008. Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Molecular Ecology* **17**: 3939–3951 [DOI: 10.1111/j.1365-294X.2008.03880.x].
- Dabney J, Knapp M, Glocke I *et al.* 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences* **110**: 15758–15763 [DOI: 10.1073/pnas.1314445110].
- Darrénougué N, De Deckker P, Fitzsimmons KE *et al.* 2009. A Late Pleistocene record of aeolian sedimentation in Blanche Cave, Naracoorte, South Australia. *Quaternary Science Reviews* **28**: 2600–2615 [DOI: 10.1016/j.quascirev.2009.05.021].
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460e2461.
- Edgar RC, Haas BJ, Clemente JC, *et al.* 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194e2200.
- Eglinton G, Logan GA, Ambler RP *et al.* 1991. Molecular preservation [and Discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences* **333**: 315–328 [DOI: 10.1098/rstb.1991.0081].
- Elsner J, Schibler J, Hofreiter M *et al.* 2015. Burial condition is the most important factor for mtDNA PCR amplification success in Palaeolithic equid remains from the Alpine foreland. *Archaeological and Anthropological Sciences* **7**: 505–515 [DOI: 10.1007/s12520-014-0213-4].
- Fallon SJ, Fifield LK, Chappell JM 2010. The next chapter in radiocarbon dating at the Australian National University: Status report on the single stage AMS. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **268**: 898–901 [DOI: 10.1016/j.nimb.2009.10.059].
- Forbes MS, Bestland EA. 2007. Origin of the sedimentary deposits of the Naracoorte Caves, South Australia. *Geomorphology* **86**: 369–392 [DOI: 10.1016/j.geomorph.2006.09.009].
- Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**: 737–748 [DOI: 10.1038/nprot.2013.038].
- Giardine B, Riemer C, Hardison RC, *et al.* 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451e1455.
- Gilbert MTP, Willerslev E, Hansen AJ *et al.* 2003. Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics* **72**: 32–47 [DOI: 10.1086/345378].
- Goecks J, Nekrutenko A, Taylor J *et al.* 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**: R86 [DOI: 10.1186/gb-2010-11-8-r86].
- Grealy AC, McDowell MC, Scofield P *et al.* 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. *Quaternary Science Reviews* **128**: 37–47 [DOI: 10.1016/j.quascirev.2015.09.014].
- Gutiérrez-García TA, Vázquez-Domínguez E, Arroyo-Cabrales J *et al.* 2014. Ancient DNA and the tropics: a rodent's tale. *Biology Letters* **10**: 20140224.
- Hadly EA, Ramakrishnan U, Chan YL *et al.* 2004. Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLOS Biology* **2**: 1600–1609 [DOI: 10.1371/journal.pbio.0020290].
- Haile J. 2012. Ancient DNA extraction from soils and sediments. *Methods in Molecular Biology* **840**: 57–63 [DOI: 10.1007/978-1-61779-516-9\_8].
- Hansen AJ, Mitchell DL, Wiuf C, *et al.* 2006. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* **173**: 1175–1179 [DOI: 10.1534/genetics.106.057349].
- Haouchar D, Haile J, McDowell MC *et al.* 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews* **84**: 56–64 [DOI: 10.1016/j.quascirev.2013.11.007].
- Haynes S, Searle JB, Bretman A *et al.* 2002. Bone preservation and ancient DNA: the application of screening methods for predicting DNA survival. *Journal of Archaeological Science* **29**: 585–592 [DOI: 10.1006/jasc.2001.0731].
- Heintzman PD, Elias SA, Moore K *et al.* 2014. Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). *Molecular Ecology Resources* **14**: 606–615 [DOI: 10.1111/1755-0998.12205].
- Hofreiter M, Betancourt JL, Sbriller AP *et al.* 2003. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquén Province, southwest Argentina. *Quaternary Research* **59**: 364–378 [DOI: 10.1016/S0033-5894(03)00030-9].
- Hofreiter M, Jaenicke V, Serre D, *et al.* 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* **29**: 4793–4799 [DOI: 10.1093/nar/29.23.4793].
- Hogg A. 2013. SHCal13 Southern Hemisphere Calibration, 0–50,000 years cal BP. *Radiocarbon* **55**: 1889–1903 [DOI: 10.2458/azu\_js\_rc.55.16783].
- Hoss M, Jaruga P, Zastawny TH *et al.* 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Research* **24**: 1304–1307 [DOI: 10.1093/nar/24.7.1304].
- Hunter AA, Macgregor AB, Szabo TO *et al.* 2012. Yabi: an online research environment for grid, high performance and cloud computing. *Source Code for Biology and Medicine* **7**: 1–1 [DOI: 10.1186/1751-0473-7-1].
- Huson DH, Auch AF, Qi J, *et al.* 2007. MEGAN analysis of metagenomic data. *Genome Research* **17**: 377e386.
- Jaenicke-Despres V, Buckler ES, Smith BD, *et al.* 2003. Early allelic selection in maize as revealed by ancient DNA. *Science* **302**: 1206–1208 [DOI: 10.1126/science.1089056].
- Jørgensen T, Kjaer KH, Haile J *et al.* 2012. Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA meta-barcoding. *Molecular Ecology* **21**: 1980–1988 [DOI: 10.1111/j.1365-294X.2011.05278.x].
- Kearse M, Moir R, Wilson A, *et al.* 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647e1649.
- Knapp M, Clarke AC, Horsburgh KA *et al.* 2012. Setting the stage – building and working in an ancient DNA laboratory. *Annals of Anatomy - Anatomischer Anzeiger* **194**: 3–6 [DOI: 10.1016/j.aanat.2011.03.008].
- Kuhn TS, McFarlane KA, Groves P *et al.* 2010. Modern and ancient DNA reveal recent partial replacement of caribou in the southwest Yukon. *Molecular Ecology* **19**: 1312–1323 [DOI: 10.1111/j.1365-294X.2010.04565.x].
- Llamas B, Brotherton P, Mitchell KJ, *et al.* 2014. Late Pleistocene Australian marsupial DNA clarifies the affinities of extinct megafaunal kangaroos and wallabies. *Molecular Biology and Evolution* **32**: 574–584.
- Leonard JA, Wayne RK, Wheeler J, *et al.* 2002. Ancient DNA evidence for Old World origin of New World dogs. *Science* **298**: 1613–1616 [DOI: 10.1126/science.1076980].
- Macken AC. 2009. *Palaeoecological investigation of the Grant hall fossil deposit, Victoria Fossil Cave, Naracoorte, South Australia*. Flinders University: South Australia.
- Macken AC, Prideaux GJ, Reed EH. 2012. Variation and pattern in the responses of mammal faunas to Late Pleistocene climatic change in southeastern South Australia. *Journal of Quaternary Science* **27**: 415–424 [DOI: 10.1002/jqs.1563].

- Macken AC, Reed EH. 2013. *Late Quaternary small mammal faunas of the Naracoorte Caves World Heritage area*. Transactions of the Royal Society of South Australia **137**: 53–67.
- Macken AC, Reed EH. 2014. Postglacial reorganization of a small-mammal paleocommunity in southern Australia reveals thresholds of change. *Ecological Monographs* **84**: 563–577 [DOI: 10.1890/13-0713.1].
- Magyari EK, Major Á, Bálint M *et al.* 2011. Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evolutionary Biology* **11** [DOI: 10.1186/1471-2148-11-66].
- McDowell MC, Bestland EA, Bertuch F *et al.* 2013. Chronology, strigraphy and palaeoenvironmental interpretation of a Late Pleistocene to mid-Holocene cave accumulation on Kangaroo Island, South Australia. *Boreas* **42**: 974–994.
- Mitchell KJ, Llamas B, Soubrier J *et al.* 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* **344**: 898–900 [DOI: 10.1126/science.1251981].
- Murray DC, Coghlan ML, Bunce M. 2015. From benchtop to desktop: important considerations when designing amplicon sequencing workflows. *PLOS ONE* **10**: e0124671 [DOI: 10.1371/journal.pone.0124671].
- Murray DC, Haile J, Dortch J *et al.* 2013. Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Scientific Reports* **3**: 3371 [DOI: 10.1038/srep03371].
- Murray DC, Pearson SG, Fullagar R *et al.* 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews* **58**: 135–145 [DOI: 10.1016/j.quascirev.2012.10.021].
- O’Keefe K, Ramakrishnan U, van Tuinen M, Hadly EA. 2009. Source-sink dynamics structure a common montane mammal. *Molecular Ecology* **18**: 4775–4789.
- Olalde I, Allentoft ME, Sánchez-Quinto F *et al.* 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**: 225–228 [DOI: 10.1038/nature12960].
- Orlando L, Ginolhac A, Raghavan M *et al.* 2011. True single-molecule DNA sequencing of a Pleistocene horse bone. *Genome Research* **21**: 1705–1719 [DOI: 10.1101/gr.122747.111].
- Oskam CL, Haile J, McLay E *et al.* 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B: Biological Sciences* **277**: 1991–2000 [DOI: 10.1098/rspb.2009.2019].
- Pansu J, Giguet-Covex C, Ficetola GF *et al.* 2015. Reconstructing long-term human impacts on plant communities: an ecological approach based on lake sediment DNA. *Molecular Ecology* **24**: 1485–1498 [DOI: 10.1111/mec.13136].
- Prideaux GJ, Roberts RG, Megirian D *et al.* 2007. Mammalian responses to Pleistocene climate change in southeastern Australia. *Geology* **35**: 33–36 [DOI: 10.1130/G23070A.1].
- Pruvost M, Schwarz R, Correia VB *et al.* 2007. Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences* **104**: 739–744 [DOI: 10.1073/pnas.0610257104].
- Rawlence NJ, Lowe DJ, Wood JR *et al.* 2014. Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *Journal of Quaternary Science* **29**: 610–626 [DOI: 10.1002/jqs.2740].
- Rawlence NJ, Wood JR, Armstrong KN *et al.* 2009. DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society B: Biological Sciences* **276**: 3395–3402 [DOI: 10.1098/rspb.2009.0755].
- Reed EH. 2003. *Vertebrate taphonomy of large mammal bone deposits, Naracoorte Caves World Heritage Area*. PhD Thesis, Flinders University of South Australia.
- Reed EH. 2009. Decomposition and disarticulation of kangaroo carcasses in caves at Naracoorte, South Australia. *Journal of Taphonomy* **7**: 265–283.
- Reed EH. 2012. Of mice and megafauna: new insights into Naracoorte’s fossil deposits. *Journal of the Australasian Cave and Karst Management Association* **86**: 7–14.
- Renfree MB, Papenfuss AT, Deakin JE *et al.* 2011. Genome sequence of an Australian Kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biology* **12**: R81.
- Sawyer S, Krause J, Guschanski K *et al.* 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLOS ONE* **7** [DOI: 10.1371/journal.pone.0034131].
- Shapiro B, Hofreiter M. 2012. *Ancient DNA: Methods and Protocols*. Humana Press (Springer Science): New York.
- Shepherd LD, Lambert DM. 2008. Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology* **17**: 2174–2184 [DOI: 10.1111/j.1365-294X.2008.03749.x].
- Smith CI, Chamberlain AT, Riley MS *et al.* 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**: 203–217 [DOI: 10.1016/S0047-2484(03)00106-4].
- Wells RT, Moriarty K, Williams DLG. 1984. The fossil vertebrate deposits of Victoria Fossil Cave, Naracoorte: an introduction to the geology and fauna. *Australian Zoologist* **21**: 305–333.
- Willerslev E, Cooper A. 2005. Review Paper. *Ancient DNA. Proceedings of the Royal Society B: Biological Sciences* **272**: 3–16 [DOI: 10.1098/rspb.2004.2813].
- Willerslev E, Hansen AJ, Binladen J *et al.* 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**: 791–795 [DOI: 10.1126/science.1084114].
- Wood JR, Wilmshurst JM, Rawlence NJ *et al.* 2013a. A megafauna’s microfauna: gastrointestinal parasites of New Zealand’s extinct moa (Aves: Dinornithiformes). *PLOS ONE* **8** [DOI: 10.1371/journal.pone.0057315].
- Wood JR, Wilmshurst JM, Richardson SJ *et al.* 2013b. Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes). *Proceedings of the National Academy of Sciences* **110**: 16910–16915 [DOI: 10.1073/pnas.1307700110].
- Wood JR, Wilmshurst JM, Worthy TH *et al.* 2012. First coprolite evidence for the diet of *Anomalopteryx didiformis*, an extinct forest ratite from New Zealand. *New Zealand Journal of Ecology* **36**: 164–170.



# Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar

Alicia Grealy<sup>a, \*</sup>, Kristina Douglass<sup>b</sup>, James Haile<sup>c</sup>, Chriselle Bruwer<sup>d, e</sup>, Charlotte Gough<sup>f</sup>, Michael Bunce<sup>a</sup>

<sup>a</sup> Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

<sup>b</sup> Department of Anthropology, National Museum of Natural History, Smithsonian Institution, USA

<sup>c</sup> PalaeoBARN, Research Laboratory for Archaeology and the History of Art, Oxford University, UK

<sup>d</sup> Department of Anthropology and Archaeology, University of South Africa, Pretoria, South Africa

<sup>e</sup> Arc Maps and Graphics, Pretoria, South Africa

<sup>f</sup> Blue Ventures Conservation, London, N7 9DP, UK

## ARTICLE INFO

*Article history:*  
Received 15 March 2016  
Received in revised form  
25 August 2016  
Accepted 4 October 2016

*Keywords:*  
Ancient DNA  
Archaeology  
Biodiversity  
Bulk-bone  
Fish  
Madagascar  
Metabarcoding  
Subsistence

## ABSTRACT

Taxonomic identification of archaeological fish bones provides important insights into the subsistence practices of ancient coastal peoples. However, it can be difficult to execute robust morphological identification of fish bones from species-rich fossil assemblages, especially from post-cranial material with few distinguishing features. Fragmentation, weathering and burning further impede taxonomic identification, resulting in large numbers of unidentifiable bones from archaeological sites. This limitation can be somewhat mitigated by taking an ancient DNA (aDNA) bulk-bone metabarcoding (BBM) approach to faunal identification, where DNA from non-diagnostic bone fragments is extracted and sequenced in parallel. However, a large proportion of fishing communities (both past and present) live in tropical regions that have sub-optimal conditions for long-term aDNA preservation. To date, the BBM method has never been applied to fish bones before, or to fossils excavated from an exposed context within a tropical climate. Here, we demonstrate that morphologically indistinct bulk fish bone from the tropics can be identified by sequencing aDNA extracted from 100 to 300 ya archaeological midden material in southwest Madagascar. Despite the biases of the approach, we rapidly obtained family, genus, and species-level assemblage information, and used this to describe a subset of the ichthyofauna exploited by an 18th century fishing community. We identified 23 families of fish, including benthic, pelagic, and coral-dwelling fishes, suggesting a reliance on a variety of marine and brackish habitats. When possible, BBM should be used alongside osteological approaches to address the limitations of both; however, this study highlights how genetic methods can nevertheless be a valuable tool for helping resolve faunal assemblages when morphological identification is hindered by taphonomic processes, lack of adequate comparative collections, and time constraints, and can provide a temporal perspective on fish biodiversity in the context of accelerated exploitation of the marine environment.

© 2016 Elsevier Ltd. All rights reserved.

## 1. Introduction

The identification of archaeological fish bone offers important insights into the subsistence practices of ancient fishing communities. Fish are a staple food for coastal peoples throughout the world; modern estimates suggest that approximately 60% of the

global population lives within 100 km of the coast (Erlandson and Rick, 2008). As such, there is tremendous potential for archaeological data and interpretations to provide a long-term perspective that can inform present-day marine resource management and conservation policies (Braje, 2010; Braje et al., 2015; Lambrides and Weisler, 2016; Speller et al., 2012). Fine-grained archaeological investigations of resource exploitation patterns are especially important because human arrival in many regions of the world has been correlated with an increase in faunal extinctions, implying that over-exploitation of local fauna has contributed to significant

\* Corresponding author.

E-mail address: [alicia.grealy@uqconnect.edu.au](mailto:alicia.grealy@uqconnect.edu.au) (A. Grealy).

loss of biodiversity (Braje and Erlandson, 2013). The coincidence of human colonisation and declines in floral and faunal diversity is particularly acute in island contexts (Rick et al., 2013). One example is Madagascar, which—although still considered a biodiversity ‘hotspot’ (Myers et al., 2000)—has undergone a significant loss of biodiversity over the last two thousand years that has long been thought to coincide with human arrival on the island (De la Bâthie, 1921; Humbert, 1927). However, many questions remain as to the timing of environmental change in Madagascar and the role early communities played in shaping the island's land and seascapes, particularly given the challenges of investigating early forager sites (Douglass and Zinke, 2015). Furthermore, disentangling anthropogenic and climatic drivers of environmental change remains a central research concern in Madagascar, since the island's climate and environment were in constant flux well before human colonisation (Dewar and Richard, 2007; Douglass and Zinke, 2015). Moreover, despite the fact that Madagascar is an island, the historical ecology of Madagascar's marine and coastal environment has received little research attention. Instead, as is the case in other parts of the world (Erlandson and Rick, 2008), far more archaeological and paleontological work has been directed at understanding anthropogenic impacts on terrestrial ecologies.

