School of Molecular and Life Sciences

Faculty of Science and Engineering

Epigenetic age estimation of cetaceans

Matthew James Heydenrych

ORCiD: 0000-0002-8426-3400

This thesis is presented for the Degree of

Master of Research (Environmental Science)

of

Curtin University

July 2021

Declaration

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 which has been accepted for the award of any other degree or diploma in any university.

Signed: Matthew James Heydenrych

Dated: 09 July 2021

Abstract

The ecological application of genomic technologies has seen a rapid expansion over the past decade. This has been a result of large reductions in sequencing costs, the growing importance of genomic information for biological investigations, and increasing computing power. This expansion has allowed for complex interactions within natural ecosystems and animal populations to be assessed on a molecular level. There is now an increasing interest in the use of epigenetic analyses as a potential tool for assessing a range of biological questions. This includes the determination of age, an important biological characteristic which is often difficult to assess through traditional methodologies. Age estimation is especially relevant in many cetacean species (whales, porpoises, and dolphins), where it is difficult to estimate in mature animals. Molecular approaches, such as epigenetic analyses, offer a viable and low-cost alternative. This thesis focusses on assessing the utility and considerations of epigenetic analyses for determining age in wild animals, specifically in cetaceans.

I initially review the current and potential uses of epigenetic analyses in relation to vertebrate population biology, specifically in determining age, sex, and sexual maturity (Chapter 2). Many wild species have now had epigenetic age estimation assays successfully developed and applied, with accuracy comparable to traditional methodologies. The benefits of this approach are apparent, and recent research into epigenetic mechanisms within sex determination and sexual maturity indicate the potential for epigenetic assays for these biological parameters.

I then assess how age is currently determined in cetacean species, highlighting the need for alternative approaches and the successes of previous epigenetic approaches (Chapter 3). For this taxonomic group, whose age estimation generally requires long-term and costly visual surveys for live animals, the value of epigenetic age estimation approaches is detailed.

Utilising a high-throughput amplicon sequencing approach, I develop and test a novel epigenetic workflow for age estimation utilising DNA methylation markers identified through homology from previous epigenetic research (Chapter 4). These analyses are applied to three whale species: southern right whales (*Eubalaena australis*), sperm whales (*Physeter macrocephalus*), and long-finned pilot whales (*Globicephala melas*;

Chapter 5). I produce age estimation methods for southern right and sperm whales based on significant correlations between age and methylation rates at various age-associated genomic sites. These methods demonstrate comparable accuracy to published age estimation methods for other cetacean species. Calibration dataset limitations, however, restrict the applicability of these methods, and further research is recommended. The issues with homology-based epigenetic analyses are highlighted by the difficulties encountered during assessment of the same epigenetic assays in pilot whales. A large and diverse calibration dataset is analysed through the epigenetic workflow, but only limited correlations between age and DNAm are found, precluding the development of a successful age estimation method. This indicates that the same age-associated relationships found in these gene regions in other species are not conserved in pilot whales. This is an issue that has been identified in other novel epigenetic age estimation studies utilising homology-based approaches. Alternative approaches for the development of a successful pilot whale epigenetic age estimator are discussed.

The novel Illumina-based epigenetic sequencing workflow developed in this thesis is also assessed for both reproducibility and preciseness (Chapter 6). It is found to represent both a high-throughput and precise sequencing approach which may greatly increase the speed and cost-effectiveness of marker-based epigenetic research. Finally, the importance of robust and sufficiently sized calibration datasets is detailed, with examples provided from the research conducted within this thesis and from external sources (Chapter 7).

This thesis highlights the complicated nature of epigenetic age estimator development. While the volume of epigenetically focused research in the ecological sphere is limited, the potential utility is apparent, especially for age estimation. Though alternative approaches to homology-based methylation analysis may be required, epigenetic analyses present an effective tool for age estimation in vertebrates. This is especially relevant for species with limited external age-related features, such as cetaceans. Here, I provide an alternative and potentially more cost-effective method for rapid epigenetic biomarker assessment, as well as a broad overview of epigenetic tools for age assessment in cetaceans. This research can be applied and adapted by environmental managers to the study of local cetacean population characteristics, guiding and aiding conservation and resource management strategies. Further research

should focus on increasing calibration datasets to develop more robust age estimators for southern right and sperm whales, where homology-based approaches were successful. For pilot whales, where they were unsuccessful, the identification of novel age-associated epigenetic sites should be a priority, and the subsequent creation of an epigenetic age estimator. Further to this, expansion of epigenetic analyses for age estimation in other taxa of conservation importance is recommended, as is the investigation and development of epigenetics-based assays for other key biological parameters. This thesis highlights the potential of epigenetic analyses for ecological purposes and their utility for the determination of various biological characteristics that are currently difficult or impossible to ascertain, especially within cetaceans.

Acknowledgements

I want to acknowledge and thank all the amazing friends, collaborators, and mentors that have been a part of my university and research experience so far, both in keeping me motivated and keeping me sane.

Firstly, I would like to thank my supervisors, Simon Jarman, Mike Bunce, and Ben Saunders, for pushing me to challenge myself, the constant support, and the needed critiquing when required. A special thanks to Mike, for initially bringing me into the TrEnD team and igniting my passion for genetics. You have been a constant mentor and motivator, and your support and encouragement will not be forgotten. A big thankyou as well to Simon, for introducing me to the world of epigenetics, for being a calm, grounding force with optimism when it was needed, and for pushing me to always improve and develop my skills. And to Ben, your support and guidance during the crunch times was invaluable, especially with everything else that has been going on these last few years.

I would also like to give a massive thanks to all of the TrEnD lab team. Friday beers, conversations over coffee, and shared stress were what made my last few years as wonderful as they were, let alone all the scientific knowledge I have managed to absorb! I would like to give special thanks to the fantastic lab rats: Katrina West, Sophie Preston, Frederik Seersholm, Adam Koziol, Georgia Nester, Miwa Takahashi, Mieke Van Der Heyde, Kristen Fernandes, Arne Adam, Jackie Jones, Alishum Ali, David Juszkiewicz, Jason Alexander, Josh Kestel, Sarah Crump, Alicia Grealy, Dáithí Murray, and Matt Power. Without you all, it would have been a far less enjoyable affair, and I look forward to many adventures and exciting scientific discoveries ahead.

I would like to acknowledge the Department of Primary Industries, Parks, Water and Environment (TAS), the Department of Biodiversity, Conservation and Attractions (WA), Emma Carroll, and give special thanks to Chandra Salgado-Kent for sample collection and whale age estimations. Various figures were created using Biorender.com. For assistance in laboratory workflows, I would like to thank Matthew Power, Katrina West, Mieke van der Heyde, Nicole White, and Frederik Seersholm. I would also like to thank Mahsa Mousavi Mousaviderazmahalleh for bioinformatic

support and the Pawsey Supercomputing Centre for assistance in high-performance computing.

My biggest appreciation as well to my family: Helen, Barry, Tim, and Benji. Without you, I would not be where I am today, and will forever be grateful for your push for me to follow my dreams. Our escapes down South and beers in the valley were always incredibly appreciated. And of course, the greatest thanks to my amazing partner, Katie. You have always been there supporting me, putting up with the long nights and stress, keeping me sane and happy, and it has meant the world to me.

Finally, I would like to acknowledge the CSIRO Environomics Future Science Platform, Curtin University, and the Trace and Environmental DNA Laboratory for their support and funding. I was personally supported by Curtin University's Masters stipend and an Australian Government Research Training Program fee offset.

Table of Contents

Statement of Contributions

Chapter 2. In Review. Please see Appendix B for a signed author contribution statement.

Chapter 2: Heydenrych, M.J., Saunders, B.J., Bunce, M., Jarman, S.N. 2020. Epigenetic measurement of key vertebrate population biology parameters. Submitted to *Frontiers in Ecology and Evolution*.

Author contribution: This review was conceptualised and designed by M.J.H and S.N.J, with M.J.H, S.N.J, B.J.S, and M.B contributing to the writing and editing of this paper.

List of Figures

List of Tables

Abbreviations

AIC	Akaike information criterion
bcDNA	Bisulfite-converted DNA
BIC	Bayesian information criterion
DNA	Deoxyribose nucleic acid
DNAm	DNA methylation
dNTPs	Deoxynucleotide triphosphates
CI	Confidence interval
CpG	Cytosine-guanine
GLG	Growth layer group
LOOCV	Leave One Out Cross Validation
NCBI	National Centre for Biotechnology Information
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
PW	Pilot whale
qPCR	Quantitative PCR
SRW	Southern right whale
SW	Sperm whale
TrEnD	Trace and Environmental DNA Laboratory
MLR	Multiple linear regression

Chapter 1 - General introduction

1.1 Brief background and rationale

The age of wild animals and population age structures provide an important metric in understanding population dynamics, such as recovery and growth potential (Baker & Clapham, 2004; Evans & Hindell, 2004; Riekkola et al., 2018). However, traditional methods of age estimation for wild populations are often invasive, expensive, and time consuming (Kukalová, Gazárková, & Adamík, 2013; Carroll et al., 2016; Jarman et al., 2015). Age estimation can be especially challenging in cetaceans (whales, porpoises, and dolphins), due to both their cryptic behaviour within a marine environment and their lack of external age identifying features (Chittleborough, 1965; Polanowski, Robbins, Chandler, & Jarman, 2014).

As we enter an era of low-cost and high-throughput sequencing, the potential for genetic and epigenetic tools for answering a range of biological questions is becoming evident (Jarman et al., 2015; Mayne, Berry, Davies, Farley, & Jarman, 2019; Stat et al., 2017). Recent advances in epigenetic analyses have allowed for age estimation assays to be designed and successfully applied to wild animals (De Paoli-Iseppi et al., 2019; Polanowski et al., 2014). This novel approach is of particular utility for cetacean research, where age is often difficult to assess in live animals (Evans & Hindell, 2004; Polanowski et al., 2014). This thesis explores the current and potential applications of epigenetic tools for ecological assessments, and the direct application of these tools to cetacean age estimation is investigated.

In addition, this thesis focusses upon the design and utilisation of a novel high-throughput epigenetic age estimation workflow. This workflow and its potential benefits are examined before the workflow's application to the development of an epigenetic age estimation assay in two odontocetes (toothed whales), sperm whales and long-finned pilot whales (*Physeter macrocephalus* and *Globicephala melas*, respectively), and a baleen whale, the southern right whale (*Eubalaena australis*). Finally, important considerations for the development of predictive models such as epigenetic age estimators, primarily in relation to calibration datasets, are discussed and detailed.

1.2 Research questions

The overarching research question addressed in this thesis is: What are the considerations and benefits of epigenetic age estimation for cetaceans? To address this question, this thesis aims to assess the utility of epigenetic age estimation assays in cetaceans and then to identify and discuss the key considerations in the design of these assays (Figure 1.1). The identification of further epigenetic tools for ecological purposes is also an important focus, as is the determination of the suitability of the novel high-throughput Illumina-based sequencing approach to homology-based epigenetic analyses developed within this thesis (Figure 1.1).

1.3 Thesis structure

This thesis is presented as a series of small chapters, each dealing with a discrete idea, and is designed to be considered as a whole, rather than chapters as standalone manuscripts. This thesis begins with an overview of both the potential benefits from epigenetic analyses in vertebrate population biology (chapter 2) and a focussed evaluation of age estimation in wild species, specifically cetaceans (chapter 3). The need for low impact age estimation alternatives is assessed, as is epigenetic analyses utility for this application. A novel epigenetic sequencing workflow for determining methylation levels in short genomic amplicons, based upon the Illumina platform, is then presented in chapter 4. This workflow is applied in the following data chapter (chapter 5), where epigenetic markers in homologous gene regions to previous age estimation assays are assessed in sperm, southern right, and long-finned pilot whales. The novel high-throughput sequencing approach used in these analyses is then assessed (chapter 6). In the final discussion chapter (chapter 7), these findings are synthesised, the importance of suitable high-fidelity calibration datasets is highlighted, and avenues for future research are detailed.

The literature review in chapter 2 is currently under review and has been formatted as a journal article and may therefore contain some repetition. Due to the interlinked nature of all chapters in this thesis, a combined reference list can be found at the end of this thesis in the References section. A visual flow diagram of prominent themes and questions linked to relevant chapters is provided in Figure 1.1.

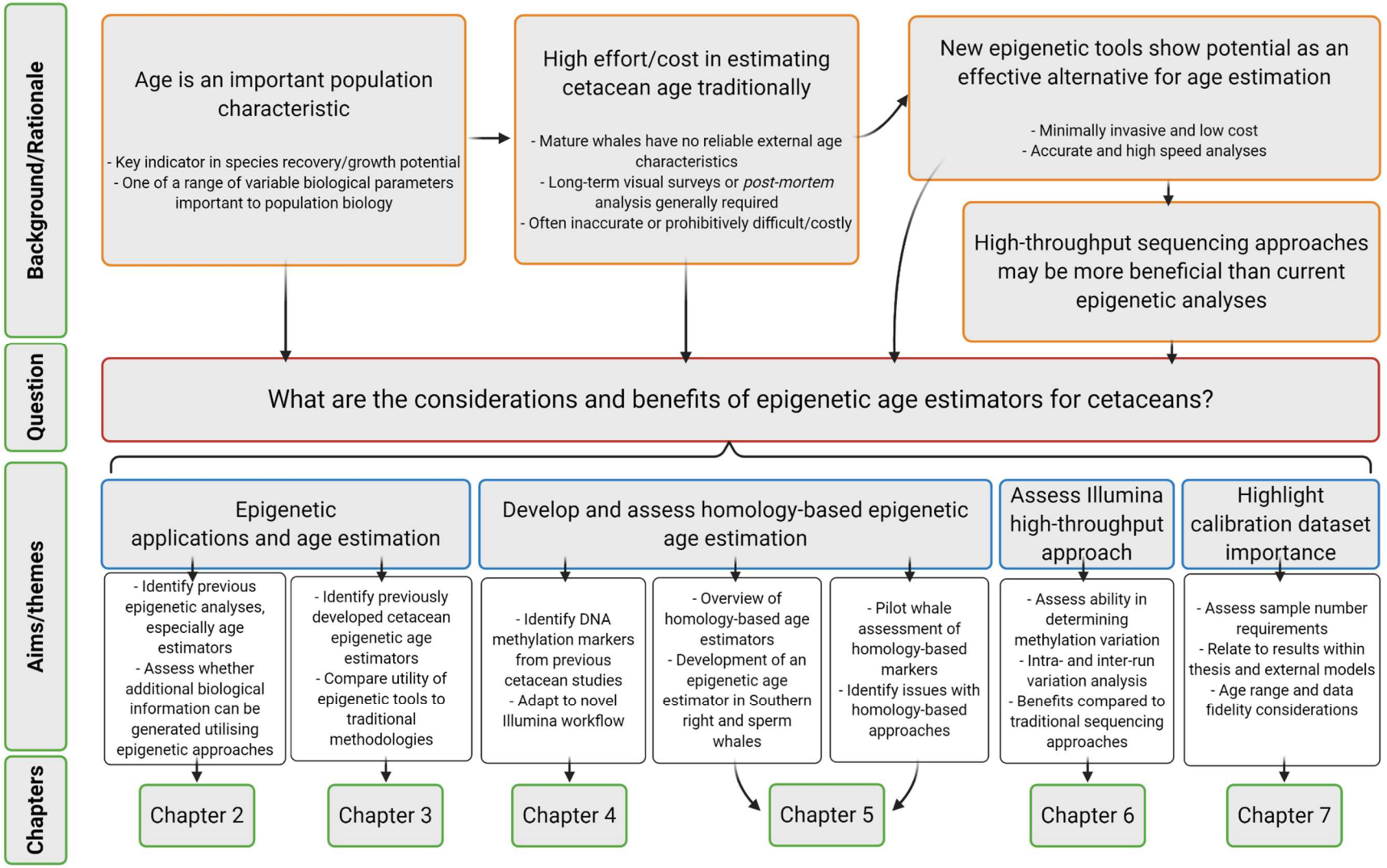

Figure 1.1 General thesis structure flow diagram.

1.3.1 Introduction to epigenetic measurement of key vertebrate population biology parameters (Chapter 2)

Chapter 2 contains a literature review of current research related to ecological applications of epigenetics in vertebrates. It assesses the potential benefits that current and proposed epigenetic assays may have in the determination of three key population biology parameters: age, sex, and sexual maturity. The current understanding of epigenetic mechanisms within these biological processes is assessed, the success of age estimation epigenetic assays highlighted, and development of assays for the other parameters proposed. Chapter 2 is currently under review in the peer-reviewed journal *Frontiers in Ecology and Evolution* (Heydenrych et al., 2020).

1.3.2 Introduction to age estimation in cetaceans (Chapter 3)

Chapter 3 outlines the importance of age estimation in ecological contexts, especially regarding recovering whale populations. Historical approaches to age estimation in cetaceans are detailed, as are recent molecular analyses that have shown promise as an alternative approach, with an emphasis on epigenetic approaches.

1.3.3 Novel high-throughput epigenetic analysis workflow (Chapter 4)

Chapter 4 details the design and methodology of the novel workflow utilised for epigenetic analysis and the development of age estimation models in this thesis. This workflow is based upon the Illumina sequencing platform and represents a high-throughput approach to homology-based epigenetic analysis and age estimation.

1.3.4 Homology-based epigenetic age estimation in sperm, southern right, and pilot whales (Chapter 5)

Chapter 5 represents the primary data chapter of the thesis, where epigenetic analysis of homology-based age-associated sites is undertaken in sperm, southern right, and long-finned pilot whales. A range of genomic sites identified in previous research, including in cetaceans, where age correlated with DNA methylation levels are assessed. This is achieved through the utilisation of the novel Illumina-based sequencing workflow detailed in chapter 4. Linear models for age estimation are developed, and limitations of the datasets, age associations, and the resulting models are investigated. Previously identified issues with the use of homology-based epigenetic marker development are also discussed.

1.3.5 Assessment of the novel Illumina-based high-throughput sequencing epigenetic analysis workflow (Chapter 6)

Chapter 6 assesses the potential benefits and suitability of the novel high-throughput Illumina-based sequencing workflow designed and utilised in this thesis for epigenetic analysis. The accuracy and reproducibility of the sequencing results are investigated, and comparisons to current epigenetic sequencing approaches are discussed.

1.3.6 Discussion and future directions (Chapter 7)

Chapter 7 synthesises the concepts identified and the research conducted throughout the thesis. Here, the advantages and challenges related to epigenetic age estimation and the novel workflow developed within the thesis are discussed. The importance of calibration datasets in the development of accurate and reliable epigenetic age estimation models is also emphasized. Finally, avenues for further epigenetically based ecological research are proposed.

Chapter 2 – Introduction to epigenetic measurement of key vertebrate population biology parameters

2.1 Abstract

The age, sex, and sexual maturity of individual animals are key parameters in assessing wild populations and informing conservation management strategies. These parameters represent the reproductive potential of a population and can indicate recovery rates or vulnerabilities. Natural populations of wild animals are difficult to study; logistically, economically, and due to the impacts of invasive biomonitoring. Genetic and epigenetic analyses offer a low impact, low cost, and information-rich alternative. As epigenetic mechanisms are intrinsically linked with both biological aging and reproductive processes, DNA methylation can be used as a suitable biomarker for population biology study. This review assesses published research utilising DNA methylation analysis in relation to three key population parameters: age, sex, and sexual maturity. We review studies on wild vertebrates that investigate epigenetic age relationships, with successful age estimation assays designed for mammals, birds, and fish. For both determination of sex and identification of sexual maturity, very little has been explored regarding DNA methylation-based assays. Related research, however, confirms the links between DNA methylation and these processes. Future development of age estimation assays for underrepresented and key conservation taxa is suggested, as is the experimental development and design of DNA methylation-based assays for both sex and sexual maturity identification, further expanding the genomics toolkit for population biology studies.

2.2 Introduction

We are currently experiencing a worldwide environmental crisis, with animal species increasingly threatened by habitat loss, pollution, climate change, and overexploitation (Darimont, Fox, Bryan, & Reimchen, 2015; Hammerschlag & Gallagher, 2017; Hooper et al., 2012; Maxwell, Fuller, Brooks, & Watson, 2016; Nabi et al., 2018; Thomas et al., 2004; Zabel et al., 2019). Anthropogenic stressors have impacted almost all vertebrate species, with more than one-fifth characterised as “threatened” by the

International Union for Conservation of Nature (Hoffmann et al., 2010). Effective conservation policies are therefore necessary to slow global biodiversity losses and maintain viable populations of vertebrates (Hoffmann et al., 2010; Reed, O'Grady, Brook, Ballou, & Frankham, 2003). Monitoring the efficacy of conservation efforts and determining current population trends in threatened species requires knowledge of key parameters of a species' population biology, such as rates of birth, death, and migration.

The size, growth rate, and distribution of an animal population is determined by various life history traits, including age distribution, sex ratio, reproductive interval, and litter size (McNab, 1980; McRae, Böhm, Deinet, Gill, & Collen, 2012; Reed et al., 2003). Population trend assessments can be made by measuring these characteristics within wild populations. Historically this has been difficult, requiring extensive longitudinal studies incorporating visual surveys or trapping programs (Amano, Yamada, Brownell, & Uni, 2011; Camacho, Canal, & Potti, 2017; Fox, Benjamins, Masden, & Miller, 2018; Thompson, 2007). Furthermore, some research methods for measuring these features, including trapping and other animal handling approaches, can have negative impacts upon study populations such as long-term behavioural changes (Kukalová, Gazárková, & Adamík, 2013; Camacho et al., 2017). Post-mortem study is also required for some characteristics, but this is rarely possible for practical or ethical reasons when threatened, endangered, or protected species are involved (Jarman et al., 2015). These considerations highlight the need for low impact alternatives.

Genomic tools provide an alternative, minimally invasive method for measurement of many population biology parameters (Figure 1). Epigenetic analyses, specifically through the analysis of DNA methylation (DNAm), are now allowing for an increased range of animal biological parameters to be studied (Bossdorf, Richards, & Pigliucci, 2008; De Paoli-Iseppi, Deagle, et al., 2017; Herrel, Joly, & Danchin, 2020; Rey et al., 2020). DNAm analysis primarily involves the determination of percentage rates of methyl group presence at the C5 position of a Cytosine nucleotide adjacent to a Guanine nucleotide, commonly referred to as CpG sites (Hannum et al., 2013; Jin & Liu, 2018; Smith, Smith, Kenny, Chaudhuri, & Ritchie, 2015). CpG sites are often clustered into 'islands' in 5' regulatory regions of vertebrate genes and are involved in regulating gene expression through hypermethylation (increased methylation,

generally decreased transcription) and hypomethylation (decreased methylation, generally increased transcription) within the regulatory region (Hannum et al., 2013). By assessing specific CpG sites, percentage rates of methylation can be determined throughout an animal's life, and this information can be used in informing on a range of characteristics. DNAm analyses have been applied to the age estimation of wild animals (Polanowski et al., 2014); population differentiation, evolutionary drivers, and phenotypic variation (Baldanzi, Watson, McQuaid, Gouws, & Porri, 2017; Caracappa et al., 2016; Liu et al., 2012); mechanisms of sexual differentiation and developmental processes (Morán & Pérez-figueroa, 2011; Road, Nagabhushana, & Mishra, 2016); and for ecotoxicological purposes (Nilsen et al., 2016; Wang, Wang, Zhang, Chen, & Zuo, 2009).

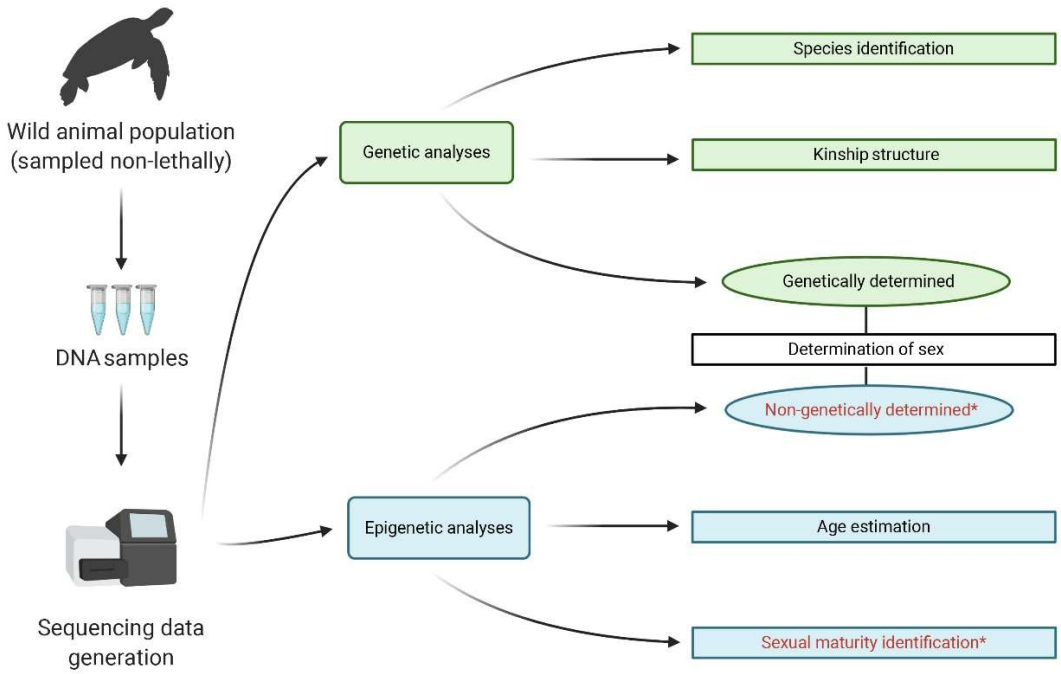

Figure 2.1 Genetic and epigenetic analyses of vertebrate population biology. Emerging epigenetic analyses hypothesised in this study are denoted with asterisks.