Determining how humans impacted the marine environment of Madagascar during the Holocene relies on a thorough understanding of the marine taxa that were targeted by ancient communities. Burnt or modified fish bone, or fish bones found in cultural deposits, are good gauges of direct human interaction with marine biota. As such, the identification of archaeological fish bone is essential to uncovering marine prey targets; to date, few studies comprehensively achieve this (Lambrides and Weisler, 2016), largely because of the limitations to morphological identification. The identification of fish bone predominantly relies on the examination of size range and diagnostic osteological features (Lambrides and Weisler, 2015). Cranial elements, such as teeth, are particularly important in refining taxonomic identification. However, the cranium and teeth generally make up a small proportion of the overall number of bones recovered from archaeological deposits (Yang et al., 2004); for each cranium, there may be over three times as many post-cranial elements (Jones, 2009), including vertebrae, ribs, spines, and rays. Articulated specimens are even rarer because fish remains are fragile and susceptible to damage during food preparation, cooking, and consumption, as well as to post-depositional weathering (Collins, 2010). Vertebrae and ribs of many fish species are often difficult, if not impossible, to distinguish from one another as they display little variation between species (Teletchea, 2009). To complicate matters further, many fish exhibit different morphology throughout their development from juvenile to adult, and can also display high intraspecific morphological variability, sexual dimorphism (Teletchea, 2009), and phenotypic plasticity (Lambrides and Weisler, 2015). Depositional bias, taphonomy, and lack of diagnostic features hinder morphological taxonomic identifications in many archaeological assemblages of fish bones, and often result in large numbers of unidentified remains.

Ancient DNA (aDNA) is a complementary method to the study of faunal remains, as it does not rely on the preservation of diagnostic morphological features. However, in Sub-Saharan African contexts, studies of archaeological aDNA are rare, despite the potential for aDNA analyses to complement traditional approaches to questions of human-environment interaction (Campana et al., 2013; Gifford-Gonzalez, 2013). For fish, DNA reference collections represent a large portion of fish diversity, and DNA analysis has been used to discriminate cryptic species and morphotypes; for example, the genus *Schindleria* consists of 21 genetically distinct but morphologically cryptic species (Kon et al., 2007), while the

morphologically different *Eumicrotremus spinosus* and *E. eggvini* constitute a single species (Byrkjedal et al., 2007). Ancient DNA has also proved to be a useful tool in studies of archaeological fish assemblages (Campana et al., 2013), albeit in a relatively small number of studies (Teletchea, 2009): in a literature search, only approximately 2.5% of articles published on archaeological aDNA relate to fish. However, the studies that have been published demonstrate the value of such an approach in garnering important information about species diversity and distribution in the past (Cannon and Yang, 2006; Grier et al., 2013; Speller et al., 2005, 2012; Yang et al., 2004), and the economic importance of different fish taxa to ancient communities (Nikulina and Schmölcke, 2015).

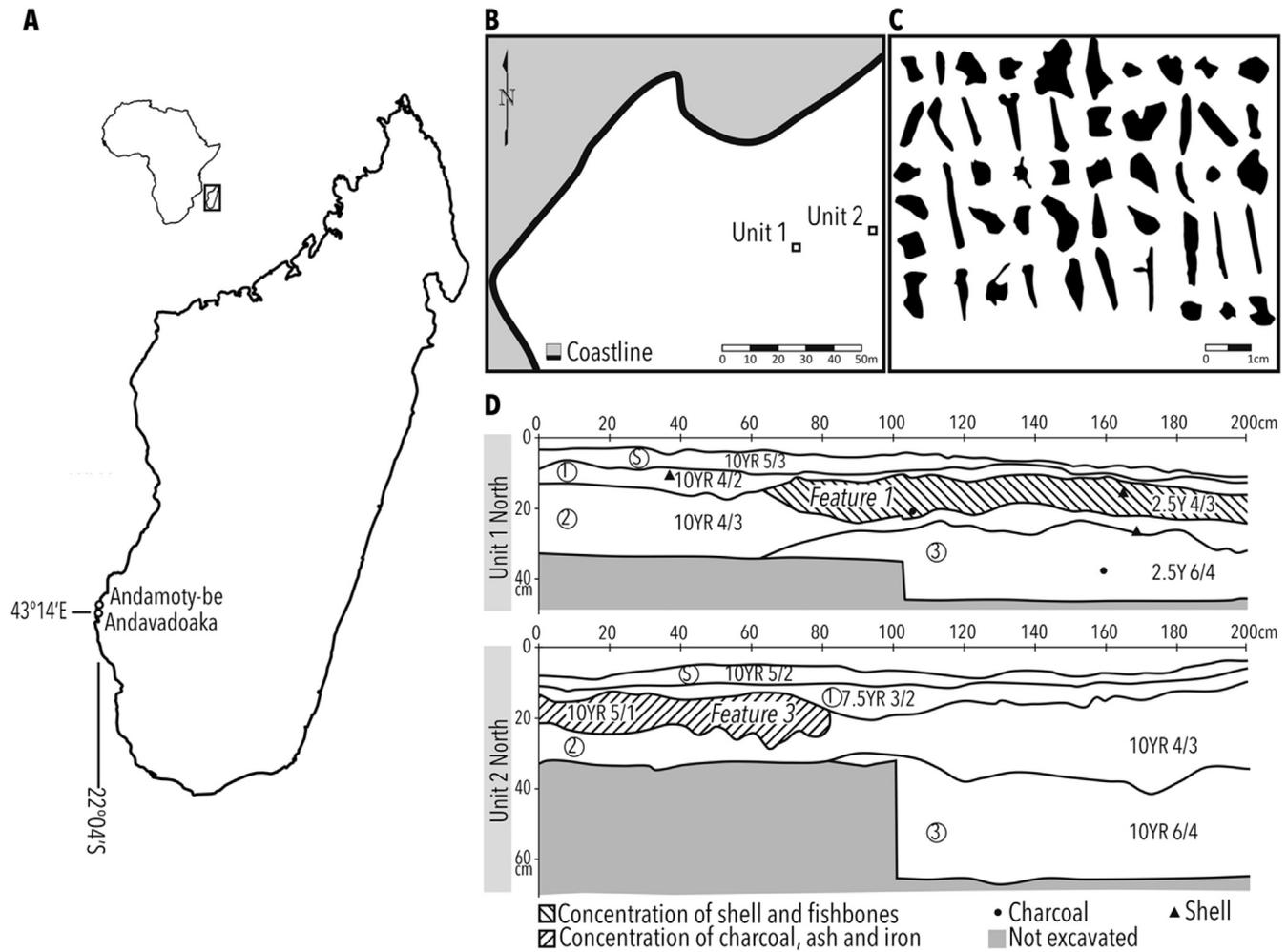
The infrequent use of aDNA techniques in the analysis of archaeological fish assemblages may be due to the fact that fish bones are often too small and numerous to warrant the high cost of individual DNA extraction and sequencing—especially in the tropics that have one of the highest biodiversities of fish in the world (Lambrides and Weisler, 2015). Furthermore, the majority of fishing communities occupy ‘exposed’ sites in coastal tropical and subtropical zones, with a consistently hot climate that fluctuates annually between dry and humid. These landscapes often lack natural and permanent shelter formations (such as caves), and are not typically conducive to aDNA preservation. Nevertheless, aDNA has been retrieved from tropical zones before (e.g., Gutiérrez-García et al., 2014; Murray et al., 2012; Nicholls et al., 2003; Schroeder et al., 2015), including Madagascar (Kistler et al., 2014; Mitchell et al., 2014; Orlando et al., 2008), and the innovation of new methods promises to increase the successful application of aDNA analysis on materials collected in tropical localities.

The recently developed ‘bulk-bone metabarcoding’ (BBM) approach is one such method that allows the DNA from many bones to be extracted, amplified, and sequenced in parallel to rapidly and accurately identify many of the taxa within a sub-fossil assemblage (Grealy et al., 2015, 2016; Haouchar et al., 2014; Murray et al., 2013), which can increase the probability of characterising tropical archeofish remains. Here, we demonstrate how the BBM method can retrieve molecular taxonomic information from Malagasy 100–300 ya archaeological fish bone fragments that can then be used to examine past interactions of humans with their marine environment.

## 2. Materials and methods

### 2.1. Site description and dating

The coastal ‘Andamoty-be’ archaeological site is located just north of the village of Andavadoaka (22° 04’S, 43° 14’E) in Toliara province, Southwest Madagascar (Fig. 1a and b), and was excavated in June 2014 by K. Douglass. The site is bordered on the east by spiny forest and by the Mozambique Channel to the west. It is located within the Velondriake Marine Protected Area, a locally managed marine area (LMMA) that encompasses the longest continuous reef system in Madagascar and is protected under the International Union for Conservation of Nature (IUCN). A decade of reef and fisheries monitoring by Blue Ventures Conservation has generated an excellent modern record of marine biodiversity within Velondriake's shallow reef flats, sand flats, macro-algae, seagrass and mangrove habitats (Cripps, 2009; Cripps et al., 2015; Hantanirina and Benbow, 2013; Harris et al., 2010; Jones et al., 2014; Nadon et al., 2007; Roy et al., 2009). Human occupation at the site is estimated to date between 100 and 300 ya based on the presence of imported 19th century British stoneware ceramics and Venetian glass trade beads found in the accumulation. The site has been described in detail elsewhere (Douglass, 2016).



**Fig. 1.** Map of Madagascar showing the location of **A** the archaeological site examined, and **B** placement of the excavation units; **C** Silhouettes of a representative pool of 50 bones as an example of the typical size and shape of bones from the archaeological accumulation; **D** North-wall profiles of the stratigraphy for each unit depicting the layers excavated (rendered by C. Bruwer; vertical axis represents depth). Note that these are examples of the stratigraphy and do not depict all contexts (for further detail refer to [Douglass, 2016](#)).

Two replicate 2 m × 2 m units were placed on areas with the highest density of surface scatter approximately 20 m apart and 500 m from the shoreline. Layers were excavated following the natural stratigraphy, resulting in four layers per unit with multiple sub-contexts within layers (Fig. 1d). Excavated material was sieved on-site using 2 mm × 2 mm mesh screens. Bones were subsequently sorted from cultural material and stored at room temperature. Gloves and facemasks were worn during excavation, sieving, and sorting to minimise contamination with modern DNA. Gloves did not come into contact with modern fish at any time during the excavation to ensure that no contamination by modern fish DNA was introduced to the samples during collection.

## 2.2. Sample preparation

Where possible, one pool of 50 bones (Fig. 1c) was randomly generated for each context for a total of 13 pools (note that two contexts contained fewer than 50 bones; for these, all bones were pooled). Three additional pools of 50 bones were generated for the first layer in each unit (6 total) that targeted fish vertebrae fragments. A total of 887 bones were sampled, with each bone having an average mass of 123.5 mg. Approximately 20 mg of bone was subsampled from each bone within a pool and these were ground into a fine powder using the *Retsch* PM200 Planetary Ball Mill at

500 rpm for 5 min. Powder was stored at  $-20^{\circ}\text{C}$ . All sample preparation was conducted in an isolated ultra-clean environment within Curtin University's TRACE facility (WA, Australia) following standard aDNA protocols for contamination avoidance ([Willerslev and Cooper, 2005](#); [Knapp et al., 2012](#)).

## 2.3. aDNA extraction

aDNA was extracted from 100 mg of bone powder for each pool, following the methods described by [Grealy et al. \(2016\)](#). DNA-free controls were included throughout the extraction processes and were carried through to sequencing.

## 2.4. Metabarcoding and next-generation sequencing

Primers targeting typical barcoding genes *CO1* and *Cytb* tend to amplify regions that are too long to capture degraded DNA fragments of ancient samples ([Jordan et al., 2010](#)). Therefore, aDNA extracts were amplified via qPCR using a primer set designed to target conserved regions of the fish 12S rRNA mitochondrial gene. At  $53^{\circ}\text{C}$ , these primers (12S 5'-CGCCTATATACCRCCGTC-3' and 5'-CRCTACACCTCGACCTG-3', flanked by unique indexes and *Illumina* sequencing adapters) amplify a 56 bp variable barcoding region from local members within more than 60 fish families. *In silico*

analysis of the primer-binding sites shows that they are conserved across modern taxa found in the area, and are not likely to be inherently more biased towards the detection of any one taxon over another (SI 1.0, Fig. S1). In most cases, the metabarcoding region differs by five or more base pairs between taxa of interest (Table S1), and it is unlikely that the combined effect DNA damage, amplification error, and sequencing error would result in taxonomic mis-identifications (SI 1.0, Table S2). Amplification, subsequent sequencing on the MiSeq platform, trimming, and quality control were performed as per Grealy et al. (2016). DNA sequences are available on the online data repository Data Dryad and can be accessed via the doi:10.5061/dryad.8d1p9.

### 2.5. Taxonomic assignment

Taxonomy was assigned to sequences by comparison with NCBI's GenBank (Benson et al., 2006) nucleotide reference database using BLASTn (default parameters; Altschul et al., 1990) implemented through the Pawsey Supercomputing Centre (WA, Australia), and examination in MetaGenome Analyser (MEGAN v. 4.70.4; Huson et al., 2007) as per Grealy et al. (2016). Identifications were based on the similarity of query and reference sequence across 100% of the query, with similarity cut-offs for species-level IDs at >98% similarity, genus-level IDs at 95–98% similarity, and family-level IDs at 90–95% similarity. Assignments were assigned a credibility rating (highly credible, credible, or unlikely; Table 1) based on whether the taxa are found in area according to species' distribution records defined by FishBase (Froese and Pauly, 2015), and whether genetic reference sequences exist in GenBank for all subtaxa within family or genus.

## 3. Results and discussion

Next-generation sequencing of 56 bp 12S rRNA sequences amplified from bulk fish bone aDNA generated a total of 77,298 reads (an average of 4024 reads per pool) and 1338 unique reads (an average of 70 unique reads per pool). After molecular taxonomic identification by comparison with a reference database, 23 families were identified with high credibility; within them, 14 genera were able to be identified with high credibility, 4 credibly, and within those, 5 species were able to be identified with high credibility and 5 credibly (Table 1). No fish DNA was amplified in any of the controls, indicating that contamination from the laboratory environment was below detectable levels.

Despite the large diversity of taxa identified, habitat associations derived from modern fisheries data (Blue Ventures) suggest a primary reliance on near-shore reef-dwelling fish, with 53% of identified families associating with coral reefs, and the remainder associating with seagrass (8%), mixed habitats (11%), or unknown (28%). These data suggest that there was a strong dependence on coral communities in terms of the exploitation of marine taxa. It appears that a range of fish sizes were targeted, and although not enough sampling has been done thus far to draw comparisons between layers, species within the family Lethrinidae (followed by Scaridae and Serranidae) were detected in more samples than any other taxa. This suggests that these typically large-bodied, high trophic-level fish may have constituted the primary staple marine food of people at Andamoty-be; in contrast, these families represent a small percentage of the catch in near-by Morombe today (Laroche et al., 1997). At Morombe, "high fishing pressure [has] led to a concentration of effort on lower trophic level species to maintain catch levels" (Laroche et al., 1997), an example of "fishing down marine food webs" (Pauly et al., 1998). This may indicate that fishing pressure in this region has increased over several hundred years, and that line fishing (the predominant method employed for

catching high trophic-level fishes) may have been more commonly practiced by ancient communities than it is at present; today, only about 6% of catch in the Morombe region is by line (Laroche et al., 1997).

Other fish families detected include reef dwelling fish of the Chaetodontidae and Pomacentridae families (a mix of corallivores, planktivores, omnivores and herbivores, some of which are small and may have been used as bait fish), carnivores such as the wrasses of the Labridae family, members of the Carangidae family, and members of the Sparidae family. *Megalops cyrinoides* (Indo-Pacific tarpon) of the Megalopidae family are typically migratory fish that move between open water and inland rivers (Merrick and Schmida, 1984). In the modern fisheries data (Blue Ventures), Megalopidae were recorded from catch in the coral habitat, suggesting that although adult fish could have been caught in the open sea beyond the barrier reef, they may have been netted, as adolescent *M. cyrinoides* migrate offshore from estuarine waters and mangroves (Coates, 1987). Pelagic fish, like members of the Chanidae family (*Chanos chanos*, or milkfish), also possibly indicate open sea fishing; these fish live in large schools in surface waters over the continental shelf and generally require sophisticated fishing methods, including nets, larger outrigger canoes than are needed for fishing around coral reefs, and potentially the co-operation of several boats (Wheeler and Jones, 1989). However, like members of the Megalopidae family, milkfish do migrate into brackish waters (including mangroves, estuaries, and lakes) as juveniles and return to the sea to sexually mature (Froese and Pauly, 2015). Requiem sharks (*Carcharhinus*) are also known to occur in brackish and freshwater habitats. The presence of demersal fish such as whiprays (*Himantura*) may indicate the practice of bottom trawling or line fishing. In addition to serving as a food source, stingrays are commonly sought out in Velondriake today as a valuable source of abrasive material, and their tails are used as a tool to shape and sand wood (Douglass, 2012).

Several taxa identified have not been recorded in modern fisheries data (Blue Ventures). These include carnivores such as *Psammodon waigiensis* (Waigeo barramundi) of the Latidae family and Ginglymostomatidae (nurse sharks), fresh-water fishes of the Cichlidae family, and forage-fishes of the Clupeidae family. The detection of nocturnal predators such as nurse shark and Waigeo barramundi may be an indication of night fishing, dive fishing, or leaving nets out overnight. In particular, nurse sharks are bottom feeders that live in shallow inshore waters with coral communities. Overfishing may be responsible for the rarity of these sharks today (Cooke, 1997), indicating that there has been significant anthropogenic impact on the environment by past people. Furthermore, the "season of abundance" for sharks is predominantly April to July (Langley, 2006), which may indicate that this site was inhabited during the cooler, dry season.