The successful use of DNAm analysis for measuring these population, evolutionary, and biological characteristics raises the possibility of its application to other life history traits (Figure 2.1). In this review, we identify recent advances within the discipline and important ecologically focused studies utilising DNAm analysis across three key population biology parameters: age, sex, and attainment of sexual maturity. The

potential applications of DNAm analysis are highlighted and key future foci identified for this new tool in the ecologist's toolbox.

2.3 Age Estimation

As an organism ages, many biological characteristics change, including size, behaviour, and sexual maturity. These changes impact an individual's ability to survive and reproduce effectively. This in turn affects the population as a whole, and population age structures can indicate past events and future growth potential (Jarman et al., 2015; Riekkola et al., 2018). Population age profiles can therefore act as an indicator of population decline or recovery (Lebreton, Burnham, Clobert, & Anderson, 1992; Riekkola et al., 2018). Historically, most age estimation methods required either post-mortem study, costly long-term visual surveys, or invasive capture and release strategies (Kukalová, Gazárková, & Adamík, 2013; Carroll et al., 2018; Chevallier, Gauthier, & Berteaux, 2017). The ability to determine age using DNAm analysis of samples obtained from live wild animals offers a minimally invasive alternative.

Age estimation by DNAm analysis is a recent development, with assays first developed for humans (Bocklandt et al., 2011; Horvath, 2013). DNAm was initially shown to correlate with age in humans at promoters of three gene regions (Bocklandt et al., 2011). Subsequently, assays were developed to age a range of human and mouse tissues through methylation profiles of a set of CpG sites (Bocklandt et al., 2011; Horvath, 2013; Horvath & Raj, 2018; Koch et al., 2011). DNAm age estimators for wild animals were adapted and produced soon after, with the first being for the humpback whale (*Megaptera novaeangliae*; Polanowski et al., 2014). The humpback epigenetic age assay achieved a mean difference between predicted and actual age of 3.8 years (when assessing whales of known age), an estimation more accurate than alternative methods at the time. This model was then successfully applied to wild whale populations of indeterminate age (Polanowski et al., 2014; Riekkola et al., 2018). This highlighted the potential of DNAm analysis for accurate age estimation of wild species, and importantly for species where age estimation through other methodologies was practically or ethically difficult (Jarman et al., 2015).

Epigenetic age assays have several constraints that are not common in most genetic analyses. Due to DNAm variation between tissue types, a result of gene expression variation between tissues, only a single tissue type can generally be used for age

estimators (Anastasiadi, Esteve-Codina, & Piferrer, 2018; Jung et al., 2019). Also, age associated CpG sites may vary between species, both in their location and strength of correlation (De Paoli-Iseppi, Polanowski, et al., 2017). While assays developed for one species have been applied to another, including a study on minke whales utilising the humpback assay, different CpG sites correlated with age than those identified for humpback whales (Tanabe et al., 2020). This indicates the potential need for complete re-identification of age associated CpG sites in each new species. The initial development effort appears to be generally justified, however, with a variety of successful age estimation models developed for captive and model organisms (Anastasiadi & Piferrer, 2020; R. Lowe et al., 2020). It is now being rapidly adopted by the ecological field, with the number of DNAm-based age associated studies in wild animals increasing steadily since the first in 2014 (Table 2.1).

Table 2.1 DNAm and age associated studies on wild animals. [A] - Model precision represented as the standard deviation of mean difference between known and estimated ages. [B] - Model precision represented as mean absolute error. [C] - Model precision represented as mean absolute difference/deviation (between estimated and actual age). [D] - The ratio between precision (stated by the age estimation model of each study) and the estimated mean lifespan of the target taxa, represented as a percentage. ND - not determined.

Target Species	Study Focus	DNAm Analysis Type	Model Precision	Author and Year
American alligator (*Alligator mississippiensis*)	Influence of age, contaminant exposure, and tissue type on global DNAm variation in wild reptiles	Global	ND	(Parrott et al., 2014)
American alligator (*Alligator mississippiensis*)	Mercury and aging impacts on DNAm levels in wild reptiles	Global	ND	(Nilsen et al., 2016)
Antarctic minke whale (*Balaenoptera bonaerensis*)	Investigation of age related CpG sites in a marine mammal	CpG Sites	ND	(Tanabe et al., 2020)
Bechstein's bats (*Myotis bechsteinii*)	Age estimation assay for wild bats	CpG Sites	1.52 years[A] (7.23% of lifespan[D])	(Wright et al., 2018)
Black grouse (*Lyrurus tetrix*)	DNAm relationships with age and sexual trait expression in wild birds	Global	ND	(Soulsbury et al., 2018)

Chimpanzees (*Pan trogolodytes*)	Age estimation assay for wild apes	CpG Sites	5.41 years[C] (12.0% of lifespan[D])	(Ito, Udono, Hirata, & Inoue-Murayama, 2018)
Common bottlenose dolphin (*Tursiops truncatus*)	Age estimation assay for a wild marine mammal	CpG Sites	4.83 years[C] (9.66% of lifespan[D])	(Beal, Kiszka, Wells, & Eirin-Lopez, 2019)
Darevskia lizards (*D. armeniaca* and *D. raddei*)	Analysis of impacts of age and pollutants on DNAm levels	Global	ND	(Sargsyan, Simonyan, Hovhannisyan, Arakelyan, & Aroutiounian, 2019)
Humpback whales (*Megaptera novaeangliae*)	Age estimation assay for wild marine mammals	CpG Sites	2.99 years[A] (5.98% of lifespan[D])	(Polanowski et al., 2014)
Humpback whales (*Megaptera novaeangliae*)	Age estimation assay application on wild marine mammals	CpG Sites	2.99 years[A] (5.98% of lifespan[D])	(Riekkola et al., 2018)
Short-tailed shearwaters (*Ardenna tenuirostris*)	Age estimation assay for wild birds (unsuccessful)	CpG Sites	ND	(De Paoli-Iseppi, Polanowski, et al., 2017)
Short-tailed shearwaters (*Ardenna tenuirostris*)	Age estimation assay for wild birds	CpG Sites	2.81 years[C] (9.37% of lifespan[D])	(De Paoli-Iseppi et al., 2019)

Wolves (*Canis lupus*)	Age estimation assay for dogs and wild wolves	CpG Sites	0.79 years[B] (9.88% of lifespan[D])	(Thompson, VonHoldt, Horvath, & Pellegrini, 2017)

While many successful DNAm age estimation assays have been developed for mammals, there has been limited research involving non-mammalian species. Only two groups have investigated DNAm and age relationships in wild birds. A study focused on the short-tailed shearwater (a seabird) found age-associated CpG sites identified in mammals were not conserved in birds, and an accurate age assay could not be generated using the selected markers (De Paoli-Iseppi, Polanowski, et al., 2017). Follow-up research, however, was able to successfully identify age-related genomic regions and develop the first epigenetic age estimation assay for birds (De Paoli-Iseppi et al., 2019). The other bird-focused study found DNAm relationships between age and heterozygosity in black grouse, but their goal was not to develop an assay for age determination (Soulsbury et al., 2018). In reptiles, age negatively correlated with global DNAm in wild alligators in two studies, a trend related to the process of epigenetic drift, but age estimation was not the objective of either study (Jones, Goodman, & Kobor, 2015; Nilsen et al., 2016; Parrott et al., 2014). Similarly, a recent study on lizards also identified relationships between increased age and global demethylation, but age determination once again was not attempted (Sargsyan et al., 2019). No DNAm-based age estimation studies on wild amphibians or fish have been completed to date. However, the first piscine epigenetic clock has been developed using captive seabass, with age estimation possible to within 2.15 years (Anastasiadi & Piferrer, 2020). These studies all highlight the efficacy of epigenetic age estimation across taxa groups, and the expansion of research into critical conservation species is recommended, especially for those taxa where traditional age estimation techniques are not appropriate.

2.4 Sex Identification and Differentiation

Sex is one of the most significant traits affecting an individual animal's phenotype. Population sex ratios are important in determining population growth rate potential and often reflect the effects of environmental pressures (Donald, 2007). However, sex is difficult to determine visually in many species (Shaw, Wilson, & White, 2003). This is exemplified by the many bird species that do not have external sexual characteristics (Griffiths & Tiwari, 1993). Genetic methods for determining bird sex by PCR testing for sex chromosomes have been developed to overcome this problem (Griffiths & Tiwari, 1993; Huynen, Millar, & Lambert, 2002). Similar tests for mammals have also been developed (Kurose, Masuda, & Tatara, 2005; Lindsay & Belant, 2008; Shaw et al., 2003; Strah & Kunej, 2019). Many vertebrates, however, do not have chromosomally determined sex (Trukhina, Lukina, Wackerow-Kouzova, & Smirnov, 2013). Turtles undergo temperature-dependent sex determination, and several fish species can change sex throughout their lives (Allsop & West, 2003; Bista & Valenzuela, 2020; Radhakrishnan, Literman, Neuwald, & Valenzuela, 2018). This environmental sex determination involves large-scale epigenetic regulation of gene expression (Tachibana, 2016).

DNAm levels have been linked to sexual differentiation, maintenance, and age of sexual maturation (Domingos et al., 2018; Morán & Pérez-figueroa, 2011; Radhakrishnan et al., 2018; Wen et al., 2014). As sexual differentiation is facilitated by epigenetic mechanisms, DNAm analyses should theoretically be able to identify sex using biomarkers for sex specific DNAm variation. However, to date, no such assay has been developed, but research into epigenetic mechanisms of sex determination have demonstrated that it is possible (Kuroki & Tachibana, 2018; Matsumoto, Buemio, Chu, Vafaee, & Crews, 2013; Metzger & Schulte, 2018). For example, DNAm levels regulating the aromatase gene in turtles and fish are known to be altered by temperature, leading to either male or female gonadogenesis (Matsumoto et al., 2013; Navarro-Martín et al., 2011). By assessing methylation patterns within these genes, related sex-specific genes, and their promoter regions, markers for sex identification could be developed.

Most studies into epigenetic control of vertebrate sexual mechanisms and development are performed under controlled laboratory conditions (Laing et al., 2018; Ortega-

Recalde, Day, Gemmell, & Hore, 2019; Wen et al., 2014). Recently, however, research has begun to also focus upon wild animal populations (Domingos et al., 2018; Martín-del-campo, Bárcenas-ibarra, Sifuentes-romero, Llera-Herrera, & García-Gasca, 2018; Soulsbury et al., 2018). The first such study assessed Atlantic salmon and DNAm levels during maturation, finding high DNAm variation between two of three analysed tissues types between maturation states (Morán & Pérez-figueroa, 2011). This suggested that early maturation was mostly mediated by epigenetic processes and not genetic variation. Contaminant impact on gonadal growth was assessed in the European eel, finding a potential role of DNAm in transcriptional regulation of a gene linked to gonadal development in female eels (Pierron et al., 2014). Similar correlations were found in the Japanese flounder between DNAm levels and the sex-related genes *dmrt1* and *cyp19a*, which are responsible for testes development and enzymes involved in estrogen synthesis (Wen et al., 2014). These links were supported by previous studies under experimental conditions (Navarro-Martín et al., 2011). A recent study also found that DNAm is a likely mechanism for natural sex maintenance, with DNAm suggested to be largely responsible for regulating expression of *cyp19a* in Chinese sea perch (Chen, He, Wang, & Li, 2018). Correlations have also been found between *dmrt1* and *cyp19a* expression and DNAm, as well as sex-specific alternative splicing, within wild barramundi (Domingos et al., 2018). A whole genome approach in stickleback fish demonstrated hypermethylation of the X chromosome in females, suggesting DNAm was playing a role in supressing X and Y chromosome recombination and possible sex-specific gene expression levels (Metzger & Schulte, 2018). Further research on zebrafish linked DNAm with both gonad transformation and identified that demethylation of female-linked genomic regions was critical in sex determination for that species (Ortega-Recalde et al., 2019). These studies all support the relationship between sex and DNAm variation, highlighting both the importance of epigenetic processes in sex determination and the future potential of DNAm biomarkers for sex identification.

2.5 Sexual Maturity

Sexual maturity is reached when an organism can first successfully reproduce. The age that sexual maturity is reached varies greatly in vertebrates, both within and between

species (Angelini & Ghiara, 1984; Ricklefs, 2010). The age of first reproduction (AFR) is significant in determining the reproductive potential of a population (Angelini & Ghiara, 1984; Mourocq et al., 2016). The AFR can be used to determine the total fecundity of an individual by deducting the AFR from the total lifespan, or reproductive senescence in some mammals (Ellis, Franks, Nattrass, Cant, et al., 2018; Mourocq et al., 2016). Determining the AFR is therefore key in understanding a species' reproductive capability and thus an important consideration in conservation management strategies.

In humans, and many other vertebrates, puberty involves the release of gonadotropin-releasing hormone (GnRH) and the activation of the hypothalamic-pituitary-gonadal axis, a set of endocrine glands which are critical to the functioning of the reproductive system. This leads to the production of sexual hormones and the transition into sexual maturity (Lomniczi et al., 2013; Ojeda et al., 2010; Rzeczkowska, Hou, Wilson, & Palmert, 2014). Epigenetic controls have recently been identified as a key mechanism within this process (Aylwin, Toro, Shirtcliff, & Lomniczi, 2019; Lomniczi et al., 2013).

Sexual maturation in humans has been found to be mediated by epigenetic controls, with hypermethylation of key puberty-linked genes *Eed* and *Cbx7* preceding puberty (Lomniczi et al., 2013). Inhibiting DNA methylation at these sites caused pubertal failure in rats. In rhesus monkeys, increased methylation of the *GnRH* gene and increased levels of *GnRH* mRNA (messenger RNA) as puberty progressed has been demonstrated (Kurian & Terasawa, 2013). Similarly, dosing zebrafish with dexamethasone to imitate early life stress altered methylation in a *GnRH* gene promoter region, resulting in decreased spermatogenesis and reduced reproductive functions (Khor, Soga, & Parhar, 2016). More recently, conserved patterns of epigenetic controls in rat and goat sexual maturation have been indicated (Yang et al., 2018). Two genes linked with pubertal pathways, *Edn3* and *PTPRN2*, showed similar increases in expression but significant variation in methylation patterns between species, indicating their potential role in the timing of species-specific puberty onset. In bovines, DNAm within sperm cells changes between early pubertal and late pubertal states (Lambert et al., 2018). Bulls aged 10 months (early pubertal) were shown to have altered methylation profiles compared to sexually mature bulls. This potentially resulted in the value of early pubertal bull's spermatozoa being compromised and

leading to negative phenotypic traits due to altered epigenetic landscapes. This is a phenomenon initially identified in human studies (Kobayashi et al., 2009). Recent research has also identified varied DNAm patterns across the pubertal boundary, with clear differences in DNAm levels between pubertal states reported in pigs (Yuan et al., 2019). This indicates the potential presence of differentially methylated CpGs for the development of DNAm assays to assess sexual maturity.

2.6 Conclusion and Future Perspective

Genetic tools are well established in conservation and ecological applications. This is due to their ability to generate large amounts of data from small tissue samples harvested from wild animals with minimal impact (Angeloni, Wagemaker, Vergeer, & Ouborg, 2012). Their widespread uptake has been further facilitated by the continued development of high-throughput DNA sequencing platforms and related reductions in sequencing costs during the last decade (Schwarze et al., 2020). The importance of epigenetic processes in an ecological context is now also becoming clear (Herrel et al., 2020; Parrott & Bertucci, 2019). This is attributable to both the bridging role between environments and the individual that epigenetics fills, and its intrinsic links with many developmental mechanisms (Rey et al., 2020). The information contained within the methylome can help researchers understand many of these processes, as well as be utilised as a valuable biomarker for informing on individual characteristics. Epigenetic research has, to date, been largely restricted to human and model organism study. These analyses are now being extended to non-model and wild animal species. They can now be utilised to assess key population parameters which have previously been prohibitively difficult or impossible to determine.

A range of wild species have now had age estimation assays based upon DNAm markers successfully designed and applied (De Paoli-Iseppi et al., 2019; Polanowski et al., 2014; Wright et al., 2018). This has allowed for the determination of age structures within populations and as a tool for assessing population recovery (Riekkola et al., 2018). The development of DNAm-based age estimation assays for wild reptiles, amphibians, and fish, as well as further increasing those for mammals and birds, would further facilitate the assessment of this important biological trait in wild populations. This is especially true for threatened, endangered, or protected species where traditional methods are not viable. Age estimation assay development for key

threatened species should therefore be a future focus, as should the attempted development of multi-species assays for more cost-effective applications of this tool.

Epigenetic processes involved with both sexual maturity and sex determination in vertebrates have now been identified, opening the doors to a range of research prospects (Lomniczi et al., 2013; Rzeczkowska et al., 2014). The identification of sex specific DNAm patterns could be utilised to determine animal sex quickly and accurately. This is especially important in species where sex is not chromosomally determined and impossible to test for with current genomic tools or visually assess and should be a research priority. Additionally, continued research into DNAm processes involved with sexual maturity could allow for the development of assays able to determine the pubertal stage of animals. DNAm biomarkers as an indicator for species-specific sexual maturity would facilitate a greater understanding of this key parameter for population reproductive capability (Casale, Mazaris, & Freggi, 2011). Future research should focus upon the analysis of DNAm differentiation between prepubertal and sexually mature vertebrate taxa, with subsequent assays for maturity developed.

The next important steps are to continue to develop and assess these epigenetically focused assays for a greater range of vertebrate species and applications. A combined epigenetic toolkit would be able to rapidly identify three important parameters of population biology from one tissue sample. The age, sex, and pubertal stage could all be assessed. These could potentially be multiplexed into a singular assay and be complemented by conventional population genetic or kinship analyses (Figure 2.1). This would allow for a rapid, cost-effective, and in-depth assessment of a wild population with minimal sampling requirements. This toolkit would facilitate effective and informed management decisions, allowing for expanded biomonitoring options in a rapidly shifting and heavily impacted global biodiversity environment.

Chapter 3 – Introduction to age estimation in cetaceans

3.1 Importance of age estimation

An animal's age is intrinsically linked with its biology, and age determination can be vital in understanding species ecology. Age information can be used to investigate relationships in developmental processes, reproductive potential and success, mortality rates, and overall population health (Campana, 2001; Chaloupka, Osmond, & Kaufman, 1999; Clapham, 1992; Cook et al., 2006; De Silva et al., 2013; Low & Pärt, 2009). Understanding these relationships can allow for the development of effective management policies and in guiding conservation efforts (De Silva et al., 2013; Riekkola et al., 2018). Several age estimation methods have been developed for wild animals, generally focusing on physical and superficial characteristics, such as tooth growth, animal size, or bone ossification (Chevallier et al., 2017; Martin, Reynolds, & Richardson, 1987; Serrano, Gallego, & Perez, 2004; Toïgo et al., 2007; White et al., 2016). Drawbacks to these methods are that either the animal must be captured and euthanised or found deceased to determine age, or alternatively the animal must have visible external features that allow for age estimation (Toïgo et al., 2007). For many marine species, however, including cetaceans, individuals have no outward indicators of age once maturity is achieved, and visual age estimation of a live, unknown adult individual is impossible (Chittleborough, 1965; Polanowski et al., 2014).

3.2 Age estimation in cetaceans

In cetaceans, the age of an unknown individual can be determined through the analysis of ear plug growth layer groups (GLGs), baleen plate thickness, or tooth striation counts, dependant on species (Amano et al., 2011; Chittleborough, 1959, 1965; Gabriele, Lockyer, Straley, Jurasz, & Kato, 2010; Hamilton, Knowlton, Marx, & Kraus, 1998; Lockyer, Hohn, Doidge, Heide-Jørgensen, & Suydam, 2007). These methods all require post-mortem study, which is not always compatible with live population monitoring, and especially unsuitable for species with low population numbers or an endangered status.

For live whales, there are only two methods of estimating animal age. The first involves tracking whales from birth or juvenile stages, artificially tagging or identifying unique features such as tail fluke markings, and then maintaining databases of each whale's age (Chaloupka et al., 1999; Elwen et al., 2014; Kniest, Burns, & Harrison, 2009). This process is resource intensive, requiring long-term photographic identification projects, and is further complicated by the cryptic nature of whales and inherent difficulties of biomonitoring within the marine environment (Amos, Schlötterer, & Tautz, 1993; Chaloupka et al., 1999; De Silva et al., 2013; Ellis, Franks, Nattrass, Currie, et al., 2018). The second method involves molecular analysis, whereby tissue samples are collected through minimally-invasive strategies and analysed, allowing not only for age estimation, but also for the identification of population sizes, structures, and migratory routes; kinship; and sex determination (Griffiths & Tiwari, 1993; Morin, Nestler, Rubio-Cisneros, Robertson, & Mesnick, 2005; Palsbøll et al., 1997, 1995; Pomilla & Rosenbaum, 2005; Valsecchi, Hale, Corkeron, & Amos, 2002).

3.3 Molecular approaches to age estimation

The ability of molecular methods to estimate age in wild animals facilitates the collection of important ecological data. They are especially valuable for studies of live populations of endangered or protected species as they allow age to be determined through minimally invasive strategies. Various molecular assays exist to estimate age, including the analysis of telomere lengths, T-cell rearrangements, lipid profiling, and the analysis of mRNA level variations, but all of these have limitations that restrict their use (Cook et al., 2006; Herman et al., 2009; Karlsson, Svensson, Marklund, & Holmlund, 2008; Zubakov et al., 2010). Previous attempts to implement age determination using molecular analysis have often focused on telomeres, repetitive nucleotide caps on the ends of chromosomes, whose tendency to shorten over an animal's lifespan can allow for the estimation of age (Herman et al., 2009; Karlsson et al., 2008; Olsen, Bérubé, Robbins, & Palsbøll, 2012). However, this rate of shortening tends to vary greatly between individuals, as do the initial telomere lengths at birth, factors which greatly reduce applicability and accuracy of age estimation assays across populations (Karlsson et al., 2008; Nussey et al., 2014; Olsen et al., 2012).

Subsequently, telomere assays have proven ineffective for population level age estimation in almost all cases (Dunshea et al., 2011). Another molecular-based method involves the analysis of lipid profiles in blubber, but results can be variable in precision, and specialised equipment and extensive calibration for each population analysed is required (Herman et al., 2009). For mRNA analyses, a major drawback is that mRNA levels are expensive to measure and provide limited accuracy (Cook et al., 2006). In T-cell rearrangement analysis, there is the requirement for extensive molecular data on a species and it has been difficult to transfer to non-human species (Miller, 2001; Zubakov et al., 2010). Adding to these drawbacks, all molecular methods report on 'biological' age, and not exact 'chronological' age, which acts as a proxy for an animal's true age. This leads to variability across populations in the relationships between chronological and biological age based on a range of factors, such as environmental conditions; external stressors and pollutants; and diet (Chatterjee, Gim, & Choi, 2018; Jarman et al., 2015; Lea, Altmann, Alberts, & Tung, 2016; Matsumoto & Crews, 2012; Pilsner et al., 2010). This discrepancy between a measured biological age and true chronological age is the main source of error for molecular age estimation assays and may cause differences in measured age responses between populations (Jarman et al., 2015; Polanowski et al., 2014). Using DNAm as a biomarker for age offers an alternative molecular approach that can minimise this biological noise identified between populations, with DNAm relationships shown to correlate similarly with age across multiple populations (Gopalan et al., 2017). By identifying clock-like methylation that occurs at genomic sites that correlate with age, accurate age estimation methods can be developed. These sites typically occur in gene promoters, and DNAm levels are therefore preserved across populations due to the intrinsic links between methylation within gene promoters and their functions (Gonzalez-Zulueta et al., 1995; Horvath, 2013). DNAm analysis has since shown its potential as an accurate age estimator in a range of vertebrate species, from humans to bats, that can be applied successfully between populations (Ito et al., 2018; Polanowski et al., 2014; Weidner et al., 2014; Wright et al., 2018).

3.4 Epigenetic age estimation

DNA methylation is an epigenetic mechanism for regulating gene expression levels and is one of the most studied epigenetic modifications in vertebrates (Bossdorf et al., 2008). Methylation analysis enables the study of a range of biological characteristics, from pollutant exposure to evolutionary processes, with significant recent advancements around age estimation (Bekaert, Kamalandua, Zapico, Van de Voorde, & Decorte, 2015; Hannum et al., 2013; Horvath, 2013; Jarman et al., 2015; Lea et al., 2016; Nilsen et al., 2016). DNAm analysis involves the determination of percentage rates of methyl group presence/absence at the C5 position of a cytosine nucleotide adjacent to a guanine nucleotide, commonly referred to as 'CpG' sites (Jin & Liu, 2018; Polanowski et al., 2014). This process is discussed in detail in chapter 2. A process called epigenetic drift has been shown to cause genome-wide hypomethylation as an organism ages (Teschendorff, West, & Beck, 2013). While epigenetic drift is not predictable enough to age an organism, 'clock-like' methylation patterns have been identified at select genomic loci where cytosine methylation at specific CpG sites correlate directly with an animal's age, especially in age-related functional genes (Hannum et al., 2013; Horvath, 2013; Thompson et al., 2017). There are also tissue specific DNAm patterns, as different tissue types have varied cell cycles and functions, and epigenetic analyses must account for this in their methodological approach (Chun Yang, Zhang, Liu, Lu, & Li, 2018).