Detection of the Cichlidae family is interesting as this is the only non-marine family identified in the archaeological assemblage thus far, and no cichlid catch is recorded in the modern fisheries data (Blue Ventures). Cichlids are a diverse family, with 28 endemic and 9 introduced species known from Madagascar (Froese and Pauly, 2015). This identification is strongly suggestive of fresh and/or brackish water fishing by local people. The closest occurrence of cichlid species to Andamoty-be is the Onilahy river basin's *Ptychochromoides betsileanus* and *Ptychochromis onilahy*: these species are classified by the IUCN as critically endangered and extinct (respectively) as a result of habitat loss, fishing, and competition or predation by introduced species. The Onilahy River is located approximately 180 km south of Andamoty-be, so the presence of a cichlid at Andamoty-be could be an indication that the ranges of one or both of these species extended as far as Andamoty-be in the past, but underwent a range contraction as a result of human



pressures. With more research, a historic range could potentially be established for these species, which may inform conservation efforts (Hofman et al., 2015; Speller et al., 2012). To confirm the identification of cichlids, additional samples and metabarcoding genes should be sequenced.

Finally, the detection of Clupeidae DNA (100% sequence similarity to *Sardina pilchardus*) in only the surface scatter layer is likely to have been derived from contamination by imported sardines: they are not native species but are a common component of human diet in Madagascar today. Although other sardine genera have been recorded in the modern fisheries data, these are generally well represented in genetic databases and their sequences differ from *Sardina pilchardus* by more than 9%, making it unlikely that this DNA originated from native sardine species.

Like osteological approaches, not all taxa can be identified to the species-level, and some taxa are likely to have not been detected at all (c.f., Grealy et al., 2015 for an in-depth discussion of the biases and limitations of the method where a direct comparison with a morphological approach was possible): Table 1 shows that some taxa are not consistently detected between replicates, highlighting that the ability to detect a given taxon in a complex mixture can be variable. Similar to other metabarcoding approaches such as bacterial metagenomics or environmental DNA monitoring, this ability to identify taxa genetically is influenced by: (1) the intensity of sampling, (2) unique taphonomic biases that affect DNA preservation (including pH, temperature fluctuations, and exposure to humidity), (3) collection and storage (such as handling that can introduce contamination), (4) choice of barcoding region (high intraspecific variation at a locus can affect taxonomic resolution), (5) incomplete reference genetic databases (while comprehensive, some taxa may not be represented), and (6) DNA damage, PCR bias, and sequencing error (although some of these can be mitigated by adequate sequencing depth, PCR replication, diluting inhibitors, and stringent quality control). Nevertheless, the detection of one taxon is not undermined by an inability to detect another, although we cannot confidently estimate what we did *not* find. In addition, although the amplification of longer fragments may have resulted in more refined taxonomic identifications, the degradation of aDNA in tropical environments typically results in the majority of fragments being very short; as such, there is a trade-off between the breadth of taxa identified and the specificity of identification (Grealy et al., 2015). Analysis of more samples, amplification of additional barcoding genes, and revisiting the existing data as genetic databases become more complete, will also deliver more fine-scaled molecular identifications and identify additional diversity. While not a complete audit of the past fish diversity, this is the first published description of an archaeological fish assemblage from southwest Madagascar, and demonstrates that a genetic approach provides useful zooarchaeological information in the absence of an alternative. The analysis of additional DNA extracts in the future will allow us to potentially compare archaeofish biodiversity between Andamoty-be and other archaeological sites in Madagascar.

#### 4. Conclusions

This study has established the first published marine zooarchaeological record for Velondriake, offering insights into how past coastal communities derived a livelihood from local marine resources. This is particularly important to establish as accelerated rates of environmental degradation, resource over-exploitation and loss of faunal diversity in recent times have generated important concerns about the future of Madagascar's natural communities and the ability of human communities to derive sustainable livelihoods, especially in a region where more than a third of the population currently engage in sea fishing (Laroche et al., 1997). The

data presented here provide a baseline upon which future data collection and analysis may build, and knowledge of historic biodiversity and human exploitation of the marine environment may assist in conservation and management decisions. *Post-hoc* comparisons with morphological analysis of fossil assemblages in Velondriake will confirm the accuracy (or otherwise) of the molecular identifications. Nevertheless, this study suggests that other archaeological sites around Madagascar, and in other tropical regions, may benefit from aDNA analysis of bulk bone to expand the taxonomic identifications obtained through traditional methods, and hopefully will encourage more fruitful collaborations between geneticists and archaeologists.

#### Author contributions

KD organised and directed the archaeological excavation. AG and JH assisted with the collection of bulk bone material. AG conducted genetic analyses with assistance from MB. JH designed the primers used. CG provided modern fisheries data for comparison. CB rendered the line drawings and figure. AG and KD drafted the manuscript with contributions and edits from all co-authors.

#### Acknowledgements

The archaeological investigations carried out in the Velondriake Marine Protected Area, were made possible with funding from the National Science Foundation Graduate Research Fellowship Program, the P.E.O. Scholar Award, the Yale Institute of Biospheric Studies, the Yale MacMillan Centre for International and Area Studies and the Yale Council on Archaeological Studies. Research permissions were granted by the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Autorisation Numéro 128/13-MESupReS/SG/DGRP and by the Centre de Documentation et de Recherche sur l'Art et les Traditions Orales Malgaches (CEDRATOM), under the auspices of the Memorandum of Understanding between the University of Toliara, under the direction of Dr. Barthélémy Manjakahery, Director of the CEDRATOM, and Yale University, under the direction of Dr. Roderick McIntosh, Professor of Anthropology. Local permission to carry out archaeological research was granted by the Office du Maire, Commune de Befandefa and by the Chefs de Fokontany of Andavadoaka, Nosy Ve, Antsaragnangy, Lamboara, Ampasilava and Salary. Permits for the export of archaeological materials for the purposes of laboratory analysis were granted by the Secretariat Général of the Ministère de l'Artisanat de la Culture et des Patrimoines, Direction Régionale de la Culture et du Patrimoine Atsimo Andrefana, Visas de Sorties Numéro 09/06-MCP/SG/DRCP.AA; Numéro 05/14-MACP/SG/DRCP.AA; Numéro 08/14-MACP/SG/DRCP.AA in accordance with Avis Numéro 375, 02/02/1978. We acknowledge the support of the Australian Research Council Discovery Project grant DP160104473 (MB).

We give special thanks to the Morombe Archaeological Project (MAP) team, to the people of Andavadoake, Madagascar, and to Blue Ventures conservation for sharing data pertaining to modern Velondriake Fisheries. We would also like to thank James Taylor for his assistance with sample preparation and sequencing, and Joey DiBattista for offering insights into fish behavior. High-throughput BLAST searches were performed through the Pawsey Supercomputing Centre (Perth).

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jas.2016.10.001>.

## References

- Altschul, S.F., Gish, W., Miller, W., et al., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Benson, D.A., Karsch-Mizrachi, L., Lipman, D.J., et al., 2006. Gen-Bank. *Nucleic Acids Res.* 34, D16eD20.
- Braje, T.J., 2010. Modern Oceans, Ancient Sites: Archaeology and Marine Conservation on San Miguel Island. University of Utah Press, California, Salt Lake City.
- Braje, T.J., Erlandson, 2013. Human acceleration of animal and plant extinctions: a Late Pleistocene, Holocene and Anthropocene continuum. *Anthropocene* 4, 14–23.
- Braje, T.J., Rick, T.C., Erlandson, J.M., Rogers-Bennett, L., Catton, C.A., 2015. Historical ecology can inform restoration site selection: the case of black abalone (*Haliotis cracherodii*) along California's Channel Islands. *Aquatic Conserv. Mar. Freshw. Ecosyst.* 26, 470–481.
- Byrkjedal, I., Rees, D.J., Willassen, E., 2007. Lumping lumpsuckers: molecular and morphological insights into the taxonomic status of *Eumicrotremus spinosus* (Fabricius, 1776) and *Eumicrotremus eggvinii* (Koefoed, 1956) (Teleostei: Cyclopteridae). *J. Fish Biol.* 71, 111–131.
- Campana, M.G., Bower, M.A., Crabtree, P.J., 2013. Ancient DNA for the archaeologist: the future of African research. *Afr. Archaeol. Rev.* 30, 21–37.
- Cannon, A., Yang, D.Y., 2006. Early storage and sedimentation on the Pacific Northwest coast: ancient DNA analysis of salmon remains from Namu, British Columbia. *Am. Antiq.* 71, 123–140.
- Coates, D., 1987. Observations on the biology of tarpon, *Megalops cyprinoides* (Broussonet) (Pisces: Megalopidae), in the Sepik river, Northern Papua New Guinea. *Aust. J. Mar. Freshw. Res.* 38, 529–535.
- Collins, B.R., 2010. Element survivability of *Salmo salar*. *J. Taphon.* 8, 291–300.
- Cooke, A.J., 1997. Survey of Elasmobranch fisheries and trade in Madagascar. In: Marshall, N.T., Barnett, R. (Eds.), *The Trade Shark and Shark Products in the Western Indian and Southeast Atlantic Oceans*. Graphcor, South Africa.
- Cripps, G., 2009. Understanding Migration Amongst the Traditional Fishers of West Madagascar. Blue Ventures Conservation Report for ReCoMaP.
- Cripps, G., Harris, A., Humber, F., Harding, S., Thomas, T., 2015. A Preliminary Value Chain Analysis of Shark Fisheries in Madagascar. Indian Ocean Commission. SF/2–15/34.
- De la Bâthie, H.P., 1921. La végétation malgache. *Ann. Muséum Colon. Marseille* 9, 1–266.
- Dewar, R.E., Richard, A.F., 2007. Evolution in the hypervariable environment of Madagascar. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13723–13727.
- Douglas S (2012) Fisher's of a song: community music education in Andavadoaka, Madagascar. A thesis presented to the Faculty of the Steinhardt School of Culture, Education and Human Development at New York University in Candidacy for the Degree of Master of Arts in Music Education.
- Douglass, K., Zinke, J., 2015. Forging ahead by land and by sea: archaeology and paleoclimate reconstruction in Madagascar. *Afr. Archaeol. Rev.* 32, 267–299.
- Douglass, K., 2016. An Archaeological Investigation of Settlement and Resource Exploitation Patterns in the Velondriake Marine Protected Area, Southwest Madagascar, Ca. 900 BC to AD 1900. A dissertation presented to the Faculty of the Graduate School of Yale University in Candidacy for the Degree of Doctor of Philosophy.
- Erlandson, J.M., Rick, T.C., 2008. *Archaeology, Marine Ecology, and Human Impacts on Marine Environments*. University of California Press, Berkeley, pp. 1–19.
- Froese, R., Pauly, D., 2015. FishBase. Available. [www.fishbase.org](http://www.fishbase.org). Accessed: Jul 2015.
- Gifford-Gonzalez, D., 2013. Animal genetics and African archaeology: why it matters. *Afr. Archaeol. Rev.* 30, 1–20.
- Grealy, A., McDowell, M., Scofield, P., Murray, D., Fusco, D., Haile, J., Prideaux, G., Bunce, M., 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional palaeontological methods. *Quat. Sci. Rev.* 128, 37–47.
- Grealy, A., Macken, A., Allentoft, M.E., Rawlence, N.J., Reed, E., Bunce, M., 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *J. Quat. Sci.* 9999, 1–13.
- Grier, C., Flanagan, K., Winters, M., Jordan, L.G., Lukowski, S., Kemp, B.M., 2013. Using ancient DNA identification and osteometric measures of archaeological Pacific salmon vertebrae for reconstructing salmon fisheries and site seasonality at Dionisio Point, British Columbia. *J. Archaeol. Sci.* 40, 544–555.
- Gutiérrez-García, T.A., Vázquez-Domínguez, E., Arroyo-Cabrales, J., Kuch, M., Enk, J., King, C., Poinar, H.N., 2014. Ancient DNA and the tropics: a rodent's tale. *Biol. Lett.* 10, 20140224.
- Hantaniirina, J.M.O., Benbow, S., 2013. Diversity and coverage of seagrass ecosystems in south-west Madagascar. *Afr. J. Mar. Sci.* 35, 291–297.
- Haouchar, D., Haile, J., McDowell, M.C., Murray, D.C., White, N.E., Allcock, R.J.N., Phillips, M.J., Prideaux, G.J., Bunce, M., 2014. Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene-Holocene cave deposit on Kangaroo Island, South Australia. *Quat. Sci. Rev.* 84, 56–64.
- Harris, A., Manahira, G., Sheppard, A., Gough, C., Sheppard, C., 2010. Demise of Madagascar's once great barrier reef: change in coral reef condition over 40 years. *Atoll Res. Bull.* 574, 1–16.
- Hofman, C.A., Rick, T.C., Fleischer, R.C., Maldonado, J.E., 2015. Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends Ecol. Evol.* 30, 540–549.
- Humbert, H., 1927. La Destruction D'une Flore Insulaire Par Le Feu. Principaux Aspects de la Végétation À Madagascar. Documents Photographiques Et Notices. Mémoires l'Académie Malgache 5, 1–80.
- Huson, D.H., Auch, A.F., Qi, J., et al., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377e386.
- Jones, S., 2009. *Food and Gender in Fiji: Ethnoarchaeological Explorations*. Lexington Books, USA.
- Jones, T., Ratsimba, H., Ravaorinoroisoarana, L., Cripps, G., Bey, A., 2014. Ecological variability and carbon stock estimates of mangrove ecosystems in North-western Madagascar. *Forests* 5, 177–205.
- Jordan, L.G., Steele, C.A., Thorgaard, G.H., 2010. Universal mtDNA primers for species identification of degraded bony fish samples. *Mol. Ecol. Resour.* 10, 225–228.
- Kistler, L., Ratan, A., Godfrey, L.R., Crowley, B.E., Hughes, C.E., Lei, R., Cui, Y., Wood, M.L., Muldoon, K.M., Andriamialison, H., McGraw, J.J., Tomsho, L.P., Schuster, S.C., Miller, W., Louis, E.E., Yoder, A.D., Malhi, R.S., Perry, G.H., 2014. Comparative and population mitogenomic analyses of Madagascar's extinct, giant 'subfossil' lemurs. *J. Hum. Evol.* 79, 45–54.
- Knapp, M., Clarke, A.C., Horsburgh, K.A., Matisoo-Smith, E.A., 2012. Setting the stage—building and working in an ancient DNA laboratory. *Ann. Anat. Anat. Anz.* 194, 3–6.
- Kon, T., Yoshinto, T., Mukai, T., et al., 2007. DNA sequences identify numerous cryptic species of vertebrate: a lesson from the gobioid fish *Schindleria*. *Mol. Phylogenet. Evol.* 44, 53–62.
- Lambrides, A.B.J., Weisler, M.I., 2015. Applications of vertebral morphometrics in Pacific Island archaeological fishing studies. *Archaeol. Ocean.* 50, 43–70.
- Lambrides, A.B.J., Weisler, M.I., 2016. Pacific Islands ichthyoarchaeology: implications for the development of prehistoric fishing studies and global sustainability. *J. Archaeol. Res.* 1–50.
- Langley, J., 2006. *Vezo Knowledge: Traditional Ecological Knowledge in Andavadoaka, Southwest Madagascar*. Blue Ventures, London.
- Laroche, J., Razanoelisoa, J., Fauroux, E., Rabenevanana, M.W., 1997. The reef fisheries surrounding the south-west coastal cities of Madagascar. *Fish. Manag. Ecol.* 4, 285–299.
- Merrick, J.R., Schmida, G.E., 1984. *Australian Freshwater Fishes—biology and Management*. Griffin Press, South Australia.
- Mitchell, K., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S.Y., Cooper, A., 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* 344, 898–900.
- Murray, D.C., Pearson, S.G., Fullagar, R., Chase, B.M., Houston, J., Atchison, J., White, N.E., Bellgard, M.L., Clarke, E., Macphail, M., Gilbert, M.T.P., Haile, J., Bunce, M., 2012. High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quat. Sci. Rev.* 58, 135–145.
- Murray, D.C., Haile, J., Dortch, J., White, N.E., Haouchar, D., Bellgard, M.L., Allcock, R.J., Prideaux, G.J., Bunce, M., 2013. Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. *Sci. Rep.* 3, 3371.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.
- Nadon, M.O., Griffiths, D., Doherty, E., Harris, A., 2007. The status of coral reefs in the remote region of Andavadoaka, Southwest Madagascar. *Western Indian Ocean. J. Mar. Sci.* 6, 207–218.
- Nicholls, A., Matisoo-Smith, E., Allen, M.S., 2003. A novel application of molecular techniques to Pacific archaeofish remains. *Archaeometry* 45, 133–147.
- Nikulina, E.A., Schmölcke, U., 2015. Archaeogenetic evidence for medieval occurrence of Atlantic sturgeon *Acipenser oxyrinchus* in North Sea. *Environ. Archaeol.* <http://dx.doi.org/10.1179/1749631415Y.0000000022>.
- Orlando, L., Calvignac, S., Schnebelen, C., Douady, C.J., Godfrey, L.R., Hänni, C., 2008. DNA from extinct giant lemurs links archaolemurids to extant indriids. *BMC Evol. Biol.* 8, 121.
- Pauly, D., Christensen, V., Dalsgaard, J., Froese, R., Torres, F.J., 1998. Fishing down the marine food webs. *Sci. Mag.* 279.
- Rick, T.C., Kirch, P.V., Erlandson, J.M., Fitzpatrick, S.M., 2013. Archaeology, deep history, and the human transformation of island ecosystems. *Anthropocene* 4, 33–45.
- Roy, R., Dunn, S., Purkis, S., 2009. Mapping Velondriake: the application of bathymetric and marine habitat mapping to support conservation planning, south-west Madagascar. *Blue Ventur. Conserv.* 18.
- Schroeder, H., Ávila-Arcos, M., Malaspina, A., Poznik, G.D., Sandoval-Velasco, M., Carpenter, M.L., Moreno-Mayer, J.V., Sikora, M., Johnson, P.L.F., Allentoft, M.E., Samaniego, J.A., Havis, J.B., Dee, M.W., Stafford Jr., T.W., Salas, A., Orlando, L., Willerslev, E., Bustamante, C.D., Gilbert, M.T.P., 2015. Genome-wide ancestry of 17<sup>th</sup>-century enslaved Africans from the Caribbean. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3669–3673.
- Speller, C.F., Yang, D.Y., Hayden, B., 2005. Ancient DNA investigation of prehistoric almon resource utilization at Keatley Creek, British Columbia, Canada. *J. Archaeol. Sci.* 32, 1378–1389.
- Speller, C.F., Hauser, L., Lepofsky, D., Moore, J., Rodrigues, A.T., Moss, M.L., McKechnie, I., Yang, D.Y., 2012. High potential for using DNA from ancient herring bones to inform modern fisheries management and conservation. *PLoS One.* <http://dx.doi.org/10.1371/journal.pone.0051122>.
- Teletchea, F., 2009. Molecular identification methods of fish species: reassessment and possible applications. *Rev. Fish. Biol. Fish.* 19, 265–293.
- Wheeler, A., Jones, A.K.G., 1989. *Fishes*. Cambridge University Press, Cambridge.
- Willerslev, E., Cooper, A., 2005. Ancient DNA. *Proc. R. Soc. B Biol. Sci.* 272, 3–16.
- Yang, D., Cannon, A., Saunders, S., 2004. DNA species identification of archaeological salmon bone from the Pacific Northwest Coast of North America. *J. Archaeol. Sci.* 31, 619–631.



# Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell



Alicia Grealy<sup>a,\*</sup>, Matthew Phillips<sup>b</sup>, Gifford Miller<sup>a,c</sup>, M. Thomas P. Gilbert<sup>a,d,e</sup>, Jean-Marie Rouillard<sup>f</sup>, David Lambert<sup>g</sup>, Michael Bunce<sup>a</sup>, James Haile<sup>h</sup>

<sup>a</sup>Trace and Environmental DNA (TrEnD) Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia

<sup>b</sup>Vertebrate Evolution Group, School of Earth, Environmental and Biological Sciences, Queensland University of Technology, Brisbane, QLD 4000, Australia

<sup>c</sup>INSTAAR and Geological Sciences, University of Colorado, Boulder, CO 80309, USA

<sup>d</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350, Denmark

<sup>e</sup>NTNU University Museum, N-7491 Trondheim, Norway

<sup>f</sup>Chemical Engineering Department, University of Michigan, Ann Arbor, MI 48109, USA

<sup>g</sup>Environmental Futures Research Institute, School of Environment, Griffith University, Brisbane, QLD 4111, Australia

<sup>h</sup>Research Laboratory for Archaeology and the History of Art, Oxford University, Oxford OX12JD, UK

## ARTICLE INFO

### Article history:

Received 31 August 2016

Revised 20 December 2016

Accepted 7 January 2017

Available online 9 January 2017

### Keywords:

Ancient DNA

Biogeography

Eggshell

High-throughput sequencing

Palaeognath

Phylogeny

## ABSTRACT

Palaeognaths, the sister group of all other living birds (neognaths), were once considered to be vicariant relics from the breakup of the Gondwanan supercontinent. However, recent molecular studies instead argue for dispersal of volant ancestors across marine barriers. Resolving this debate hinges upon accurately reconstructing their evolutionary relationships and dating their divergences, which often relies on phylogenetic information from extinct relatives and nuclear genomes. Mitogenomes from the extinct elephant birds of Madagascar have helped inform the palaeognath phylogeny; however, nuclear information has remained unavailable. Here, we use ancient DNA (aDNA) extracted from fossil eggshell, together with target enrichment and next-generation sequencing techniques, to reconstruct an additional new mitogenome from *Aepyornis* sp. with 33.5X coverage. We also recover the first elephant bird nuclear aDNA, represented by 12,500 bp of exonic information. While we confirm that elephant birds are sister taxa to the kiwi, our data suggests that, like neognaths, palaeognaths underwent an explosive radiation between 69 and 52 Ma—well after the break-up of Gondwana, and more rapidly than previously estimated from mitochondrial data alone. These results further support the idea that ratites primarily diversified immediately following the Cretaceous–Palaeogene mass extinction and convergently evolved flightlessness. Our study reinforces the importance of including information from the nuclear genome of extinct taxa for recovering deep evolutionary relationships. Furthermore, with approximately 3% endogenous aDNA retrieved, avian eggshell can be a valuable substrate for recovering high quality aDNA. We suggest that elephant bird whole genome recovery is ultimately achievable, and will provide future insights into the evolution these birds.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Palaeognaths (including the volant tinamous and the flightless ratites) are distinguished from other birds by the retention of an ‘ancient palate’ (Cracraft, 1974). As such, the Palaeognathae are ubiquitously recognised as the sister group of all other living birds

(neognaths). The common ancestor of palaeognaths split from neognaths during the middle of the Cretaceous period (Clarke et al., 2005), after which neognaths underwent an explosive radiation around the Cretaceous–Palaeogene boundary (Jarvis et al., 2014; Claramunt and Cracraft, 2015; Prum et al., 2015); however, the subsequent diversification of palaeognaths remains poorly understood. Ratites have historically been seen as archetypal relics from the breakup of the Gondwanan supercontinent (Cracraft, 1973), where populations of a flightless ancestor ‘rafted’ with the drifting continents and diverged from one another in isolation to

\* Corresponding author at: Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, WA 6102, Australia.  
E-mail address: [alicia.grealy@uqconnect.edu.au](mailto:alicia.grealy@uqconnect.edu.au) (A. Grealy).

give rise to rheas (*Rhea* spp.) in South America, the ostrich (*Struthio* spp.) in Africa, emus (*Dromaius* spp.) in Australia, cassowaries (*Casuarius* spp.) in Australia and New Guinea, kiwi (*Apteryx* spp.) and moa (*Dinornis* spp., *Anomalopteryx* sp., *Emeus* sp., *Euryapteryx* sp., *Pachyornis* spp., *Megalapteryx* sp.) in New Zealand, and elephant birds (*Aepyornis* spp., *Mullerornis* spp.) in Madagascar. As such, ratites were assumed to form a monophyletic group, sister to tinamous (Cracraft, 1974; Cooper et al., 2001). While aspects of this phylogeny gain some support from morphological studies (Johnston, 2011; Worthy and Scofield, 2012), morphology has been shown to be poor at resolving evolutionary relationships from characters that are functionally correlated (e.g., Springer et al., 2007). Morphological characters such as flightlessness could be interpreted equally as synapomorphies (inherited from a common ancestor) or homoplasies (arising independently via convergent evolution), confounding phylogenetic inference (Hackett et al., 2008; Harshman et al., 2008; Allentoft and Rawlence, 2012; Baker et al., 2014).

In contrast, DNA has been used to resolve the evolutionary relationships of many organisms where morphology has been inconclusive (e.g., Bunce et al., 2003; Krause et al., 2010). Ratite monophyly has been challenged by the inclusion of DNA sequence information into the phylogeny (Harshman et al., 2008), particularly with the inclusion of ancient DNA (aDNA) sequence information from the extinct moa. The discovery that the volant South American tinamous (*Tinamus* spp., *Nothocercus* spp., *Crypturellus* spp., *Rhynchotus* spp., *Nothoprocta* spp., *Nothura* spp., *Taoniscus* spp., *Eudromia* spp., *Tinamotis* spp.) are nested within ratites as the sister taxon of moa, and that ostriches are sister to all other palaeognaths, suggests that flightlessness evolved several times among ratites (Harshman et al., 2008; Phillips et al., 2010; Smith et al., 2013; Haddrath and Baker, 2012; Baker et al., 2014). Similarly, mitochondrial aDNA places Madagascar's extinct giant elephant birds as sister to New Zealand's kiwi (Phillips et al., 2010; Mitchell et al., 2014), which further supports the idea of multiple losses of flight among ratites because the timing of divergences suggests that vicariance is unlikely to explain their biogeographic distribution. Consequently, there is renewed skepticism of the extent to which vicariance played a role in palaeognath evolution because their distribution could alternatively be explained by dispersal via flying. Resolving this debate hinges on accurately reconstructing the evolutionary relationships and timing of divergence within this group. Reconstructions of palaeognath evolutionary history have further been mired by: uncertainty in the timing of the Gondwanan break-up; controversy surrounding the dating and classification of avian fossil specimens (Ksepka et al., 2014); and methods used to construct phylogenies (Ksepka and Phillips, 2015), including choice of characters (mitochondrial DNA vs. nuclear DNA; Mitchell et al., 2014), choice of outgroup (Johnston, 2011), choice of analytical method (Johnston, 2011), choice of molecular clock calibration (Phillips et al., 2010), and taxon sampling (Cooper et al., 2001; Phillips et al., 2010; Mitchell et al., 2014), specifically the underrepresentation of extinct taxa in phylogenetic reconstructions.

Most studies that do include DNA from extinct taxa in phylogenetic analyses rely on mitochondrial aDNA alone as ancient nuclear DNA is typically more challenging to recover. The likelihood of retrieving endogenous aDNA with high coverage from the multi-copy mitochondrial genome is many-fold greater than the likelihood of retrieving endogenous aDNA from a diploid nuclear genome. Nuclear DNA has also been shown to degrade faster than mitochondrial DNA (Allentoft et al., 2012), which may be protected from microbial decay due to its stable circular structure and the double membrane of the surrounding organelle. Nevertheless, nuclear information often proves crucial to accurately recover deep evolutionary relationships (Hackett et al., 2008): mitochondrial

DNA typically evolves over an order of magnitude faster than nuclear DNA, and the resulting saturation tends to erode phylogenetic signal at deeper divergences in the tree and leaves both phylogeny and branch-length inference more susceptible to biases, such as from compositional non-stationarity (Ksepka and Phillips, 2015). In addition, maternally inherited mitochondrial gene histories may not represent whole genome evolution. For example, many groupings based on both genomic sequences and retroposons of neognaths (Jarvis et al., 2014; Prum et al., 2015) differ from earlier mitochondrial DNA-based phylogenetic studies, demonstrating that nuclear DNA is often needed to reconstruct accurate phylogenies. As emphasised by Haddrath and Baker (2012), erroneous phylogenetic topologies will result in misleading interpretations of evolutionary history and biogeography.

While nuclear DNA from moa has been retrieved, nuclear DNA sequences have not been obtained from elephant birds, presumably due to the rarity of skeletal remains, particularly those with well-preserved DNA. In addition, DNA preservation in bone is typically poor in hot, tropical climates (Hoss et al., 1996) and non-endogenous DNA levels in bone are often high compared with other substrates (Gilbert et al., 2007; Ramirez et al., 2009; Schroeder et al., 2015). As such, it remains to be seen whether the relationships among elephant birds and other palaeognaths can be consistently recovered with the inclusion of elephant bird nuclear aDNA as many avian relationships based on mtDNA have been overturned by nuclear DNA analyses (e.g., Hackett et al., 2008). Because Madagascar has been isolated for over 85 Ma (McLoughlin, 2001; Yoder and Nowak, 2006), accurate placement of elephant birds within the palaeognath phylogeny using aDNA has the potential to better inform their evolutionary history as it offers a unique opportunity to test biogeographical hypotheses regarding the roles of vicariance and dispersal in ratite evolution (Phillips et al., 2010; Mitchell et al., 2014); if the lineage leading to elephant birds arose after Madagascar was isolated, a vicariance model of evolution could not explain their presence there.

As an alternative to bone, eggshell has been shown to provide a good source of high-quality, endogenous aDNA from extinct birds, such as moa (Oskam et al., 2010, 2011; Allentoft et al., 2011). This may be because eggshell is water resistant and therefore DNA in eggshell is better protected from microbial decay and hydrolytic damage than the DNA in porous bone (Oskam et al., 2010). To date, there has been some success with recovering aDNA from elephant bird eggshell (Oskam et al., 2010). In addition, elephant bird eggshell is highly robust and abundant; certain beaches are "nearly paved with broken eggshells" (Dewar, 1984), with many buried within dune systems that might offer additional protection of DNA from the environment. Thus, eggshell is the most promising substrate for the retrieval of both mitochondrial and nuclear elephant bird DNA without the destructive sampling of scarce and valuable skeletal specimens. Coupled with recent advances in aDNA methods (Gansauge and Meyer, 2013; Li et al., 2013) and high-throughput sequencing, we demonstrate how aDNA can be extracted from elephant bird eggshell to independently reconstruct the whole mitochondrial genome of *Aepyornis* sp., as well as capture nuclear loci for the first time. In doing so, we aim to confirm the relationship between elephant birds and kiwi and more accurately date the divergences among the palaeognaths, in order to clarify the processes that shaped their evolution.

## 2. Materials and methods

### 2.1. Site description and collection of eggshell samples

Eggshell samples screened for elephant bird aDNA (13; q.v., Sections 2.3 and 2.4) were collected from various coastal sites in

Southern Madagascar (data available upon request). The sample chosen for further genomic analysis, sample M06-M028 (Fig. 1a and b), was excavated from a rapidly deflating coastal dune system on the southern coast of Madagascar at 25.46250°S, 45.70083°E (Talaky). Eggshell was buried beneath 5–10 cm of aeolian sand; however, the long-term depth of burial is estimated to have been over a meter. The tight concentration and lack of any abrasion along the edges of eggshell fragments in this deposit suggest they were continuously buried (no exposure to the surface) since their initial burial.

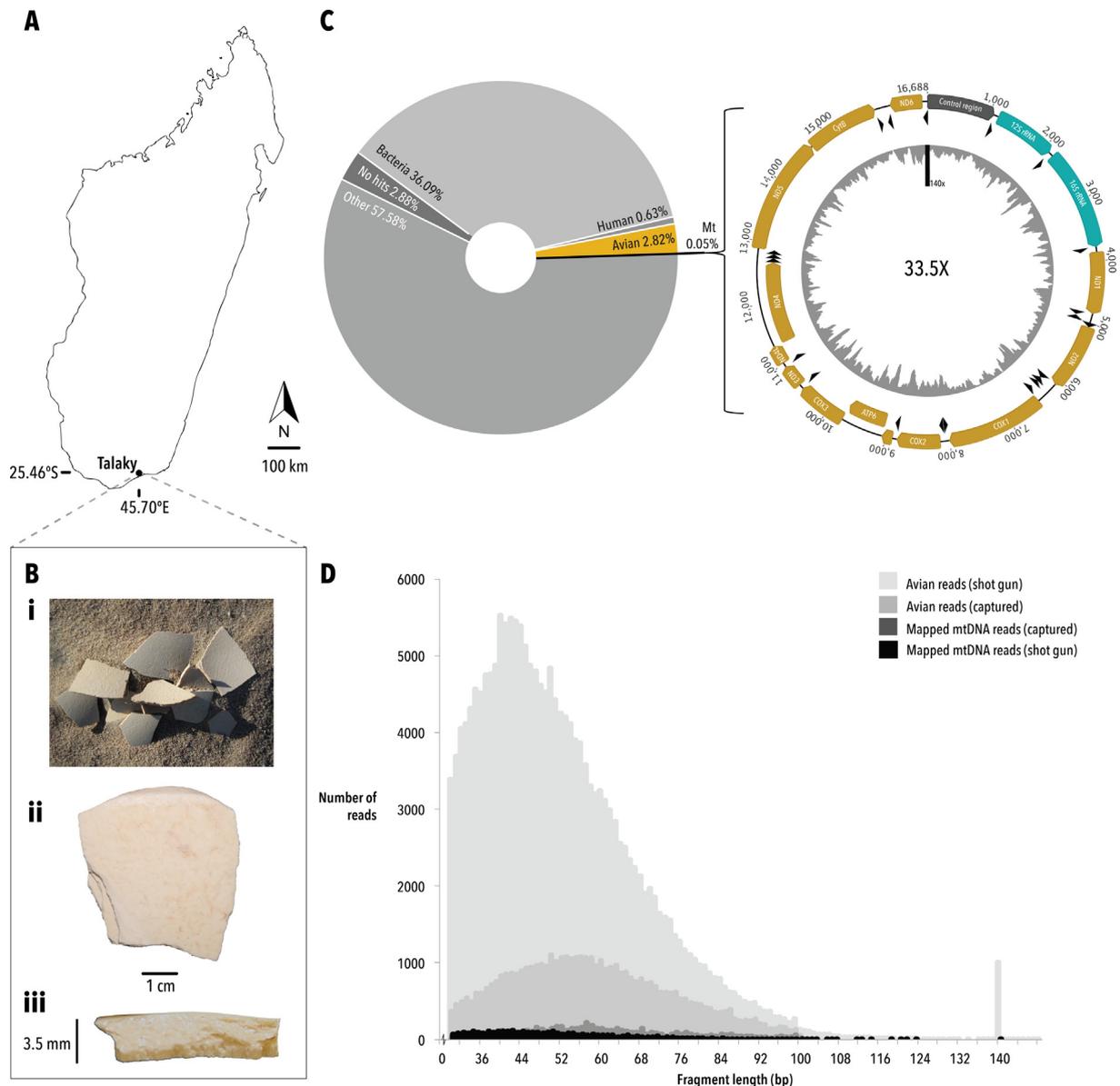
## 2.2. Dating and morphological identification of eggshell

Genus identification of eggshell in the accumulation described above (Section 2.1) was based on the thickness; fragments ranged from 3.6 to 3.9 mm thick, and were identified as *Aepyornis* sp. as there is a clear distinction in the thickness of eggshell between ele-

phant bird genera (Tovondrafae et al., 2014). The age of specimen M06-M028 was estimated to be 1100 years BP, based on  $^{14}\text{C}$  dating (SI 1.1).

## 2.3. aDNA extraction

Genomic DNA was extracted from 200 mg of powdered *Aepyornis* sp. eggshell (13 samples) using a previously optimised protocol with minor changes (Shapiro and Hofreiter, 2012; SI 1.2). Sample preparation and DNA extraction were performed in a designated ultra-clean facility at Murdoch University, WA, Australia using sterile reagents and equipment in keeping with standard aDNA practice (Willerslev and Cooper, 2005). All post-PCR methods were performed in a physically separated laboratory. Extraction and subsequent sequencing of sample M06-M028 was replicated in a separate ultra-clean facility at Curtin University, WA, Australia for authentication; however, a different extraction method was



**Fig. 1.** (A) Map of Madagascar showing the collection site of the eggshell sample; (B) photograph of *Aepyornis* sp. eggshell specimen M06-M028 showing it (i) as collected (not to scale), (ii) as viewed from above, and (iii) in cross-section; (C) annotated schematic of the mitochondrial genome of *Aepyornis* sp.: grey peaks represent the coverage across the genome with the average coverage depicted, and a pie chart displaying the percentage of endogenous (avian) DNA versus exogenous DNA from shotgun sequencing is shown left; (D) fragment length distribution of total avian and mapped mitochondrial reads obtained from both shotgun and capture enrichment.

employed with minor changes (Dabney et al., 2013; SI 1.2). Extraction and PCR negative controls were included and carried through to sequencing.

#### 2.4. Screening samples via amplicon sequencing

Samples were screened for endogenous aDNA by qPCR-amplifying and sequencing in parallel a 53 bp avian metacoding region of the mitochondrial 12S rRNA gene, with unique 5' and 3' indexes flanking the insert (SI 1.3). Sequencing was carried out using the Ion Torrent PGM (Life Technologies) 200 bp kit V2 on a 314 chip following the manufacturer's instructions. The sequences were trimmed, quality filtered, abundance filtered, dereplicated (see Section 2.7; SI 1.4), and were aligned to the NCBI's GenBank nucleotide database (Benson et al., 2006) using BLAST v.2.2.30+ (Altschul et al., 1990, default parameters) to confirm that the DNA was most similar to published reference sequences of *Aepyornis* and free from avian contaminants. Of all the samples that contained amplifiable elephant bird aDNA (seven), the sample exhibiting the lowest relative qPCR cycle-threshold value (and therefore, the sample with the most endogenous DNA, sample M06-M028) was selected for further genomic analyses.

#### 2.5. Shotgun sequencing

Four single-stranded shotgun libraries (two per extract) were prepared independently in ultra-clean facilities at both Murdoch University and Curtin University by following Gansauge and Meyer (2013) with minor changes (SI 1.5). Extraction and negative control libraries were carried through to sequencing in order to assess potential contamination. Libraries were double-indexed and sequenced from one end on both the Ion Torrent PGM (Life Technologies; 200 v.2 kit) and MiSeq (Illumina; 150 v.3 kit) platforms, with the best library further sequenced on the high-throughput sequencing platform, Ion Torrent Proton (Life Technologies; 200 v.2 kit) at Lottery West, Royal Perth Hospital, following the manufacturer's instructions.

#### 2.6. Targeted enrichment through hybridisation capture

A relaxed hybridisation capture was used to enrich the two best libraries for avian mitochondrial and nuclear sequences by following Li et al. (2013) and the MYcroarray MYbaits Sequence Enrichment for Targeted Sequencing kit protocol, with minor changes (SI 1.6). Whole genome baits (MYbaits, MYcroarray) were synthesised by globally transcribing tinamou genomic DNA (*Nothocercus bonapartei*) into RNA baits using biotinylated rUTP (SI 1.6), and 100mer mitochondrial baits with 50 bp tiling were designed using the reference mitochondrial genomes of several avian taxa (Table S4). Enriched libraries were sequenced on the MiSeq (Illumina) platform as above.

#### 2.7. Quality control

Reads were sorted by index and trimmed in Geneious v.7.1.2 (www.geneious.com; Kearsse et al., 2012), allowing no mismatches in the expected sequence of the index or library adapters. FASTQ files were exported and uploaded on to the web-based bioinformatics platform Galaxy (usegalaxy.org; Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010) for quality filtering following Grealy et al. (2016) with minor changes (SI 1.4): sequences were required to have an average quality score of at least Q25, with all bases above Q10, 95% of bases above Q15, and 90% of bases above Q20. Sequences were then combined and dereplicated such that only unique sequences remained. Reads below 30 bp in length were discarded. To conservatively estimate

the proportion of endogenous DNA extracted from *Aepyornis* eggshell relative to bacterial and contaminating DNA, unenriched (shotgun) sequences were assigned taxonomy using BLAST and MEGAN v.4.70.4 (Huson et al., 2007; SI 1.7).

#### 2.8. Mitochondrial genome assembly

Quality filtered, unique sequences were iteratively mapped against a ratite consensus mitochondrial reference genome in Geneious (i.e., a consensus derived from aligning kiwi, emu, cassowary, ostrich, rhea, moa, and tinamou mitochondrial genomes; SI 1.7). Mapped reads were then aligned with BLAST to a custom aggregate database containing both the GenBank nucleotide database and unpublished *Struthio camelus* whole genome contigs (SI 1.7). Taxonomic assignments were examined in MEGAN. Sequences that were assigned to class Aves were extracted and re-mapped onto the consensus mitochondrial genome from the previous round of mapping in order to ensure that no contaminating sequences were mapped. Mapped tRNA-Phe was used as a seed to complete assembly of the control region in MITObim v.1.8 (Hahn et al., 2013), using all unique sequences as the read pool. Bases with coverage >2 and matching at least 50% of sequences were called as the consensus as per Dabney et al. (2013), in order to generate a conservative contiguous reconstruction of the elephant bird mitochondrial genome. aDNA damage patterns were assessed using mapDamage v.2.0 (Jonsson et al., 2013) to authenticate the antiquity of sequences (SI 1.8).

#### 2.9. Identification of nuclear sequences

In order to identify orthologous, phylogenetically informative nuclear loci, unique reads were mapped, in Geneious, to a custom database of ostrich (*S. camelus*) reference sequences containing 44 nuclear protein-coding exons from Baker et al. (2014), 67 nuclear protein-coding exons from Haddrath and Baker (2012), Chojnowski et al. (2007), Hackett et al. (2008), Harshman et al. (2008), and Smith et al. (2013), and 7976 nuclear protein-coding exons from Jarvis et al. (2014) and Zhang et al. (2014) (SI 1.9). These genes are known to be single-copy and have been used in prior avian phylogenetic studies. With relatively short sequence reads and limited coverage, any non-coding data that is substantially variable also tends to provide far less certainty about orthology; rather, single-copy, protein coding genes allow frame-checking, as well as a mix of conserved and variable sites that balances homology considerations with signal considerations. Hits were accepted as putative orthologues only if: (1) the read was 80 bp or longer and aligned across the entire length of the read; (2) the percent similarity for the alignment with ostrich was higher than the similarity between ostrich and chicken for the same locus; and (3) reciprocal comparison of the ostrich genome to elephant bird sequence set identified the same read as its closest match (Vallender, 2009). The minimum coverage accepted per base was one as there are far fewer copies of nuclear loci than mitochondrial loci; this is not unusual for aDNA (Rohland et al., 2010). The ostrich orthologues were then mapped against tinamou (*T. major*/*T. guttatus*; Jarvis et al., 2014), kiwi (*A. apteryx mantelli*; Le Duc et al., 2015), and unpublished emu (*D. novaehollandiae*), and rhea (*R. americana*) whole genome contigs (generated as part of the B10k avian genome sequencing project; Zhang et al., 2015) to find the corresponding loci in these ratites using the same method. These sequences were also reciprocally mapped to the elephant bird data set to further confirm orthology (SI 1.9).

## 2.10. Alignments and partitioning

Each locus was aligned in Geneious with the corresponding locus from 16 other palaeognaths (where available), and 11 neognath outgroups using the default parameters (SI 1.10). Alignments were concatenated and translated in MUSCLE (Edgar, 2004) to find ORFs and refine the alignments. The multiple sequence alignment was then imported into Se-AL v.2.0a (Rambaut, 1996) for manual checking. Based on published partitioning schemes for palaeognaths (e.g., Phillips et al., 2010), concatenated mitochondrial protein-coding genes were partitioned into first, second, and third codon positions, while RNA-coding genes were partitioned into stem and loop sites, resulting in a total of five mitochondrial partitions (m1, m2, m3, stems, loops; SI 1.10). Due to differing taxonomic composition, nuclear data was partitioned by source as well as by codon position, with loci derived from Jarvis et al. (2014) and Zhang et al. (2014) concatenated and separated from loci derived from Haddrath and Baker (2012) and Baker et al. (2014), resulting in a total of six nuclear partitions (n1, n2, n3 and hb1, hb2, hb3, respectively; SI 1.10).

## 2.11. Phylogenetic analyses

Relative composition variability (RCV) and stemminess tests (Phillips and Pratt, 2008) were performed in PAUP v.4.0b (Swofford, 2003) to assess base composition bias and the extent of phylogenetic signal erosion in order to determine which partitions may benefit from RY coding that will alleviate the biases (SI 1.11.1, SI 1.11.2). Mitochondrial third codon positions had both the highest compositional heterogeneity across taxa (Table S13, Fig. S4) and the lowest uncorrected stemminess (greatest phylogenetic signal erosion) and hence, were RY-coded; however, we also considered how an additional two extreme mitochondrial coding schemes impact phylogenetic support by performing analyses when (1) all mitochondrial partitions coded by standard nucleotides, and (2) first and second mitochondrial codon positions, as well as RNA stem and loop partitions, all RY-coded, with mitochondrial third codons excluded (SI 1.11.2).

jModelTest v.2.1.17 (Guindon and Gascuel, 2003; Darriba et al., 2012) was used to test the best-fitting substitution model for each partition (SI 1.11.3). Mitochondrial, nuclear and total evidence phylogenetic trees were constructed using both a maximum likelihood and Bayesian approach for each coding scheme. Phylogeny reconstruction was performed in RAxML v.1.5 (with 500 bootstrap replicates; Stamatakis, 2014; Silvestro and Michalak, 2012) and MrBayes v.3.2.6 (Huelsenbeck and Ronquist, 2001; SI 1.11.4–1.11.5) implemented through the CIPRES online bioinformatics pipeline (Miller et al., 2010). Maximum likelihood analyses were run under a GTR + I +  $\Gamma$  model with 500 bootstrap replicates (SI 1.11.4). Bayesian analyses were run twice after a burn-in of 0.2 for 10 million generations sampling every 2500 generations and employing three chains (SI 1.11.5); m1, m2, stems, loops, and n1 and n2 partitions were run under a GTR + I +  $\Gamma$  model, while RY-coded mitochondrial third codon positions (m3) were run under an CF87 + I +  $\Gamma$  model, with n3 and hb1 run under a GTR +  $\Gamma$  model, and hb2 and hb3 run under a HKY +  $\Gamma$  model (Table S16). Tracer v1.6.1 was used to examine the convergence of Bayesian runs (Rambaut et al., 2003; SI 1.11.6).

## 2.12. Molecular dating

Molecular dating was performed using MCMCTree (Yang and Rannala, 2006) within PAML v.4.8 (Yang, 2007) on the mitochondrial, nuclear, and combined data with both independent and auto-correlated models of rate variation across the respective tree topologies described above (Fig. 2; SI 1.12). In each case, seven

fossil-based age priors were used for calibration: avian root 66.5–124.1 Ma, Galloanserae node >66.5<83.8 Ma, Penguin/tubenose node >60.5<72.3 Ma, Core land/water birds node >60.5<72.3 Ma, non-ostrich palaeognaths node >56.0<72.3 Ma, Parrots/Passeriformes node >53.5 Ma, Emu/kiwi node >24.5 Ma (SI 1.12.1, Table S18). For each analysis the GTR +  $\Gamma_4$  substitution model was employed (and collapsed to effectively 2-state F81 +  $\Gamma_4$  for RY partitions). Default parameters were generally used, although sigma2\_gamma scale (=1) reflects the weak prior of 100 Ma for the avian root (cf., Jarvis et al., 2014; Ksepka and Phillips, 2015), and rgene\_gamma scale (=2) is derived as the inverse of the approximately 1/2 substitutions per 100 Ma averaged across the data (SI 1.12.2).

## 3. Results

### 3.1. Sample screening via amplicon sequencing

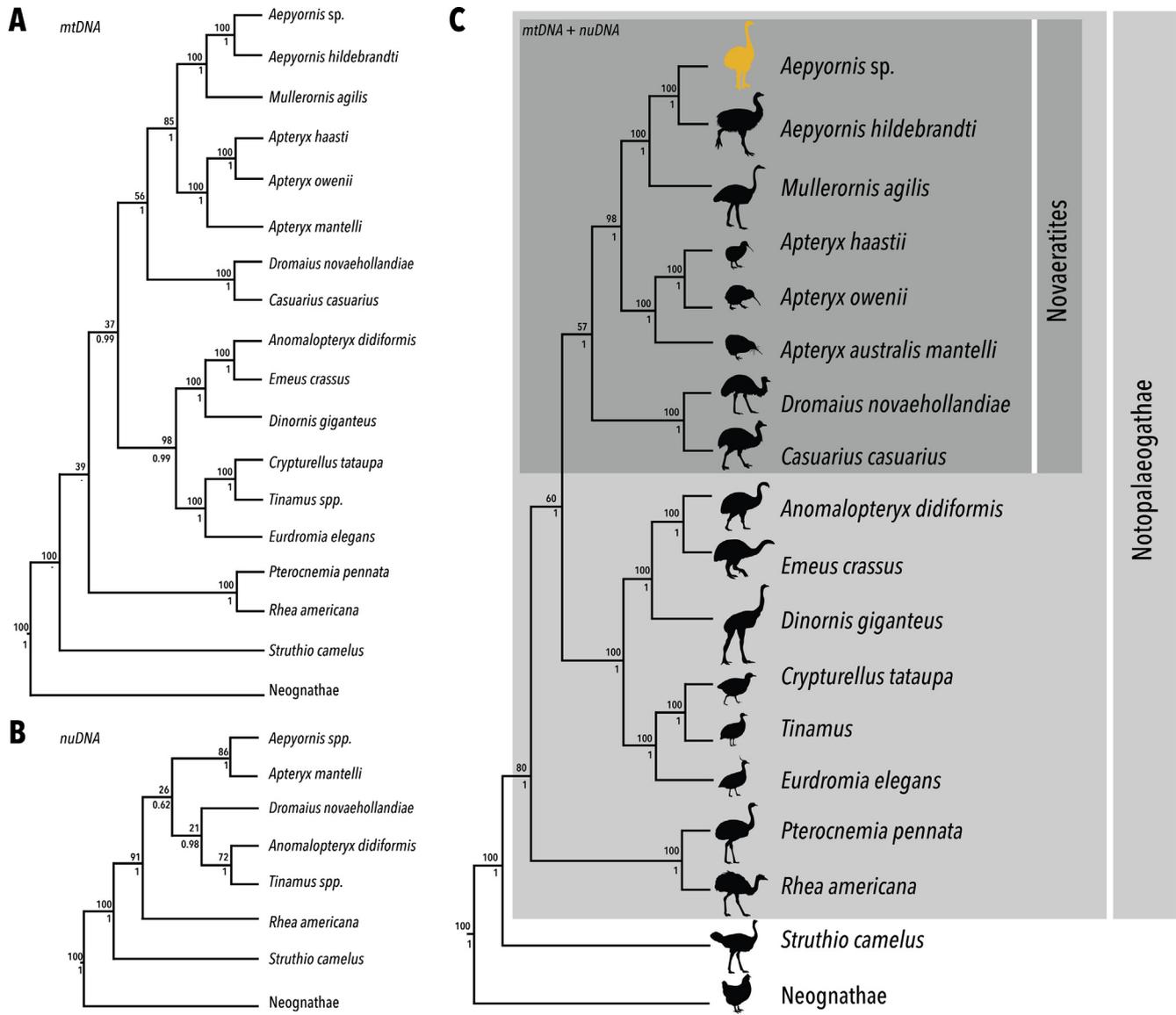
Seven out of 13 eggshells (approximately 54%) screened for endogenous aDNA via sequencing of a 12S rRNA barcoding region yielded elephant bird aDNA. For the most promising sample, M06-M028, 29 unique sequences were generated post quality and abundance filtering, all of which aligned closest with the published *Aepyornis hildebrandti* (KJ749824) and *Mullerornis agilis* (KJ749825) 12S rRNA gene sequences, suggesting that the sample contained endogenous DNA and was largely free of contamination by other birds.

### 3.2. Shotgun sequencing

Shotgun sequencing of aDNA extracted from the eggshell sample above (SI 1.5) yielded a total of 6,628,956 quality filtered, unique sequences above 30 bp, with 97.2% of reads having an average quality above Q20, and 52.5% of reads having an average quality above Q30 (Table S8). 2.82% (187,044; Table S7) of reads were taxonomically assigned to class Aves (Fig. 1c), with about one third assigned to bacterial taxa (36.09%, Fig. 1c), and less than one percent (0.63%, Fig. 1c) assigned to human. The mean length of avian shotgun reads was 50.8 bp (St Dev 17.4 bp; Fig. 1d). Of these avian reads (where over 80% assigned to Palaeognathae; SI 1.7), 1.87% (3505; 0.05% of the total reads) mapped onto a ratite consensus reference mitochondrial genome (Table S7). Mitochondrial mapped reads had a mean length of 49.3 bp (St Dev 15.40 bp), with 97.6% of reads having an average quality over Q20, and 64.3% of reads having an average quality over Q30. 34,706 sequences were obtained from the extraction and negative control libraries, seven of which mapped to the ratite mitochondrial reference genome. However, only one of these sequences had an identical counterpart in the dataset, and this sequence was removed.