The age of many vertebrates has been linked with methylation changes at various CpG sites, specifically those within genes with cell cycle expression changes and therefore clock-like methylation patterns (Bocklandt et al., 2011; Horvath & Raj, 2018; Maegawa et al., 2010). Accurate age estimation assays have subsequently been developed for model species by utilising these relationships (Bekaert et al., 2015; Bocklandt et al., 2011; Thompson et al., 2017). By adapting age-associated genes identified in model organisms to homologous genes in non-model vertebrates, age estimation assays have also been developed and applied to a range of wild species (De Paoli-Iseppi, Polanowski, et al., 2017; Ito et al., 2018; Jarman et al., 2015; Polanowski et al., 2014; Riekkola et al., 2018; Wright et al., 2018). However, this is not globally applicable, with a recent study focused on a long-lived seabird, genes orthologous to mammalian genes with age relationships in model organisms presented limited correlations between DNAm and age, indicating variability between the methylation

profiles of distantly related taxa, but not excluding the possibility of other, unidentified age-related genes (De Paoli-Iseppi, Polanowski, et al., 2017). This was confirmed to be the case through further research, where the same group identified novel age-related markers, developing the first successful epigenetic age estimation assay for birds (De Paoli-Iseppi et al., 2019).

3.5 Epigenetic age estimation in cetaceans

In cetaceans, humpback whales (*Megaptera novaeangliae*) were the first species to have an age estimation assay successfully developed, with population age structures determined and the species' recovery post-whaling prohibition investigated (Polanowski et al., 2014; Riekkola et al., 2018). This analysis was applied to wild populations by 'biopsy darting' individuals at sea, a minimally invasive and cost-effective sampling strategy to obtain a standardised tissue sample. These biopsies were able to age live animals with a mean average difference of 2.99 years, a comparable result to a human assay's ability to predict age to within 3.75 years (Bekaert et al., 2015; Polanowski et al., 2014). This demonstrated the robustness of the assay and its potential in estimating age in cetaceans. A successful DNAm-based age estimator has now also been developed for bottlenose dolphins, and in minke whales, the same age-associated sites identified in the humpback studies were investigated, finding similar relationships but at different CpG sites (Beal et al., 2019; Tanabe et al., 2020). Most recently, an epigenetic age estimation assay has been developed for Beluga whales, utilising a whole genome CpG array to determine novel differentially methylated CpGs, achieving a mean absolute error of 2.9 years in their age predictive model (Bors et al., 2020). These studies highlight the utility of epigenetic methods for age estimation in cetaceans due to their accuracy, reduced invasiveness, and low cost in comparison to previously applied methodologies.

Chapter 4 – Novel high-throughput epigenetic analysis workflow

4.1 Illumina sequencing platform potential

Recent breakthroughs with next-generation sequencing (NGS) platforms have facilitated a rapid expansion of ecologically focused genetic studies (Beng & Corlett, 2020). Large increases in sequencing ability and decreases in costs for sequencing have allowed for tools initially reserved for human and model organism study to be applied to other fields (Wang, Srivathsan, Foo, Yamane, & Meier, 2018; Young & Linacre, 2020). These have included large scale environmental DNA analysis and now the utilisation of epigenetic analyses, as outlined in chapters 2 and 3.

A promising development for utilising NGS platforms such as Illumina for ecological studies is the use of a fusion tag workflow for short amplicon analyses. In this approach, samples can be multiplexed together into large libraries for massively parallel sequencing utilising a single round of PCR amplification (Stat et al., 2017, 2019). This is a cost and time-effective high-throughput alternative to traditional library preparation workflows, which often require ligation or 2-step PCR approaches (Aird et al., 2011; Wang et al., 2018). This workflow has been pioneered by the Trace and Environmental DNA (TrEnD) laboratory at Curtin University for use predominately with ancient and environmental DNA analyses across a range of ecological applications and substrates (Koziol et al., 2019; Seersholm et al., 2020; Stat et al., 2017).

4.2 Fusion tagging architecture

The fusion tagging workflow utilises the addition of unique 'barcode' tags, called fusion tags, to the primers used for amplicon sequencing of targeted sites. The barcoded primers, which also contain Illumina sequencing and platform adaptors, allow for a single round of PCR amplification of samples which can then be multiplexed together onto a single sequencing run. The fusion tags can then be used to demultiplex sequencing results, matching unique sequences with individual samples through bioinformatic pipelines. This allows for rapid amplification, pooling, and sequencing of hundreds of samples without the need for a second round of PCRs and/or

ligation of barcodes and adaptors, a traditional approach for multiplexing samples that can be more costly and increase biases (Aird et al., 2011; Seguin-Orlando et al., 2013).

This thesis presents the first known adaptation of this workflow to epigenetic analyses, and a novel methodology for assessing site specific DNAm levels. By adapting fusion tags to PCR primers designed to amplify genomic regions with DNAm associations with age, a high-throughput approach to multiplex and rapidly sequence samples for epigenetic analysis is achieved and detailed below.

4.3 Methods

4.3.1 Sample collection and characterisation

I developed this epigenetic age estimation workflow through the analysis of tissue samples from three different species of whale: sperm whales (*Physeter macrocephalus*), southern right whales (*Eubalaena australis*), and long-finned pilot whales (*Globicephala melas*). The selection of these species was opportunistic based upon available tissue samples.

For sperm whales, skin tissue samples were obtained from beached, deceased whales at three sites from around Tasmania, Australia, by the Tasmanian Government's Department of Primary Industries, Parks, Water and Environment staff. Whale characteristics including sex and size were recorded *in situ* and teeth and skin biopsy samples were collected, with tissues stored in 70% ethanol and frozen at -18°C. A total of 26 samples were obtained; 21 samples (16 individuals) from Ocean Beach, Strahan in 2004; four samples from Perkins Island in 2009; and one sample from Croppies Beach, Waterhouse in 2002.

For southern right whales, 32 skin samples were biopsied at sea from 16 individual wild animals during a long-term monitoring project. Their characteristics were determined and DNA extracted as described in the original research conducted by Carroll et al., 2016, before DNA aliquots were transferred to the TrEnD facilities at Curtin University in 2019.

Long-finned pilot whale (herein referred to as pilot whale) skin tissue samples were obtained from beached, deceased whales at two sites: from Marion Bay, Tasmania,

Australia, by the Tasmanian Government's Department of Primary Industries, Parks, Water and Environment staff; and from Hamelin Bay, Western Australia, Australia, by the Department of Biodiversity, Conservation, and Attractions staff. Whale characteristics including sex and size were recorded *in situ* and teeth and skin biopsy samples were collected, with tissues stored in 70% ethanol and frozen at -18°C. A total of 140 samples were collected from 139 individuals: 96 skin samples (95 individuals) were collected from the mass stranding event of 145 whales in Marion Bay in 2005, and 44 samples were collected from the stranding event of roughly 150 pilot whales at Hamelin Bay in 2018.

4.3.2 Age estimation of whales

Age estimation for use as calibration 'known age' was achieved for sperm and pilot whales using tooth GLG analysis. This was performed by two separate commercial service providers using industry standard methods for the 20 sperm whale and 75 Marion Bay pilot whale samples (Bloch, Lockyer, & Zachariassen, 1993). Nine Hamelin Bay pilot whale samples had their age estimated through similar GLG estimations by an independent collaborator. Age estimations made in this way have a degree of error. Based on previously published accuracy estimates, the estimated age likely varied from true age by up to two years standard deviation within sperm whales, and within a likely standard deviation of one year in pilot whales (Bloch et al., 1993; Lockyer, 1993).

For southern right whales, age was determined through a photographic and sampling recapture program starting from the birth of each individual for 16 individual whales, as described in Carroll et al., 2016. A total of 32 known age tissue samples were obtained utilising biopsy darting for the 16 individuals. Age estimates had a standard deviation of less than one year, as an almost exact age for each sample was known. All age estimations are recorded in Table App A 1.

4.3.3 Sample processing and DNA extraction

Sperm whale and pilot whale samples were removed from ethanol storage buffer and black epidermal tissue was sub-sampled, separating it from blubber. The epidermis was then dried and weighed. Weights of subsamples were between 1.0 and 65 mg, dependant on available tissue, with a target weight of 20 mg (DNA extraction kit

recommendation). Extraction of DNA from the subsampled tissue was automated using the QIAcube extraction platform (Qiagen, Germany) and the DNeasy Blood and Tissue Kit (Qiagen) with standard protocols. Extracted DNA was then quantified using the QIAxpert system (Qiagen). DNA extracts were stored at -18°C. For southern right whales, DNA was extracted as outlined in Carroll et al., 2016, with DNA aliquots transferred to the TrEnD laboratory and stored at -18°C.

4.3.4 Bisulfite conversion

All samples were bisulfite-converted using the EZ DNA Methylation-Lightning™ Kit (Zymo, USA) with standard protocols and 200 ng of DNA input per sample. Expected conversion efficiency was >99.5% with DNA recovery of >80%. Bisulfite-converted DNA (bcDNA) was frozen at -18°C for no more than four weeks before PCR amplification.

4.3.5 PCR design and amplification

Target genomic regions shown to have age related CpG variation were identified from the literature, specifically from model organism studies and a humpback whale age estimation study (Table 4.1). Primers were developed for these age-related genes by adapting primers used in previous studies to a sperm whale scaffold genome (no pilot whale or southern right whale genome existed at the time of writing) present on NCBI (Arnason, Gullberg, Gretarsdottir, Ursing, & Janke, 2000; T. M. Lowe & Eddy, 1997). *In silico* tests using Geneious R10.1 (Kearse et al., 2012) identified the binding of primers to target gene regulatory 5' regions on the sperm whale scaffold (transformed *in silico* to a bcDNA state), and primer sequences were adjusted to reduce mismatches where required. A total of six primer sets were utilised and optimised (Table 4.1).

Table 4.1 Bisulfite-converted DNA specific primers for amplification of target age-associated gene regions in cetaceans.

Gene	Direction (5' – 3')	Sequence (5' – 3')	Product size (bp)	Optimum T_A (°C)	References of loci
PDE4C	Forward	TTTGTGGAAGGAG GAGGGTTAGGAG	177	55	(Freire-Aradas et al., 2016; Weidner et al., 2014)
	Reverse	CCAAAACATAACC AAAACTC			
KLF14	Forward	GGTAGGTTGTTTAG AAGTTA	115	52	(Hannum et al., 2013; Spólnicka et al., 2018)
	Reverse	CATTTTTAACAACC TAAAAAATTATCTT ATC			
CCDC105	Forward	GGTGGTAGTATTAT GGAGAAGTTAT	295	52	(Galamb et al., 2016; Horvath, 2013)
	Reverse	TCTACTAATTAATA AACAAAACCTTAC			
TET2	Forward	GTGGTTAAAGTAA ATAGAAGGT	138	53	(Grönniger et al., 2010; Polanowski et al., 2014)
	Reverse	CAAAAACACTCCC CWATTTC			
18S	Forward	GATTGAGTAATAA TAGGTTTGTGATG	227	53	(Wang & Lemos, 2019)
	Reverse	GGTGTGTATAAAG GGTAGGGA			
CDKN2A	Forward	AATGATTTTTGGTA AAGGGGAGAT	121	54	(Gonzalez-Zulueta et al., 1995; Horvath, 2013; Koch et al., 2011; Maegawa et al., 2010; Polanowski et al., 2014)
	Reverse	CCCCATATACTTTT CAATCCTCC			

T_A represents annealing temperature. References indicate studies identifying age related DNAm correlation in each specific locus.

To allow multiplexing of samples and for sequencing on the Illumina Miseq platform (Illumina, USA), specialised fusion-tagged primers were developed, consisting of Illumina sequencing adaptors for 5' binding to the Miseq flow cell, 5' binding regions for Illumina sequencing primers, a unique multiple identifier sequence (MID) tag for demultiplexing, and the gene-specific bcDNA targeted primer. These fusion primers were used for all subsequent steps.

Using pilot and sperm whale bcDNA as a template, temperature gradient quantitative PCRs (qPCR) were performed on a StepOne Plus (Applied Biosystems, USA) real-time qPCR instrument to identify optimum annealing temperatures (T_A) for each primer set (Table 4.1). All qPCRs consisted of 25 µl total volume containing the following: 2.5 mM/L $MgCl^2$ (Applied Biosystems), 1X PCR Gold buffer (Applied Biosystems), 0.25 mM/L dNTPs (Astral Scientific, Australia), 0.4 mg/ml BSA (Fisher Biotec, Australia), 0.4 µmol/L forward and reverse primer, 0.6 µl of a 1:10,000 solution of SYBR® Green dye (Life Technologies, USA), 1U AmpliTaq Gold® DNA polymerase (Applied Biosystems), and made up to volume with Ultrapure™ Distilled Water (Life Technologies). A clean, DNA-extract free laboratory was used during PCR setup, and 2µL of bcDNA was added to each reaction in a separate laboratory. The amplification of desired regions was confirmed using gel electrophoresis and amplicon length comparison to their expected lengths.