### 3.3. Bait enrichment

Together, enrichment of shotgun libraries by hybridisation capture with avian mitochondrial and tinamou genomic baits (SI 1.6) yielded a further 557,074 quality filtered unique sequences above 30 bp with a mean length of 50.7 bp (St Dev 15.6 bp), 99.8% of sequences having an average quality above Q20, and 98.7% of sequences having an average quality above Q30. These sequences were comprised of 8.12% (45,236; Table S7) avian reads, while 1.08% of the total captured reads (6021; Table S7) mapped onto the ratite consensus mitochondrial reference genome, with an average length of 60.2 bp (St Dev 16.3 bp; Fig. 1d). Although fewer reads were obtained by capture compared with direct shotgun sequencing, nearly 4 times as many were target (avian) DNA, with an approximately 20-fold increase in mitochondrial DNA reads.



**Fig. 2.** (A) Mitochondrial (with third codon positions RY-coded; where a node is not supported by Bayesian analyses, an ‘-’ symbol is displayed—note these unsupported nodes are favoured under ML and Bayesian inference with the most conservative coding scheme that has third codon positions excluded and all other sites RY-coded); (B) nuclear; and (C) combined mitochondrial (with third codon positions RY-coded) and nuclear palaeognath phylogenetic trees. Maximum Likelihood bootstrap and Bayesian posterior probability estimates are displayed above each node. Neognath outgroup relationships are not shown for simplicity (see Table S17).

Furthermore, a greater proportion of captured avian sequences were long (i.e., above 60 bp) in comparison to non-captured reads (Fig. 1d).

### 3.4. Mitochondrial genome reconstruction and nuclear mapping

In total (combined shotgun and bait-enriched datasets), 9478 unique avian reads above 30 bp mapped onto a ratite consensus mitochondrial reference genome (SI 1.7) with 641 sequences assembled across the control region (D-loop). This allowed us to independently reconstruct 16,668 bp of the *Aepyornis* sp. mitochondrial genome with an average coverage of 33.5X (Fig. 1c). Regions with lower coverage largely correspond to the hyper-variable control region, although 1077 bp of this region was recovered. This is consistent with the length of the control region in other ratites, which ranges from 1034 bp in ostrich (*Struthio camelus*) to 1362 bp in kiwi (*Apteryx owenii*). The characteristic pattern of a greater proportion of C-to-T misincorporations at both ends of the read that is expected from libraries built from single stranded

aDNA (Schroeder et al., 2015), was observed in the mitochondrial mapped reads and suggests that the DNA is likely to be of ancient origin (SI 1.8, Fig. S2). The influence of aDNA damage and errors on the final consensus sequence is likely to be minimal given the high coverage obtained, as well as the use of both *Life Technologies* and *Illumina* platforms to sequence several independent libraries.

In addition to a complete mitochondrial genome, we retrieved sequence information from 154 protein-coding regions (Table S10) of *Aepyornis* sp. that totaled 12,519 bp of phylogenetically informative nuclear loci with an average coverage of approximately 2.36X (with 52% of bases having a coverage of 1X and the remainder having a coverage of 2X or more), and average quality of Q24.8 (SI 1.9).

### 3.5. Phylogeny reconstruction

Relative composition variability (RCV) and stemminess tests (Phillips and Pratt, 2008) respectively indicated base composition bias and phylogenetic signal erosion in the mitochondrial third

codon positions that could lead to inaccurate phylogeny reconstruction (SI 1.11.1–1.11.2). RY-coding the mitochondrial third codon positions saw RCV fall from 0.1 to 0.04 (Table S11), while stemminess increased from 0.17 to 0.25 (Table S14), suggesting that there is both improved base compositional homogeneity and less substitution saturation when mitochondrial third codon positions are RY-coded. Nevertheless, we additionally performed phylogenetic analyses with no mitochondrial RY-coding (standard coding), as well as full mitochondrial RY-coding with mitochondrial third codon positions excluded entirely (SI 1.11). Regardless of the level of mitochondrial RY-coding, the relationships among the neognath outgroups largely agree with extant avian phylogenies derived from genomic analyses (Jarvis et al., 2014; SI 1.11.7, Table S17); in particular, the outgroup relationships inferred from the combined mitochondrial and nuclear analyses fully agree with those of Jarvis et al. (2014), which adds credibility to the topology of the palaeognath phylogeny that we determined hereafter.

We confirm, with high statistical support, that elephant birds are sister to kiwi (Fig. 2). Mitochondrial and nuclear DNA support this relationship, both independently and together: maximum likelihood bootstrap estimates range between 86% and 100%, while posterior probabilities are all 1, regardless of whether the mitochondrial third codon positions were RY-coded or not (Fig. 2; Table S17). However, support decreases as more mitochondrial positions are RY-coded, most likely due to loss of phylogenetic signal: when all mitochondrial positions are RY-coded with third codon positions excluded, bootstrap support decreases to 51% (Table S17) and posterior probability decreases to 0.93. It should be noted that this is not the case in the combined dataset; the addition of nuclear information restores support for this grouping, even with the highest level of mitochondrial RY-coding (i.e., full mitochondrial RY-coding with third codon positions excluded entirely; Table S17).

While most analyses placed ostrich as sister to all other palaeognaths with high support (Fig. 2; Table S17), the monophyly of the “Notopalaeognathae”—all palaeognaths to the exclusion of ostrich—is best supported by the nuclear data (Fig. 2b), with a combination of both nuclear and mitochondrial DNA resulting in 80% bootstrap support for this node (Fig. 2c). Although Notopalaeognathae monophyly was unsupported by standard nucleotide mitochondrial data, fully RY-coding the mitochondrial genome increased bootstrap support for this grouping, up to 98% in combination with the nuclear data (Table S17). Bayesian analyses also show high support for this topology, but only when nuclear data is included (adding nuclear data increases support for every mitochondrial RY-coded grouping that was not already at 100%; Fig. 2), or with moderate to high levels of mitochondrial RY-coding (Table S17). In fact, mitochondrial incongruence with the nuclear phylogeny is largely eliminated by RY-coding; ‘approximately unbiased’ (AU) testing on the nuclear data finds significant incongruence with the standard nucleotide mitochondrial topology ( $p = 0.034$ ), but not with the RY-coded mitochondrial topology ( $p = 0.589$ , S1.11.8).

The branching order of the emu/cassowary, rhea, and moa/tinamou clades within Notopalaeognathae are less robust. The clade grouping moa/tinamou with the “Novaeratitae” to the exclusion of the rheas does not garner high bootstrap support from either mitochondrial or nuclear analyses (Fig. 2a and b); rather, this relationship gains the most support from a combination of both nuclear and mitochondrial DNA (Fig. 2c). Similarly, bootstrap support for novaeratite monophyly (i.e., Casuariiformes sister to the elephant birds/kiwi) to the exclusion of moa/tinamou is highest when both mitochondrial and nuclear data are combined (Fig. 2c). Both bootstrap support and posterior probability estimates for this grouping tend to decrease as more mitochondrial positions are RY-coded (Table S17). Although there is no support for this topology from nuclear data alone, there is extremely little bootstrap support for the alternative grouping of emus sister to moa/tinamou (21%; Fig. 2b). Although some clades in the nuclear tree are poorly resolved, there is good complementarity between the mitochondrial and nuclear datasets, as shown by the results of the AU test (above).

### 3.6. Molecular dating

In general, divergence dates estimated from alternative analyses (SI 1.12.3, Table S19) are in broad agreement with one another, suggesting that the node ages determined below remain largely unaffected by alternative analyses.

With mitochondrial third codon positions RY-coded, divergence estimates for all palaeognath orders span 16.3 Ma, between about 69 Ma and 52.7 Ma (Table 1; Fig. S5). Without RY-coding, divergence dates differ from these estimates by less than 1 Ma (Table S19). In contrast, stem divergences of all palaeognath orders (except Dinornithiformes and Tinamiformes) estimated by nuclear data are older than the mitochondrial estimates (occurring between 72.1 Ma and 48.8 Ma; Table 1, Fig. S5), and span 23.3 Ma–7 Ma slower than mitochondrial estimates. However, the nuclear data indicates that divergence from the base of Notopalaeognathae to the elephant bird/kiwi clade spans only 5.7 Ma, as opposed to 11.5 Ma estimated from the mitochondrial data—5.8 Ma faster. The divergence between elephant birds and kiwi occurs approximately 59.2 Ma based on the nuclear data, and approximately 53.5 Ma based on the mitochondrial data (Table 1). In contrast to the mitochondrial estimate, the nuclear estimate for the elephant bird/kiwi split is closer to the K-Pg boundary, occurring approximately 10.4 Ma before the divergence between moa and tinamou (rather than contemporaneously), and comes a mere 0.8 Ma after the Casuariiformes split from their common ancestor (as opposed to 4.5 Ma; Table 1).

Nevertheless, the dating estimates for the mitochondrial and nuclear phylogenies are consistent; in fact, the combined analysis estimates divergence between elephant birds and kiwi to be between these, approximately 54.2 Ma, and total divergence time of all palaeognath orders is approximately 17.3 Ma, between 69.2 and 51.9 Ma (Table 1). The divergence between the two genera of

**Table 1**

Divergence dates and their uncertainty for palaeognath orders estimated through molecular clock dating using mitochondrial data (\* with third codon positions RY-coded), nuclear data, and combined mitochondrial and nuclear data. Nodes labeled A–E correspond to those in Fig. 3.

Node	Divergence time (Ma) ± 95% HPD		
	Mitochondrial*	Nuclear	Mitochondrial* + Nuclear
A. Palaeognathae	69.0 (77.2–60.5)	72.1 (82.1–62.4)	69.2 (78.0–61.3)
B. Notopalaeognathae	65.0 (72.3–57.2)	64.9 (72.3–56.2)	64.2 (72.2–56.8)
C. Moa/tinamou/Novaeratites	62.6 (70.0–55.1)	62.6 (70.2–53.8)	62.1 (70.0–55.0)
D. Casuariidae/elephant bird/kiwi OR Casuariidae/moa/tinamou	58.1 (65.7–51.1)	60.0 (67.8–50.9)	58.3 (66.0–51.3)
E. Elephant bird/Kiwi	53.5 (61.0–46.6)	59.2 (66.6–49.8)	54.2 (61.7–47.5)
Moa/tinamou	52.7 (60.1–45.7)	48.8 (58.5–39.1)	51.9 (59.4–45.5)

elephant birds (*Aepyornis* and *Mullerornis*) is estimated to have occurred approximately 27.6 Ma (95% HPD 33.9–21.9 Ma).

## 4. Discussion

### 4.1. Mode and tempo of palaeognath evolution

We present the first palaeognath phylogeny based on nuclear and mitochondrial DNA data for all recent orders, and our analyses differ from other recent studies in several ways, which we detail below. Hence, nuclear and mitochondrial aDNA from the extinct Madagascan elephant bird, *Aepyornis* sp. has important implications for understanding palaeognath evolution.

Both nuclear and mitochondrial DNA unequivocally place elephant birds sister to kiwi, confirming the mtDNA findings of Mitchell et al. (2014); this is even despite the low coverage of the nuclear loci. Although the palaeognath root is ubiquitously placed between the ostrich and notopalaeognaths, its position is stabilised by the addition of nuclear DNA. Similarly, relative to the nuclear data alone, including mitochondrial DNA substantially improved the phylogenetic resolution (i.e., bootstrap support and Bayesian posterior probability), and brings the topology in agreement with most other recent studies (e.g., Mitchell et al., 2014).

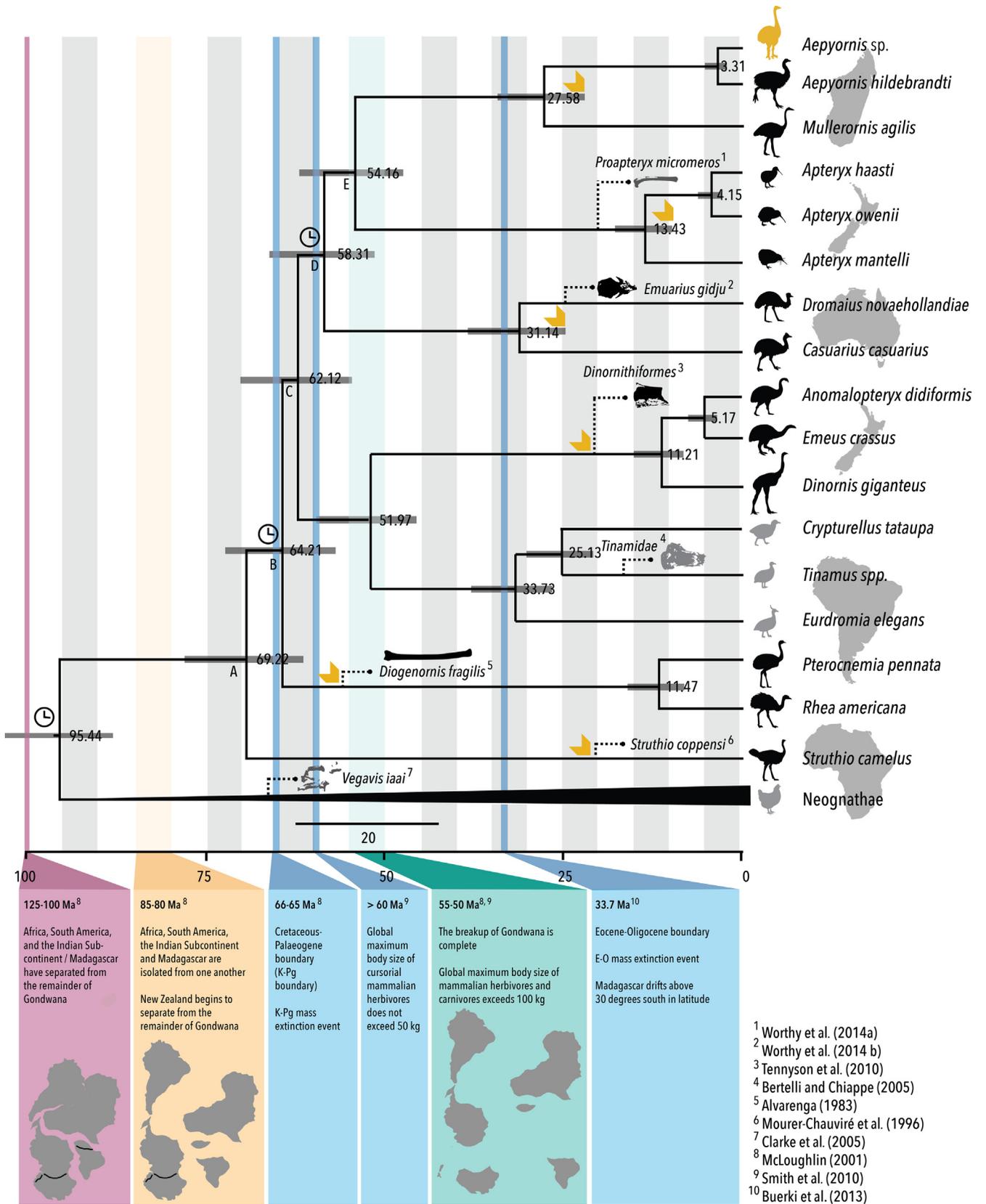
However, the branching order of major clades within Notopalaeognathae (moa/tinamous, rheas, and emus/cassowaries) is less certain. In no analysis are moa/tinamous placed basal among Notopalaeognathae as suggested by Haddrath and Baker (2012) and Baker et al. (2014) (Table S17). While the position of the rheas as deepest among the notopalaeognaths is in agreement with Mitchell et al. (2014) and Phillips et al. (2010), their placement has very low statistical support that is only restored by the inclusion of nuclear DNA. Low statistical support and short internal branches at the base of Notopalaeognathae (Table S17) indicate that its constituent clades may have diverged near-simultaneously. The apparent polytomy may potentially be better resolved with the addition of data from more nuclear loci, particularly for taxa with little nuclear data (e.g., cassowary; though it is worth noting that our nuclear topology is similar to that of Prum et al. (2015) whose tree was based upon a much larger genomic data set, albeit lacking the extinct ratites). However, even with many thousands of DNA base pairs of across hundreds of loci, statistical support for these internal branches remains poor (Harshman et al., 2008; Smith et al., 2013; Baker et al., 2014). In order to achieve fine scale resolution of these nodes, DNA from retroelements and faster-evolving sequences such as introns may be beneficial. Retroelements are generally not homoplastic and analysis of insertion events tends to be immune to certain biases (Haddrath and Baker, 2012) that potentially arise when using protein-coding sequence (Jarvis et al., 2014), such as GC content among taxa in third codon positions (although here, GC content in nuclear third codon positions is unlikely to generate bias, being similar among palaeognaths; Table S12, Fig. S3). Loci that have a faster rate of evolution potentially contain greater phylogenetic signal and have better resolved the relationships among neognaths (Jarvis et al., 2014). However, more data is not guaranteed to robustly resolve all palaeognath relationships; some neognath polytomies persist despite the inclusion of data from over 10,000 orthologous nuclear sequences (Jarvis et al., 2014). Resolving the basal divergence within Notopalaeognathae may also be confounded by incomplete lineage sorting; accordingly, multispecies coalescent models may benefit phylogenetic reconstructions. However, the short, low information aDNA sequences obtained here make implementing such approaches inappropriate (Mirarab et al., 2014; Tonini et al., 2015). Nevertheless, the elephant bird/kiwi grouping would be far less likely conflated with deep coales-

cence issues, given the high support values and a longer inferred stem lineage.

The combined analysis is currently our best estimate for the tempo of the palaeognath radiation, being based on our best-resolved phylogeny and most precise divergence estimates (i.e., smallest Bayesian 95% HPDs on average). This mirrors the findings of Ksepka and Phillips (2015) that nuclear data and RY-coded mitochondrial genomes provide closely comparable divergence estimates for birds, and together provide a more precise estimate of divergence time than either alone, which advocates for the combined approach taken here. The total divergence time among palaeognaths estimated from the combined mitochondrial and nuclear data was about 6 Ma more rapid than previously suggested by Mitchell et al. (2014), who estimated the divergence time to occur across approximately 23 Ma based on mitochondrial DNA only. Similarly, divergence across the notopalaeognath polytomy spans 5.9 Ma–4.3 Ma faster than estimated by Mitchell et al. (2014) and occurring closer to the K-Pg boundary. This is consistent with the rapid radiation of notopalaeognaths implied by the lack of phylogenetic resolution at the base of this group. Furthermore, our data date the first three continental divergences within Notopalaeognathae (rheas from the remaining notopalaeognaths, moa/tinamous from novaeratites, and Casuariiformes from elephant birds/kiwi) to the Paleocene, coincident with the larger Neoaves radiation (Jarvis et al., 2014; Prum et al., 2015). In comparison, this diversification was inferred to take twice as long by mtDNA alone (Phillips et al., 2010; Mitchell et al., 2014), or to have occurred approximately 20 Ma earlier (Baker et al., 2014; Haddrath and Baker, 2012) or 20 Ma later (Prum et al., 2015) with nuclear genomic data. The stem divergences of several palaeognath orders also occur after 66 Ma: the stem lineage of Tinamiformes, Dinornithiformes, Casuariiformes, Aepyornithiformes, and Apterygiformes all have origins in the Late Palaeocene–Early Eocene. This strengthens the argument that the Cretaceous–Palaeogene extinction event was the catalyst for a rapid radiation of palaeognaths, as well as for the neognaths, due to ‘ecological opportunity’, where an abundance of resources (Myers, 2012) that were available to surviving organisms in the wake of the K-Pg mass extinction triggered a selective shift that led to adaptive divergence.

The divergence time we estimate between elephant birds and kiwi from the combined nuclear and mitochondrial data (Table 1; Fig. 3) is not significantly different from that estimated by Mitchell et al. (2014) (50 Ma, 95% HPD 61.5–40.1 Ma); however, our estimate is more precise (i.e., Bayesian 95% HPDs are smaller), and occurs closer to the K-Pg boundary, further supporting the idea of a rapid radiation of palaeognaths. In addition, we confirm that the divergence between elephant birds and kiwi comes well after Madagascar and New Zealand were isolated (ca. 30 Ma later), suggesting that vicariance cannot explain the current distribution of ratites. Although our results reject vicariance to explain the distribution of elephant birds, dispersal may have occurred by a volant common ancestor that evolved flightlessness multiple times (Mitchell et al., 2014; Haddrath and Baker, 2012; Phillips et al., 2010), or by ‘island hopping’ via land bridges that may have existed at this time. While the latter hypothesis is unlikely (Ali and Krause, 2011), new fossil evidence would be required to definitively distinguish between these scenarios. The idea of dispersal by a volant common ancestor is consistent with the finding of a potentially volant “proto” kiwi from the early Miocene of New Zealand (Worthy et al., 2013), as well as Eocene fossils of other volant palaeognaths (e.g., *Lithornis*) found in Europe and the Northern Hemisphere (Houde, 1986; Mayr, 2008). If this is the case, flightlessness evolved at least 6 times among palaeognaths (Fig. 3).

Thus, our results suggest that palaeognaths rapidly diversified but converged upon flightlessness independently in response to similar environmental pressures; for example, the sudden, simul-



**Fig. 3.** Combined mitochondrial (with third codon positions RY-coded) and nuclear molecular clock dating tree with palaeognath divergence times shown next to each node with time (Ma) along the x-axis. Grey bars represent 95% highest posterior density intervals. The timings of major geological events are plotted along the time line. Volant taxa are represented by a grey silhouette as opposed to flightless taxa in black. Fossils used to calibrate the molecular clock are designated by an adjacent clock symbol (n.b., some among neognath calibrations are not depicted). Yellow arrows indicate minimum possible estimates for the evolution of flightlessness along each major palaeognath lineage. The continental distribution of each taxa is depicted by the adjacent transparency (n.b., the distribution for neognath outgroups is global). Nodes labeled A–E refer to those named in Table 1.

taneous generation of novel ecospace across multiple landmasses that would have occurred with the extinction of non-avian dinosaurs at the K-Pg boundary 66 Ma. This mass extinction eliminated all known large terrestrial predators, and new vegetative habitats would have been readily exploitable by palaeognath ancestors with little competition from mammals. Our data is consistent with this hypothesis as all major stem notopalaeognath clades (rheas, moa/tinamous, and novaeratites) diverged before 55 Ma; only after this time did the maximum body size for cursorial mammalian herbivores reach a comparable size, and mammalian carnivores remained under 100 kg until ca. 50 Ma (Smith et al., 2010; Saarinen et al., 2014). This scenario may be analogous to the repeated evolution of flightlessness among Gruiformes, such as rails, that have also undergone multiple rapid losses-of-flight in the recent past in response to island habitats lacking mammalian predators and competitors. Thus, the derivation of flightlessness in ratites appears to be another exemplary case of parallel evolution (Mitchell et al., 2014; Harshman et al., 2008).

Finally, the diversification of elephant bird genera is estimated to be about 10 Ma earlier than previous estimates (Mitchell et al., 2014). The divergence between *Aepyornis* and *Mullerornis* seems to coincide with the Eocene-Oligocene (E-O) boundary around 33.9 Ma—a time characterized by global cooling, widespread extinction, and floral turnover (also known as the ‘Grande Coupure biotic turnover’). At this time, Madagascar drifted above 30°S in latitude into the warm and humid subtropics (Yoder and Yang, 2004), establishing its major biomes as well as many of Madagascar’s endemic flora (Buerki et al., 2013). A change in plant communities brought about by this environmental upset is hypothesized to have prompted the diversification of lemurs on Madagascar (Yoder and Yang, 2004), and similarly, may have triggered the divergence between elephant bird genera. The E-O boundary also may have punctuated speciation within other notopalaeognath lineages, as this time seems to also correspond with a diversification within the Casuariiformes (i.e., divergence between emu and cassowary) and Tinamiformes. However, testing this hypothesis would require more precise dating of these crown divergences. Shallow divergences that are poorly informed by mainly deeper calibrations ages tend to be more susceptible to error and different coding strategies than deep-divergences (Ksepka et al., 2014).

#### 4.2. Eggshell palaeogenomics and future directions

DNA preservation in fossils from tropical-temperate environments like Madagascar will always be challenging. For the first time, we explore the viability of eggshell as a substrate for the retrieval of whole genomes. In comparison to the partial mitogenome published by Mitchell et al. (2014), we retrieved a mitochondrial genome that is about 2271 bp more complete with roughly twice the average coverage and less than 100 ambiguous sites. With almost 3% avian DNA recovered from shotgun data alone, a higher percentage of endogenous DNA was recovered from eggshell than has been recovered from bone, which typically results in less than 1% endogenous DNA for other taxa (Shapiro and Hofreiter, 2014). The mapped M06–M028 eggshell reads are likely conservative estimates due to requirements to map to a distant relative. Furthermore, capture could be optimised by enrichment with genomic baits from a closer relative (as opposed to tinamou whole-genome baits). Considering this, it is not beyond the realm of possibility to obtain a full genome of *Aepyornis*: based on reads mapping to aves (0.0012 Gb from 0.4 Gb) and assuming the size of the elephant bird genome is about the same as the kiwi (1.6 Gb; Le Duc et al., 2015), 1300 times the amount of sequence data generated here would be needed to obtain a draft 1X genome if the library does not sequence to saturation (estimated to be the equivalent of two full HiSeq runs).

Furthermore, eggshell offers probably the only opportunity to study the genetic diversity within and between species of elephant birds, and examine how this relates to their geographic distribution within Madagascar, because it is readily available in many places where elephant birds once nested, while fossil bone is rare. In addition, over half the eggshell samples tested yielded amplifiable elephant bird aDNA, which is a high proportion for ancient samples, particularly in hot environments; these samples, particularly those that were found *in situ*, may also yield genomic information in the future. Whole mitochondrial genomes of elephant birds from across their former range will allow us to address many unsolved questions about elephant bird evolution. For example, we could potentially predict how many ‘molecular species’ of elephant bird actually existed given the paucity of fossil bone. Considering that the biodiversity of other ratites (notably moa) has been grossly overestimated based on morphology, it is entirely possible that fewer species of elephant birds existed than have been named (approximately 8; Brodkorb, 1963; Hume and Walters, 2012). Alternatively, more species or genera of elephant birds may exist than have been described. In fact, studies on other organisms, such as hominids, have described new species based solely on aDNA (e.g., Krause et al., 2010). The integration of eggshell aDNA, 14C dating and eggshell stable isotope data provides a means by which to study species boundaries and extinction timelines.

The ability we now have with fossil eggshell to characterise parts of the elephant bird nuclear genome will not only improve our ability to further resolve tree topologies, but also will provide insights into elephant bird evolution that are not possible to glean from maternally inherited mitochondrial DNA alone: that is, nuclear DNA (including non-coding DNA) will allow us to examine the forces driving evolution, such natural selection and genetic drift by examining microsatellites (as has been done in moa; Allentoft et al., 2011) and genes that show signatures of selective sweeps or population bottlenecks. A better understanding of the population history of elephant birds will be important for determining the factors responsible for their speciation and extinction. There is also the potential to uncover functional mutations, including those responsible for island gigantism, flightlessness, and egg development, which may give insight into behavior and life-history strategies. For example, a genome-wide study of the kiwi found mutations in opsin genes that could be responsible for nocturnal adaptations (Le Duc et al., 2015), and other studies have uncovered similar adaptive mutations in extinct organisms using aDNA (Rompler et al., 2006; Lalueza-Fox et al., 2007, 2008, 2009). Unique mutations among ratites in flightlessness genes may lend further support for the independent evolution of flightlessness from a volant common ancestor.

## 5. Conclusion

The evolutionary history of palaeognaths has historically been contentious but is finally reaching a consensus as our results affirm those of recent studies. We provide independent evidence that elephant birds are the closest relatives of the kiwi, and have refined the dating of the divergences among palaeognaths, through the addition of nuclear aDNA that is essential for clarifying deep evolutionary relationships. This study has also shown that eggshell can be a valuable source of high-quality aDNA and is an excellent substrate for studying the palaeogenomics of extinct organisms that lived in climates that are suboptimal for the preservation of aDNA. For elephant birds, aDNA from eggshell provides an avenue for future research into their phylogeography and functional genomics that will further shed light on the evolution of these understudied birds.

## Author declarations

The authors declare no competing or financial interests.

## Author information

The mitochondrial genome sequence for the studied *Aepyornis* sp. specimen has been deposited within GenBank (available ncbi.nlm.nih.gov) under the accession code KY412176. Data, including nuclear sequence data, can be accessed through DataDryad at doi:10.5061/dryad.6h3q7. Correspondence and requests for materials should be address to AG (alicia.grealy@uqconnect.edu.au).

## Author contributions

JH and MB conceived and supervised the study. JH and AG designed the experiments. AG performed the experiments. AG and MP analysed the data. J-MR synthesised and supplied tinamou nuclear MYbaits. GM supplied and dated eggshell samples. MTPG and DL provided databases of *D. novaehollandiae*, *R. americana* and *A. mantelli* nuclear sequences. All authors contributed to writing and editing the manuscript.

## Note added in proof

In the time since the work described in this paper was accepted, Yonezawa et al. (2016) published a study in Current Biology (doi:10.1016/j.cub.2016.10.029) reporting nuclear DNA recovery from elephant bird fossil bone. This study recovers the same phylogenetic topology but differs in some of the inferred dates within the palaeognath radiation.

## Acknowledgements

This work was funded by a grant from the Australian Research Council (ARC) grants awarded to J. H. (DE120100107). MB was supported in this research by an ARC future fellowship (FT0991741). MTPG was supported by ERC Consolidator Grant (681396-Extinction Genomics). GF wishes to acknowledge the National Geography Society, as well as the Easterbrook Distinguished Scientist Award (from the Division of Quaternary Geology and Geomorphology Division of the Geological Society of America), for funding sample collecting trips in Madagascar. Richard Allcock at Lottery West performed high-throughput sequencing at Royal Perth Hospital, WA. Computing resources at the University of Copenhagen and Pawsey Supercomputer Centre (WA) were used to perform BLAST searches. We thank Guojie Zhang and Erich Jarvis for access to avian genome data, and Leon Huynen for constructive comments on the manuscript. The authors would also like to acknowledge the contribution of Jon Fjelds  and George Pacheco who provided tinamou genomic DNA used for the construction of capture baits.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ympev.2017.01.005>.

## References

Ali, J.R., Krause, D.W., 2011. Late Cretaceous bioconnections between Indo-Madagascar and Antarctica: refutation of the Gunnerus Ridge causeway hypothesis. *J. Biogeogr.* 38, 1855–1872.

Allentoft, M.E., Collins, M., Harker, D., Haile, J., Oskam, C.L., Hale, M.L., Campos, P.F., Samaneigo, J.A., Gilbert, M.T.P., Willerslev, E., Zhang, G., Scofield, P., Hodaway, R.

N., Bunce, M., 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. Roy. Soc. B – Biol. Sci.*

Allentoft, M.E., Oskam, C., Houston, J., et al., 2011. Profiling the dead: generating microsatellite data from fossil bones of extinct megafauna—protocols, problems, and prospects. *PLoS ONE* 6. doi: e1667010.1371/journal.pone.0016670.

Allentoft, M.E., Rawlence, N.J., 2012. Moa's Ark or volant ghosts of Gondwana? Insights from nineteen years of ancient DNA research on the extinct moa (Aves: Dinornithiformes) of New Zealand. *Ann. Anat. – Anat. Anz.* 194, 36–51. <http://dx.doi.org/10.1016/j.aanat.2011.04.002>.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. <http://dx.doi.org/10.1006/jmbi.1990.9999>.

Baker, A.J., Haddrath, O., McPherson, J.D., Cloutier, A., 2014. Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. *Mol. Biol. Evol.* 31, 1686–1696. <http://dx.doi.org/10.1093/molbev/msu153>.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2006. GenBank. *Nucleic Acids Res.* 34, D16–D20. <http://dx.doi.org/10.1093/nar/gkj157>.

Blankenberg, D.V.K., Coraro, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor, J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. In: al. FMAe (Ed.), *Current Protocols in Molecular Biology*. Unit 19.10.11–21.

Brodtkorb, P., 1963. Catalogue of fossil birds. *Bull. Florida State Mus. Biol. Sci.* 7, 206–207.

Buerki, S., Devey, D.S., Callmander, M.W., Phillipson, P.B., Forest, F., 2013. Spatio-temporal history of the endemic genera of Madagascar. *Bot. J. Linn. Soc.* 171, 304–329.

Bunce, M., Worthy, T.H., Ford, T., Hoppitt, W., Willerslev, E., Drummond, A., Cooper, A., 2003. Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* 425, 172–175.

Chojnowski, J.L., Franklin, J., Katsu, Y., Iguchi, T., Guillelte, L.J., Kimball, R.T., Braun, E. L., 2007. Patterns of vertebrate isochore evolution revealed by comparison of expressed mammalian, avian, and crocodylian genes. *J. Mol. Evol.* 65, 259–266. <http://dx.doi.org/10.1007/S00239-007-9003-2>.

Claramunt, S., Cracraft, J., 2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. *Sci. Adv.* 1, e1501005.

Clarke, J.A., Tambussi, C.P., Noriega, J.I., Erickson, G.M., Ketchum, R.A., 2005. Definitive fossil evidence for the extant avian radiation in the Cretaceous. *Nature* 433, 305–308. <http://dx.doi.org/10.1038/nature03150>.

Cooper, A., Lalueza-Fox, C., Anderson, S., Rambaut, A., Austin, J., Ward, R., 2001. Complete mitochondrial genome sequences of two extinct moa clarify ratite evolution. *Nature* 409, 704–707. <http://dx.doi.org/10.1038/35055536>.

Cracraft, J., 1973. Continental drift, palaeoclimatology, and evolution and biogeography of birds. *J. Zool.* 169, 455–545.

Cracraft, J., 1974. Phylogeny and evolution of ratite birds. *Ibis* 116, 494–521. <http://dx.doi.org/10.1111/j.1474-919X.1974.tb07648.x>.

Dabney, J., Knapp, M., Glocke, I., et al., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* 110, 15758–15763. <http://dx.doi.org/10.1073/pnas.1314445110>.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772.

Dewar, R., 1984. Extinctions in Madagascar: the loss of the subfossil fauna. In: *Quaternary Extinctions: A Prehistoric Revolution*. University of Arizona Press, Tucson.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Gansauge, M.T., Meyer, M., 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* 8, 737–748.

Giardine, B.R., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A., 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15, 1451–1455.

Gilbert, M.T.P., Tomsho, L.P., Rendulic, S., Packard, M., Drautz, D.I., Sher, A., Tikhonov, A., Dal n, L., Kuznetsova, T., Kosintsev, P., Campos, P.F., Higham, T., Collins, M.J., Wilson, A.S., Shidlovskiy, F., Buigues, B., Ericson, P.G., Germonpr , M., G therstr m, A., Lacumin, P., Nikolaev, V., Nowak-Kemp, M., Willerslev, E., Knight, J.R., Irzyk, G.P., Perbost, C.S., Fredrikson, K.M., Harkins, T.T., Sheridan, S., Miller, W., Schuster, S.C., 2007. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* 317, 1927–1930.

Goecks, J., Nekrutenko, A., Taylor, J., Team, The.Galaxy., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biol.* 11, R86.

Grealy, A., Macken, A., Allentoft, M.E., Rawlence, N.J., Reed, E., Bunce, M., 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *J. Quat. Sci.* 9999, 1–13.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* 52, 696–704.