All bisulfite-converted samples were amplified across a selection of assays through qPCR using fusion-tagged primers, with thermocycler conditions as follows: 95°C for 5 minutes followed by 50 cycles of 95°C for 30 seconds, assay specific annealing temperature (Table 4.1) for 30 seconds, 72°C for 45 seconds, and a final 10 minutes at 72°C.

Sperm whale samples were amplified using *PDE4C*, *CCDC105*, and *KLF14* assays, while pilot whales were amplified across *18S*, *CDKN2A*, and *TET2*. Southern right whales were sequenced across all six assays. A secondary set of pilot whale samples were amplified and sequenced utilising *CCDC105* and *KLF14* assays for the determination of inter- and intra-run precision, with results presented in chapter 6. These assay variations between species were due to a combination of the availability of assays, samples, success of un-tagged PCRs on different species, and resources for sequencing.

4.3.6 Amplicon library preparation and sequencing

Amplified samples were pooled together into equimolar libraries using qPCR endpoints and quantification on a QIAxcel Advanced system (Qiagen) before being size-selected (based on target amplicon lengths) using a Pippin Prep instrument (Sage Sciences, USA) and cleaned with a QIAquick PCR Purification Kit (Qiagen). Final libraries were quantified using the Invitrogen™ Qubit™ 4 Fluorometer (Thermofischer, Australia) before sequencing on the Illumina MiSeq platform using either 300 cycle V2 or V2 Nano reagent kits (Illumina).

4.3.7 Data Processing and statistical analysis

The Obitools package (Boyer et al., 2016) was used at default settings in a bioinformatic pipeline to filter out lower-quality sequence reads with an average Q score less than 30, with reads then separated by barcode into sample files, and sequencing adaptors, gene-specific primers, and MID tags removed. Resultant sample files were then translated into FASTA format using a custom R (R Core Team, 2019) script.

A custom Python (Python Core Team, 2019) script was used to determine a consensus sequence length in each FASTA file and remove all non-consensus length reads to reduce the impact of homopolymer runs disrupting sequence alignments. The script then identified all CpG sites and percentage methylation was determined through the comparison of percentage of cytosine and thymine present at each site. To confirm CpG sites used in further analyses, each gene's sample FASTA files were also visually inspected in Jalview (Waterhouse, Procter, Martin, Clamp, & Barton, 2009).

Each sample's CpG site methylation levels were assessed against their corresponding estimated whale age, with significant correlations between age and methylation percentage identified using StatistiXL v2 (Robert & Withers, 2007). For each gene with a significant result (or the closest site to a significant result of $p < 0.05$), the CpG site with the strongest correlation was added to the multiple linear regression model, operating through another custom R (R Core Team, 2019) script. Only one CpG site per assay was utilised in multiple linear regressions. This was due to concerted methylation occurring across genomic regions, resulting in a lack of true biological replication as well as the compounding of experimental error if using multiple CpG

sites from a single locus (Grönniger et al., 2010; Koch & Wagner, 2011). This approach aligns with strategies undertaken in similar studies (Bocklandt et al., 2011; Koch & Wagner, 2011; Polanowski et al., 2014). The multiple linear regression model with the best predictive model using various combinations of sites was determined using scores derived from the Akaike information criterion (AIC) and Bayesian information criterion (BIC), but alternate models are also presented (Table App A 2).

Chapter 5 – Homology-based epigenetic age estimation of sperm, southern right, and pilot whales

5.1 Homology-based epigenetic analyses

The novel Illumina workflow for epigenetic analysis detailed in chapter 4 utilises sites identified by homology from previous age estimation studies, predominately in cetaceans. In closely related taxa, previous studies have shown that DNAm age-associations are often conserved, though the exact CpG sites that are age associated may vary (Polanowski et al., 2014; Tanabe et al., 2020; Wright et al., 2018). By assessing these genomic regions through targeted amplicon sequencing, age associations can be rapidly tested and assessed. There is therefore no requirement for novel identification of genomic regions with age-associations, and epigenetic analyses can be initiated as soon as primers are validated to amplify the targeted region. This precludes the need for methylation arrays or whole genome bisulfite sequencing, both of which can be costly or require specialist equipment (Bors et al., 2020; Olova et al., 2018). Utilising basic PCR primers on bisulfite converted DNA extracts, the assays developed in chapter 4 were amplified and sequenced successfully for all three cetacean species without the need for further sample preparation or primer variation, demonstrating the cross-species potential for primer sets.

5.2 Sperm, southern right, and pilot whale background

Sperm whales (*Physeter macrocephalus*) are the largest odontocete (toothed whale), found in all ice-free waters around the globe (Giorli & Goetz, 2019; Whitehead, 2002). Sperm whale numbers have been significantly impacted by whaling, with populations recovering slowly due to a combination of late sexual maturity attainment, slow growth rates, and low fecundity (Evans & Hindell, 2004). The determination of life-history data of sperm whales, especially age structures, has been identified as a key metric in understanding their recovery from whaling and the dynamics of current extant populations (Evans & Hindell, 2004; Ohsumi, 1966; Whitehead, 2002). Owing to their pelagic nature and the little information on their migratory routes, traditional approaches for determining life history and population characteristics of sperm whales,

such as visual surveys, have been difficult, especially for live animals (Evans & Hindell, 2004).

Southern right whales (*Eubalaena australis*) are baleen whales (mysticetes) found in the Southern hemisphere with a circumpolar distribution (Fretwell, Staniland, & Forcada, 2014; Richards, 2009). They were heavily hunted over the last 300 years until 1973, with roughly 150,000 whales processed and a reduction in population size from about 55,000–70,000 individuals pre-whaling to a low of 300 animals in the 1920s (Greig, Secchi, Zerbini, & Dalla Rosa, 2020; Fretwell et al., 2014; Jackson, Patenaude, Carroll, & Baker, 2008). Their numbers have now increased to over 7,500 individuals and are generally rising, although current populations are still fragmented and vulnerable (Fretwell et al., 2014; Stamation, Watson, Moloney, Charlton, & Bannister, 2020). Assessment of population characteristics for right whales, as discussed above for sperm whales, is vital for monitoring post-whaling recovery and informing on conservation management policies (Hamilton et al., 1998). For assessing the age of live animals, however, photographic identification and visual surveys remain some of the only tools currently available. With these animals' lifespans potentially far greater than 65 years, a viable alternative in determining population age structures and guiding conservation strategies would be of benefit (Carroll et al., 2016; Hamilton et al., 1998).

Long-finned pilot whales (*Globicephala melas*) are a widespread species of medium-sized odontocete, with extant populations in temperate waters of the North Atlantic region, ranging from the East coast of North America to the Mediterranean, and in the Southern Hemisphere in cold temperate waters (Fullard et al., 2000; Pike et al., 2019) (Figure 4.1). They are social animals, swimming in matrilineally-structured pods of up to 1000 individuals, and have estimated average lifespans of at least 35 to 45 years for males and exceeding 60 years for females (Amos et al., 1993; Bloch et al., 1993). Pilot whales are well-known as a species involved in mass-strandings (Oremus, Gales, Kettles, & Baker, 2013). Entire pods numbering in the hundreds have beached themselves on various coastlines worldwide, and they are the most common cetacean to be involved in stranding events in regions such as Tasmania, Australia (Beasley et al., 2019; Martin et al., 1987; Oremus et al., 2013). Investigations into age structures within pilot whale populations are currently restricted to post-mortem analysis, due both to their comparatively large population size and lack of unique external visual features (Martin et al., 1987). Determining age demographics within live pilot whale

populations could lead to improved determination of factors behind beaching events, a still poorly understood phenomenon, as well as informing on a range of population health indicators (Amos et al., 1993; Borrell, Bloch, & Desportes, 1995; Oremus et al., 2013; Soto, Grandi, Garcia, Crespo, & Dans Silvana, 2017).

DNAm analysis may offer a robust method to non-invasively age all three of these whale species. It could be effectively applied to live, wild populations, as well as to stranded whales to increase the speed and potentially the accuracy of age estimation (Polanowski et al., 2014). If successfully applied, epigenetic approaches could allow for large scale comparative studies of age demographics in sperm and southern right whales and also enable the study of age-related links to stranding events for pilot whales. These potential outcomes, and other concerns raised within chapter 2, highlight the need for molecular tools for the age estimation of wild sperm, southern right, and pilot whale populations. Here, I develop a homology-based epigenetic age estimator for both southern right and sperm whales utilising the novel high-throughput sequencing approach detailed in chapter 4. I also attempt the development of an age estimation assay for long-finned pilot whales.

5.3 Methods

The methodology for all three whale species' epigenetic age-associated assessments are detailed in chapter 4. The only variation in methodology between species was assay selection, sample collection, known age estimation, and DNA extraction approaches. These are all detailed within chapter 4. Specific methodology relating to each species' analysed assays and CpG site determinations are detailed below, as are the statistical tests applied.

5.3.1 Sperm whales

The bcDNA of 20 known-age sperm whales was analysed over three assay regions: *KLF14* (n = 20), *CCDC105* (n = 20), and *PDE4C* (n = 10). Only the sites with the strongest correlations from each assay were assessed in detail. An ANCOVA (analysis of covariance) assessment was utilised to test for sex-specific relationships within the regressions.

A subset of 10 samples were used for the development of a multiple linear regression (MLR) model of all three sites. This was due to the requirement that each assay needed to have representative samples present for an MLR to be produced. This necessitated the removal of 10 samples from both the *KLF14* and *CCDC105* datasets, which numbered 20 samples initially.

The three chosen assays were assessed in a series of MLR models, with only one CpG site (the strongest age to DNAm correlated site) from each assay used, as detailed in chapter 4. All possible combinations of sites were investigated in MLRs, and both AIC and BIC selection criteria calculated. Both were considered to determine which combinations of two or more sites was best at predicting age (Table App A 2).

A Leave One Out Cross Validation (LOOCV) analysis was used to validate the model and estimate the mean average difference between known and predicted age, and a Shapiro-Wilk normality test on the residuals was utilised to determine normality, and testing indicated a normal distribution ($W = 0.9111$, $p = 0.2888$).

5.3.2 Southern right whales

The bcDNA of 32 known-age southern right whale samples from 16 individuals was analysed over six assay regions: *KLF14* (n = 32), *18S* (n = 32), *CDKN2A* (n = 32), *CCDC105* (n = 23), *TET2* (n = 31), and *PDE4C* (n = 30). An ANCOVA assessment was utilised to test for sex-specific relationships within the regressions.

A subset of 22 samples was used for the development of a MLR model of all six assays, as each assay needed to have representative samples present and various samples failed during PCR and sequencing from the *CCDC105*, *TET2*, and *PDE4C* assays.

Within the subset of 22 samples, the six assays were used in a series of MLR models, with only one CpG site (the strongest age to DNAm correlated site) per assay included. All possible combinations of sites were investigated in MLRs, and both an AIC and BIC score set were developed, determining which combinations of two or more sites was best at predicting age (Table App A 2).

A LOOCV analysis was used to validate the model and estimate the mean average difference between known and predicted age, and a Shapiro-Wilk normality test on the residuals was utilised to determine normality and indicated a non-normal distribution

(W = 0.8240, p = 0.00122), an expected issue with the age range skewed heavily to younger samples, as can be seen in Figure 7.2 (chapter 7).

5.3.3 Pilot whales

A selection of three loci were investigated in pilot whales: *18S*, *TET2*, and *CDKN2A*. Two other genomic regions were analysed, *ELOVL2* and *GRIA2*, but both failed to amplify after duplicated attempts across three samples using a gradient of annealing temperatures between 48-60 ℃. This was likely due to genomic variation between the pilot whale genes and the sperm whale scaffold used for *in silico* primer design. Within each gene region analysed, all CpG sites were identified and DNAm percentages determined within a calibration set of 84 bcDNA pilot whale samples. An ANCOVA assessment was utilised to test for sex-specific relationships within the regressions.

The three analysed assays were used in a series of MLR models, with one CpG site (the strongest age to DNAm correlated site) selected from each gene region for inclusion in the analyses. All possible combinations of sites were investigated in MLRs, and both an AIC and BIC score set were developed, determining which combination of two or more sites was best at predicting age (Table App A 2).

A LOOCV analysis was used to validate the model and estimate the mean average difference between known and predicted age, and a Shapiro-Wilk normality test on the residuals was utilised to determine normality, which indicated a non-normal distribution (W = 0.9520, p = 0.0034).