Hackett, S.J., Kimball, R.T., Reddy, S., et al., 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320, 1763–1768. <http://dx.doi.org/10.1126/science.1157704>.

Haddrath, O., Baker, A.J., 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proc. Roy. Soc. B – Biol. Sci.* 279, 4617–4625. <http://dx.doi.org/10.1098/rspb.2012.1630>.

- Hahn, C., Bachmann, L., Chevreaux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Res.* 41, 9. <http://dx.doi.org/10.1093/nar/gkt371>.
- Harshman, J.B., Edward, L., Braun, M.J., Huddleston, C.J., Bowie, R.C.K., Chojnowski, J. L., Hackett, S.J., Han, K., Kimball, R.T., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy, S., Sheldon, F.H., Steadman, D.W., Steppan, S.J., Witt, C.C., Yuri, T., 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl. Acad. Sci.* 105. <http://dx.doi.org/10.1073/pnas.0803242105>. 13462–12467.
- Hoss, M., Jaruga, P., Zastawny, T.H., Dizdaroğlu, M., Paabo, S., 1996. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 24, 1304–1307. <http://dx.doi.org/10.1093/nar/24.7.1304>.
- Houde, P., 1986. Ostrich ancestors found in the northern-hemisphere suggest new hypothesis of ratite origins. *Nature* 324, 563–565. <http://dx.doi.org/10.1038/324563a0>.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. <http://dx.doi.org/10.1093/bioinformatics/17.8.754>.
- Hume, J.P., Walters, M., 2012. *Extinct Birds*. Bloomsbury Publishing.
- Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome Res.* 17, 377–386. <http://dx.doi.org/10.1101/gr.5969107>.
- Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., et al., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 345, 1320–1331.
- Johnston, P., 2011. New morphological evidence supports congruent phylogenies and Gondwana vicariance for palaeognathous birds. *Zool. J. Linn. Soc.* 163, 959–982. <http://dx.doi.org/10.1111/j.1096-3642.2011.00730.x>.
- Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684. <http://dx.doi.org/10.1093/bioinformatics/btt193>.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, P., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28.
- Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., Pääbo, S., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* 464, 894–897. <http://dx.doi.org/10.1038/nature08976>.
- Ksepka, D.T., Phillips, M.J., 2015. Avian diversification patterns across the K-Pg boundary: influence of calibrations, datasets, and model misspecification. *Ann. Mo. Bot. Gard.* 100, 300–328.
- Ksepka, D.T., Ware, J.L., Lamm, K.S., 2014. Flying rocks and flying clocks: disparity in fossil and molecular dates for birds. *Proc. Roy. Soc. B – Biol. Sci.* 281, 20140677. <http://dx.doi.org/10.1098/rspb.2014.0677>.
- Lalueza-Fox, C., Gigli, E., de la Rasililla, M., Fortea, J., Rosas, A., 2009. Bitter taste perception in Neanderthals through the analysis of the TAS2R38 gene. *Biol. Lett.* 5, 809–811. <http://dx.doi.org/10.1098/rsbl.2009.0532>.
- Lalueza-Fox, C., Gigli, E., de la Rasililla, M., Fortea, J., Rosas, A., Bertranpetit, J., Krause, J., 2008. Genetic characterization of the ABO blood group in Neandertals. *BMC Evol. Biol.* 8. <http://dx.doi.org/10.1186/1471-2148-8-342>.
- Lalueza-Fox, C., Roempler, H., Caramelli, D., et al., 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neandertals. *Science* 318, 1453–1455. <http://dx.doi.org/10.1126/science.1147417>.
- Le Duc, D., Gabriel, R., Arunkumar, K., et al., 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biol.* 16. <http://dx.doi.org/10.1186/s13059-015-0711-4>.
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J.P., 2013. Capturing protein-coding genes across highly divergent species. *Biotechniques* 54, 321–326.
- Mayr, G., 2008. First substantial middle Eocene record of the Lithornithidae (Aves): a postcranial skeleton from Messel (Germany). *Annales de Paléontologie* 94, 29–37.
- McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Aust. J. Botany* 49, 271–300.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)* New Orleans, pp. 1–8.
- Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014. Statistical binning enables an accurate coalescent-based estimator of the avia tree. *Science* 346, 1250463.
- Mitchell, K.J., Llamas, B., Soubrier, J., Rawlence, N.J., Worthy, T.H., Wood, J., Lee, M.S.Y., Cooper, A., 2014. Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution. *Science* 344, 898–900.
- Myers, E.A., 2012. Ecological opportunity: trigger of adaptive radiation. *Nat. Educ. Knowl.* 3, 23.
- Oskam, C.L., Haile, J., McLay, E., et al., 2010. Fossil avian eggshell preserves ancient DNA. *Proc. Roy. Soc. B – Biol. Sci.* 277, 1991–2000. <http://dx.doi.org/10.1098/rspb.2009.2019>.
- Oskam, C.L., Jacomb, C., Allentoft, M.E., Walter, R., Scofield, R.P., Haile, J., Holdaway, R.N., Bunce, M., 2011. Molecular and morphological analyses of avian eggshell excavated from a late thirteenth century earth oven. *J. Archaeol. Sci.* 38, 2589–2595. <http://dx.doi.org/10.1016/j.jas.2011.05.006>.
- Phillips, M.J., Gibb, G.C., Crisp, E.A., Penny, D., 2010. Tinamou and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. *Syst. Biol.* 59, 90–107. <http://dx.doi.org/10.1093/sysbio/syp079>.
- Phillips, M.J., Pratt, R.C., 2008. Family-level relationships among the Australasian marsupial “herbivores” (Diprotodontia: Koala, wombats, kangaroos and possums). *Mol. Phylogenet. Evol.* 46, 594–605.
- Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Moriarty Lemmon, E., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526.
- Rambaut, A., 1996. *Sequence Alignment (Se-AL) Program*. Version 2.0a. University of Oxford, Oxford.
- Rambaut, A., Suchard, M.A., Xie, W., Drummond, A.J., 2003. Tracer: MCMC Trace Analysis Tool. Version 1.6.1pre.
- Ramírez, O., Gigli, E., Bover, P., Alcover, J.A., Bertranpetit, J., Castresana, J., Lalueza-Fox, C., 2009. Palaeogenomics in a temperate environment: shotgun sequencing from an extinct Mediterranean caprine. *PLoS One* 4 (5), e5670.
- Rohland, N., Reich, D., Mallick, S., Meyer, M., Green, R.E., et al., 2010. Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biol.* 8.
- Rompler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T., Hofreiter, M., 2006. Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313, 62. <http://dx.doi.org/10.1126/science.1128994>.
- Saareinen, J.J., Boyer, A.G., Brown, J.H., Costa, D.P., Morgan Ernest, S.K., Evans, A., Fortelius, M., Gittleman, J.L., Hamilton, M.J., Harding, L.E., Lintulaakso, K., Lyons, S.K., Okie, J.G., Sibly, R.M., Stephens, P.R., Theodor, J., Uhen, M.D., Smith, F.A., 2014. Patterns of maximum body size evolution in Cenozoic land mammals: eco-evolutionary processes and abiotic forcing. *Proc. Roy. Soc. B – Biol. Sci.* 281. <http://dx.doi.org/10.1098/rspb.2013.2049>.
- Schroeder, H., Ávila-Arcos, M.C., Malaspina, A.S., Poznik, G.D., Sandoval-Velasco, M., Carpenter, M.L., Moreno-Mayar, J.V., Sikora, M., Johnson, P.L.F., Allentoft, M.E., Samaniego, J.A., Havisser, J.B., Dee, M.W., Stafford, T.W., Salas, A., Orlando, L., Willerslev, E., Bustamante, C.D., Gilbert, M.T.P., 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. *Proc. Natl. Acad. Sci. USA* 112, 3669–3673.
- Shapiro, B., Hofreiter, M., 2012. *Ancient DNA: Methods and Protocols*. Humana Press (Springer Science), New York.
- Shapiro, B., Hofreiter, M., 2014. A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. *Science* 343, 7. <http://dx.doi.org/10.1126/science.1236573>.
- Smith, F.A., Boyer, A.G., Brown, J.H., et al., 2010. The evolution of maximum body size of terrestrial mammals. *Science* 330, 1216–1219.
- Smith, J.V., Braun, E.L., Kimball, R.T., 2013. Ratite nonmonophyly: independent evidence from 40 novel loci. *Syst. Biol.* 62, 35–49. <http://dx.doi.org/10.1093/sysbio/syb067>.
- Springer, M.S., Burk-Herrick, A., Meredith, R., Eizirik, E., Teeling, E., O’Brien, S.J., Murphy, W.J., 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Syst. Biol.* 56, 673–684.
- Stamatidakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Silvestro, D., Michalak, I., 2012. RaxmlGUI: a graphical front-end for RAXML. *Organ. Divers. Evolut.* 12, 335–337.
- Swofford, D.L., 2003. *PAUP\**. Version 4.0b. Sinauer Associates, Sunderland, Massachusetts.
- Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLOS Curr. Tree Life* 1. <http://dx.doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be>.
- Tovondrafale, T., Razakamanana, T., Hiroko, K., Rasoamiamanan, A., 2014. Palaeoecological analysis of elephant bird (Aepyornithidae) remains from the Late Pleistocene and Holocene formation of southern Madagascar. *Malagasy Nat.* 8.
- Vallender, E.J., 2009. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. *Methods* 49, 50–55.
- Willerslev, E., Cooper, A., 2005. Ancient DNA. *Proc. Roy. Soc. B – Biol. Sci.* 272, 3–16. <http://dx.doi.org/10.1098/rspb.2004.2813>.
- Worthy, T.H., Scofield, R.P., 2012. Twenty-first century advances in knowledge of the biology of moa (Aves: Dinornithiformes): a new morphological analysis and moa diagnoses revised. *New Zealand J. Zool.* 39, 87–153.
- Worthy, T.H., Worthy, J.P., Tennyson, A.J.D., Salisbury, S.W., Hand, S.J., Scofield, R.P., 2013. Miocene fossils show that kiwi (Apteryx, Apterygidae) are probably not phyletic dwarves. *Paleornithological Research: Proceedings of the 8th International Meeting of the Society of Avian Paleontology and Evolution*, pp. 63–80.
- Yang, Z., 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, Z., Rannala, B., 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* 23.
- Yoder, A.D., Nowak, M.D., 2006. Has vicariance or dispersal been the prominent biogeographic force in Madagascar? Only time will tell. *Annu. Rev. Ecol. Evol. Syst.* 37, 405–431.
- Yoder, A.D., Yang, Z., 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. *Mol. Ecol.* 13, 757–773.
- Yonezawa, T., Segawa, T., Mori, H., Campos, P.F., Hongoh, Y., Endo, H., Akiyoshi, A., Kohno, N., Nishida, S., Wu, J., Jin, H., Adachi, J., Kishino, H., Kurokawa, K., Nogi, Y., Tanabe, H., Mukoyama, H., Yoshida, K., Rasoamiamanan, A., Yamagishi, S., Hayashi, Y., Yoshida, A., Koike, H., Akishinonmiya, F., Willerslev, E., Hasegawa,

- M., 2015. Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of the ratites. *Curr. Biol.* 27, 1–10.
- Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwald, M.J., Meredith, R.W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xion, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xion, Y., Wang, G., Wang, B., Wang, J., Fan, Y., Da Fonseca, R. R., Alfaro-Núñez, A., Schubert, M., Orlando, L., Mourier, T., Howard, J.T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M.V., Hara, E., Smith, J., Farré, M., Narayan, J., Slavov, G., Romanov, M.N., Borges, R., Machado, J.P., Khan, I., Springer, M.S., Gatesy, J., Hoffmann, F.G., Opazo, J.C., Håstad, O., Sawyer, R.H., Kim, H., Kim, K., Kim, H.J., Chio, S., Li, N., Huang, Y., Bruford, M.W., Zhan, X., Dixon, A., Bertelsen, M.F., Derryberry, E., Warren, W., Wilson, R.K., Li, S., Ray, D. A., Green, R.E., O'Brien, S.J., Griffin, D., Johnson, W.E., Haussler, D., Ryder, O.A., Willerslev, E., Graves, G.R., Alström, P., Fjeldså, J., Mindell, D.P., Edwards, S.V., Braun, E.L., Rahbek, C., Burt, D.W., Houde, P., Zhang, Y., Huanming, Y., Wang, J., Avian Genome Consortium, Jarvis, E.D., Gilbert, M.T.P., Wang, J., 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346, 1311–1320.
- Zhang, G., Rahbek, C., Graves, G.R., Lei, F., Jarvis, E.D., Gilbert, M.T., 2015. Genomics: bird sequencing project takes off. *Nature* 522, 34.

—APPENDIX II—

---

**SIGNED CO-AUTHOR PERMISSIONS**

---

## II.I CHAPTER 2

---

Grealy A, McDowell MC, Scofield P, Murray DC, Fusco DA, Haile J, Prideaux GJ, Bunce M. 2015. A critical evaluation of how ancient DNA bulk bone metabarcoding complements traditional morphological analysis of fossil assemblages. *Quaternary Science Reviews* **128**: 37-47.

**STATEMENT OF CONTRIBUTION.** AG designed experiments and wrote the manuscript, with methods written by AG, MM, DF, and PS. All authors contributed to editing of the manuscript. AG, DM, MM, MB assisted with analysis. Samples were collected by MM, DM, PS, and GP. Morphological identification was carried out by MM, DF, PS.

I, Matthew McDowell (MM), confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
MATTHEW MCDOWELL

Date | 1 August 2016

I, Paul Scofield, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
PAUL SCOFIELD

Date | 15 October 2016

I, Dáithí Conall Murray, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
DÁITHÍ MURRAY

Date | 9 August 2016

I, Diana A Fusco, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
DIANA FUSCO

Date | 1 August 2016

I, James Haile, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
JAMES HAILE

Date | 3 August 2016

I, Gavin Prideaux, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

 GAVIN PRIDEAUX

Date | 1 August 2016

I, Michael Bunce, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

 MICHAEL BUNCE

Date | 9 August 2016

## II.II CHAPTER 3

---

Grealy A, Macken A, Allentoft ME, Rawlence NJ, Reed E, Bunce M. 2016. An assessment of ancient DNA preservation in Holocene-Pleistocene fossil bone excavated from the world heritage Naracoorte Caves, South Australia. *Journal of Quaternary Science* **31**: 33-45.

**STATEMENT OF CONTRIBUTION.** ER excavated and collected materials from RCEC, provided dates, stratigraphy and faunal data from morphological analyses. AM provided fauna data from morphological analyses and palaeoecological input. AG designed, performed, and analysed experiments with input from MA, MB and NR. AG wrote the manuscript with contributions and edits from all co-authors.

I, Amy Macken, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |



Date | 14 September 2016

I, Morten E Allentoft, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |



Date | 2 August 2016

I, Nicolas J Rawlence, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
NICOLAS J RAWLENCE

Date | 14 September 2016

I, Elizabeth Reed, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
ELIZABETH REED

Date | 1 August 2016

I, Michael Bunce, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
MICHAEL BUNCE

Date | 9 August 2016

## II.III CHAPTER 4

---

Grealy A, Douglass K, Haile J, Bruwer C, Gough C, Bunce M. 2016. Tropical ancient DNA from bulk archaeological fish bone reveals the subsistence practices of a historic coastal community in southwest Madagascar. *Journal of Archaeological Science (under review)*.

**STATEMENT OF CONTRIBUTION.** KD organised and directed the archaeological excavation. AG and JH assisted with the collection of bulk bone material. AG conducted genetic analyses with assistance from MB. JH designed the primers used. CG provided modern fisheries data for comparison. CB rendered the line drawings and figure. AG and KD drafted the manuscript with contributions and edits from all co-authors.

I Kristina Douglass confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

 KRISTINA DOUGLASS

Date | 1 August 2016

I James Haile confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

 JAMES HAILE

Date | 3 August 2016

I Charlotte Gough confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

DocuSigned by:  
*CHRISTELLE BRUNER*  
67C10F17BAE54BD...

Date | 2 August 2016

I Charlotte Gough confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

*CG*  
CHARLOTTE GOUGH

Date | 2 August 2016

I Michael Bunce confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

*MB*  
MICHAEL BUNCE

Date | 9 August 2016

## II.IV CHAPTER 5

---

Grealy A, Phillips M, Miller G, Gilbert MTP, Rouillard J, Lambert D, Bunce M, Haile J. Eggshell palaeogenomics: palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Molecular Biology and Evolution* (submitted).

**STATEMENT OF CONTRIBUTION.** JH and MB conceived and supervised the study. JH and AG designed the experiments. AG performed the experiments. AG and MP analysed the data. JR synthesised and supplied tinamou nuclear MYbaits. GM supplied and dated eggshell samples. MTPG and DL provided unpublished databases of *D. novaehollandiae*, *R. americana* and *A. mantelli* nuclear sequences. All authors contributed to writing and editing the manuscript.

I Matthew Phillips confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
MATTHEW PHILLIPS

Date | 1 August 2016

I, Gifford Miller, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
GIFFORD MILLER

Date | 6 August 2016

I, Marcus Thomas Pius Gilbert, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
MTP GILBERT

Date | 1 August 2016

I, Jean-Marie Rouillard, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
JEAN-MARIE ROUILLARD

Date | 1 August 2016

I, David Lambert, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
DAVID LAMBERT

Date | 1 August 2016

I Michael Bunce confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
MICHAEL BUNCE

Date | 9 August 2016

I, James Haile, confirm my contribution to the cited publication and give permission for its inclusion in this thesis.

Signature |

  
JAMES HAILE

Date | 3 August 2016

*By hook or by crook, I'll be last in this book.*

- John Lennon