5.4 Cetacean epigenetic age estimation model results

5.4.1 Sperm whale homology-based epigenetic age assessment results

For sperm whales significantly correlated hypermethylation with increasing age was identified in CpG sites in all assays. The weakest association was *CCDC105* site 153 (n = 20, R^2 = 0.2204, p = 0.0368, Figure 5.1), followed by *KLF14* site 5 (n = 20, R^2 = 0.2491, p = 0.0251, Figure 5.1). *PDE4C* site 129 showed the strongest relationship (n = 10, R^2 = 0.6220, p = 0.00671, Figure 5.2), although it had a limited dataset of 10 sequenced samples (10 samples failed to amplify during the PCR process).

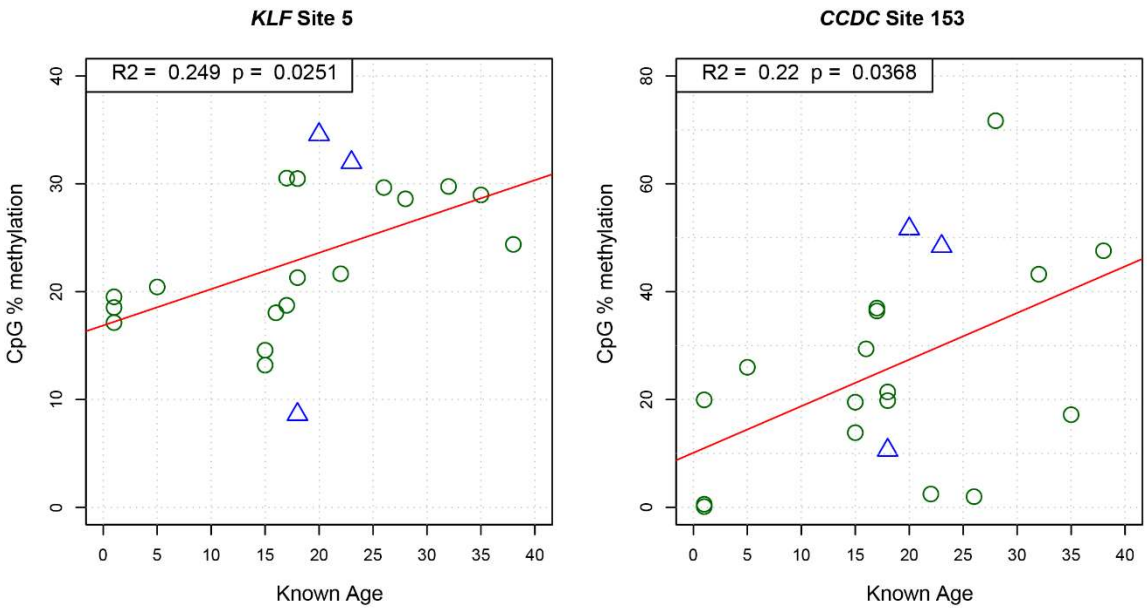

Figure 5.1 Sperm whale relationships between age and CpG site methylation of *KLF14* CpG site 5 and *CCDC105* CpG site 153 in the dataset (n = 20). Males are indicated by blue triangles, and females by green circles.

In the restricted subset of 10 samples, the statistical relationships were reduced in both the *KLF* (n = 10, R^2 = 0.2204, p = 0.4014) and *CCDC105* sites (n = 10, R^2 = 0.1034, p = 0.3649), with neither showing a significant relationship in the subset (Figure 5.2).

The multi-site model with the best R^2 and lowest scores (optimal) in both the AIC and BIC employed all three sites. No significant differences were found between male and female regressions (ANCOVA, $p > 0.5$ in all cases).

The selected multiple linear regression model below represents the final sperm whale age estimation assay and allows the accuracy of the model to be assessed (Figure 5.3 A). The equation for the sperm whale epigenetic age estimation model is: Age = 6.62 + (0.47 x *PDE4C* site 129 methylation) – (0.29 x *KLF14* site 5 methylation) + (0.22 x *CCDC105* site 153 methylation). This model indicated that approximately three quarters of the variation in age was driven by DNAm (R^2 = 0.7639). An adjusted R^2 value was used as it allows comparison to other models that employ different numbers of predictors, such as in the humpback whale age estimator, whose 3-predictor model showed an R^2 of 0.787, a very similar result to this sperm whale model (Polanowski et al., 2014). The y intercept of 4.786 indicated that younger whales will have their age slightly overestimated the younger they are. The slope of the regression line (Figure

5.3 A) indicates that whales older than roughly 25 years old may have their ages slightly underestimated, which will compound with increasing age.

The LOOCV showed a mean absolute difference between known and estimated age of 3.857 years with a standard deviation of 2.934 (Figure 5.3 B). Based on this calibration from 10 whales, the 95% confidence interval (CI) for predictions was 11.91 years.

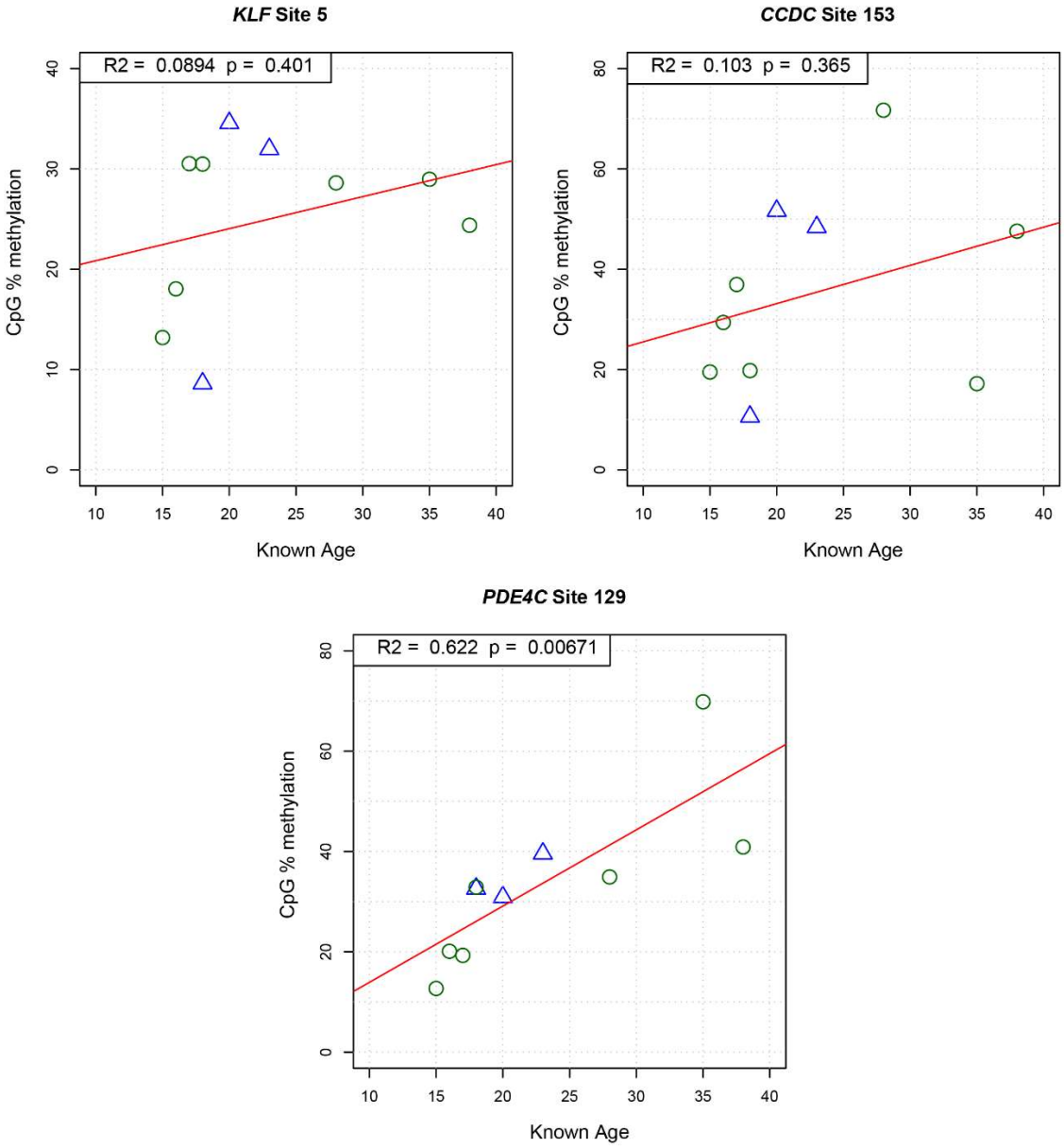

Figure 5.2 Sperm whale relationships between age and CpG site methylation of *KLF14* CpG site 5, *CCDC105* CpG site 153, and *PDE4C* CpG site 129 within the subset used in the multiple linear regression (n = 10). Males are indicated by blue triangles and females by green circles.

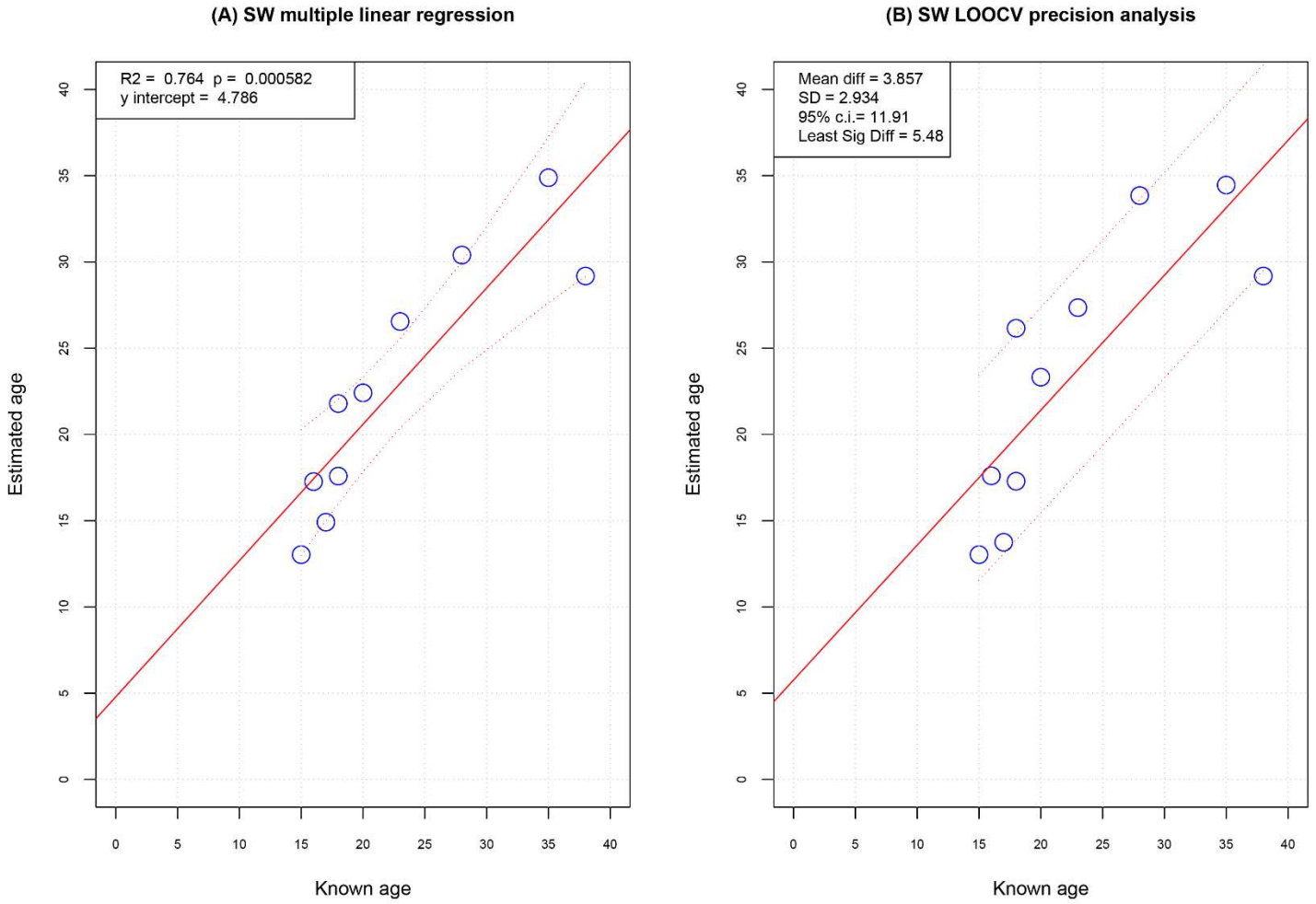

Figure 5.3 Multiple linear regression (A) and LOOCV (B) of known ages and estimated ages in sperm whales from methylation values at three CpG sites: *KLF14* site 5, *CCDC105* site 153, and *PDE4C* site 129 utilising n = 10 samples. 95% confidence limits of the regression line are shown with dashed red lines.

5.4.2 Southern right whale homology-based epigenetic age assessment results

Significantly correlated hyper- or hypomethylation with increasing age was identified in CpG sites in all assays besides *TET2*. *CDKN2A* site 38 showed the strongest relationship (n = 32, R^2 = 0.4216, $p \leq 0.0001$, Figure 5.4), followed by *CCDC105* site 164 (n = 23, R^2 = 0.2770, p = 0.0099, Figure 5.4), *PDE4C* site 47 (n = 30, R^2 = 0.2641, p = 0.0037, Figure 5.4), *KLF14* site 57 (n = 32, R^2 = 0.1966, p = 0.0110, Figure 5.4), *18S* site 27 (n = 32, R^2 = 0.1804, p = 0.0154, Figure 5.4), and finally the non-significant *TET2* site 74 (n = 31, R^2 = 0.0878, p = 0.1056, Figure 5.4).

Within the restricted subset of 22 samples, DNAm relationships changed slightly in all assays: *CDKN2A* site 38 was still the strongest relationship (n = 22, R^2 = 0.3344, p = 0.0048), followed by *PDE4C* site 47 (n = 22, R^2 = 0.3282, p = 0.0053), then *KLF14* site 57 (n = 22, R^2 = 0.2964, p = 0.0088), *CCDC105* site 164 (n = 22, R^2 = 0.2734, p = 0.0125), the now non-significant *18S* site 27 (n = 22, R^2 = 0.1778, p = 0.0506), and finally the previously non-significant *TET2* site 74 (n = 22, R^2 = 0.1643, p = 0.0613).

The multi-site model with the best R^2 and lowest scores (optimal) in both the AIC and BIC employed all six sites. No significant differences were found between male and female regressions (ANCOVA, $p > 0.5$ in all cases).

The selected multiple linear regression model below represents the final southern right whale age estimation assay and allows the accuracy of the model to be assessed (Figure 5.5 A). The equation for the southern right whale epigenetic age estimation model is: Age = 28.89 – (0.19 x *18S* site 27 methylation) – (47.33 x *PDE4C* site 47 methylation) + (0.02 x *TET2* site 74 methylation) – (0.11 x *CDKN2A* site 38 methylation) + (0.03 x *KLF14* site 57 methylation) + (0.04 x *CCDC105* site 164 methylation). The y intercept of 1.110 indicates that younger whales will have their age very slightly overestimated. The slope of the regression line (Figure 5.5 A) indicates that whales older than roughly 5 years old will have their ages slightly underestimated, with increasing underestimation as age increases.

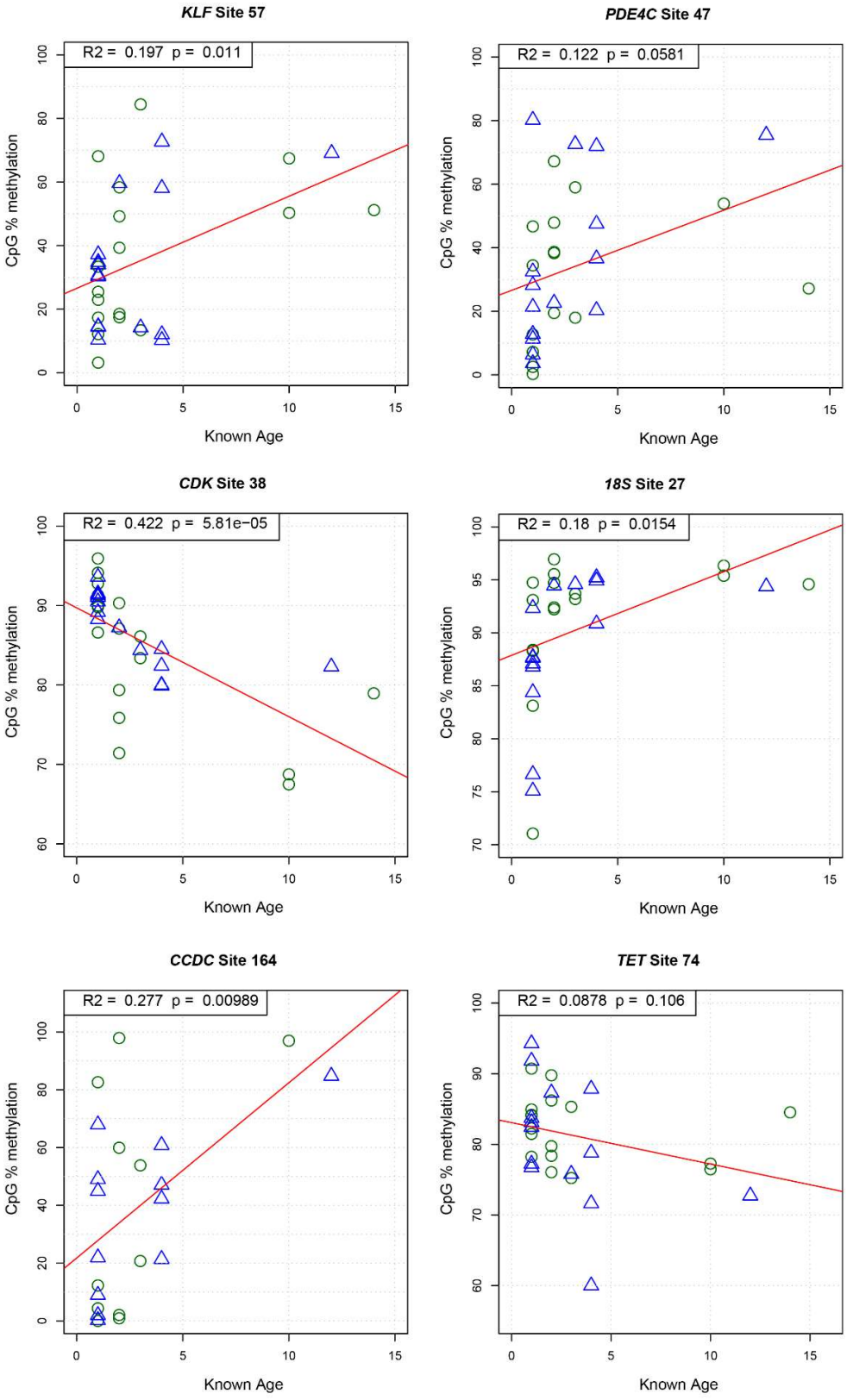

Figure 5.4 Southern right whale relationships between age and CpG site methylation of *KLF14* CpG site 57, *PDE4C* CpG site 47, *CDKN2A* CpG site 38, *18S* CpG site 27, *CCDC105* CpG site 164, and *TET2* CpG site 74, the assays used in the multiple linear regression. Males are indicated by blue triangles, and females by green circles.

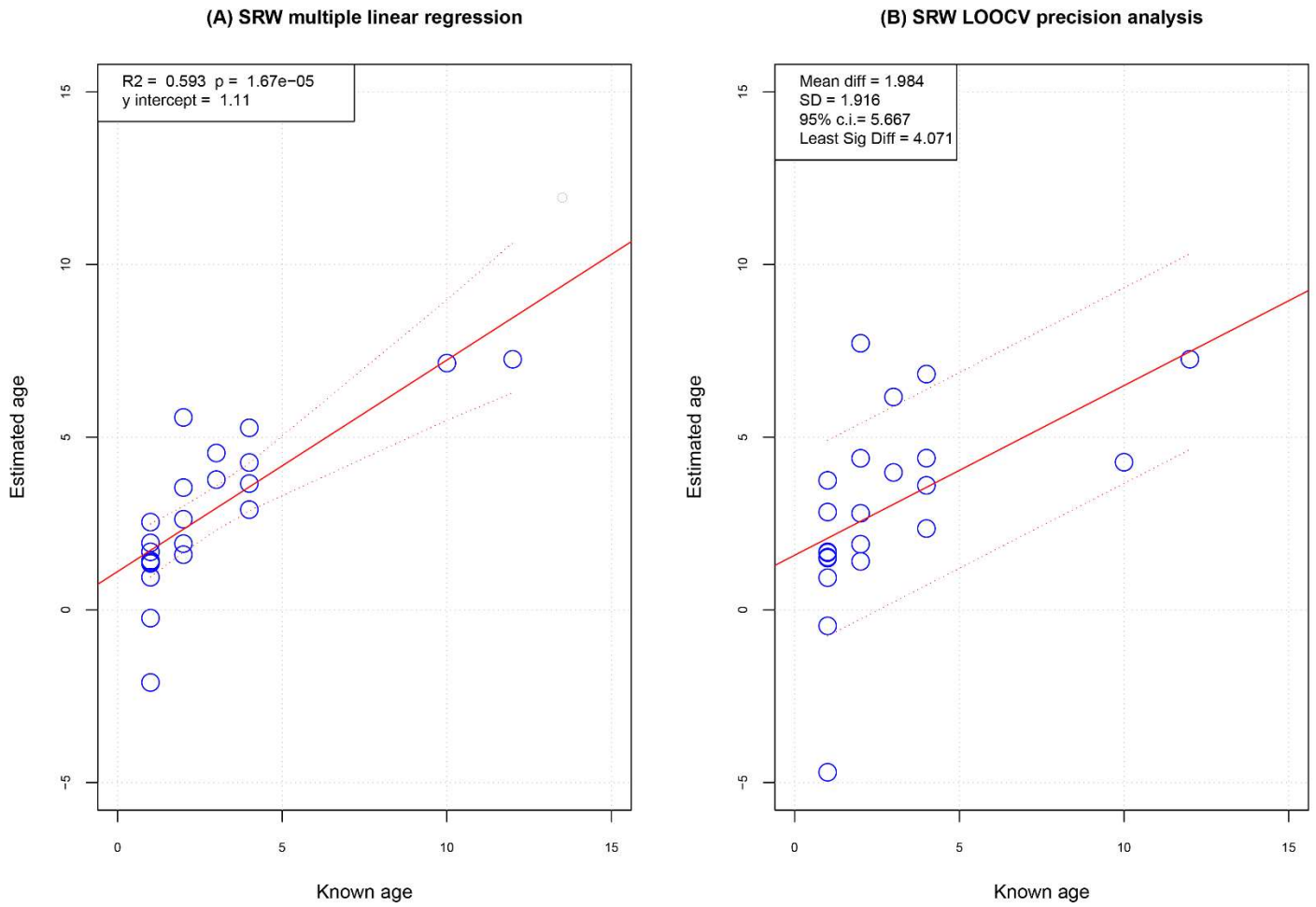

Figure 5.5 Multiple linear regression (A) and LOOCV (B) of known ages and estimated ages in southern right whales from methylation values at six CpG sites: *KLF14* site 57, *PDE4C* site 47, *CDKN2A* site 38, *18S* site 27, *CCDC105* site 164, and *TET2* site 74 in n = 22 samples. 95% confidence limits of the regression line are shown with dashed red lines.

The LOOCV analysis estimated mean average difference between known and predicted age was 1.984 years with a standard deviation of 1.916 (Figure 5.5 B). Based on the calibration dataset of 22 whales, the 95% confidence interval for predictions was 5.667 years.

5.4.3 Pilot whale homology-based epigenetic age assessment results

Across the three genes analysed in pilot whales (*18S*, *TET2*, and *CDKN2A*; n = 84), only the *CDKN2A* assay showed a significant correlation between CpG site methylation and age. The *CDKN2A* gene region showed the strongest correlation at site 53 (n = 84, R^2 = 0.0716, p = 0.0139, Figure 5.6). For the non-significant sites, the best correlations were found in the *TET2* assay at site 10 (n = 84, R^2 = 0.03147, p = 0.1064, Figure 5.6), and in *18S* at site 135 (n = 84, R^2 = 0.0245, p = 0.1555, Figure 5.6).

The multi-site model with the best (lowest) scores in both the AIC and BIC and strongest regression coefficient employed all three sites (Figure 5.6). No significant differences were found between male and female regressions (ANCOVA, $p > 0.5$ in all cases).

The selected multiple linear regression model below represents the best pilot whale age estimation assay (Figure 5.7 A). The equation for the pilot whale epigenetic age estimation model is: Age = 22.75 – (0.05 x *18S* site 135 methylation) + (0.01 x *TET2* site 10 methylation) – (0.1 x *CDKN2A* site 53 methylation). The adjusted R^2 value of 0.0622 indicated that only about 6% of the age variation is driven by DNAm in the targeted CpG sites, with almost 94% due to other, unknown factors. The y intercept of 14.23 indicated that younger whales will have their age heavily overestimated. The slope of the regression line (Figure 5.7 A) also indicates that whales younger than roughly 15 years old will have their ages heavily overestimated, and those above 15 years old will be heavily underestimated. This model is therefore not suitable for predictive purposes.

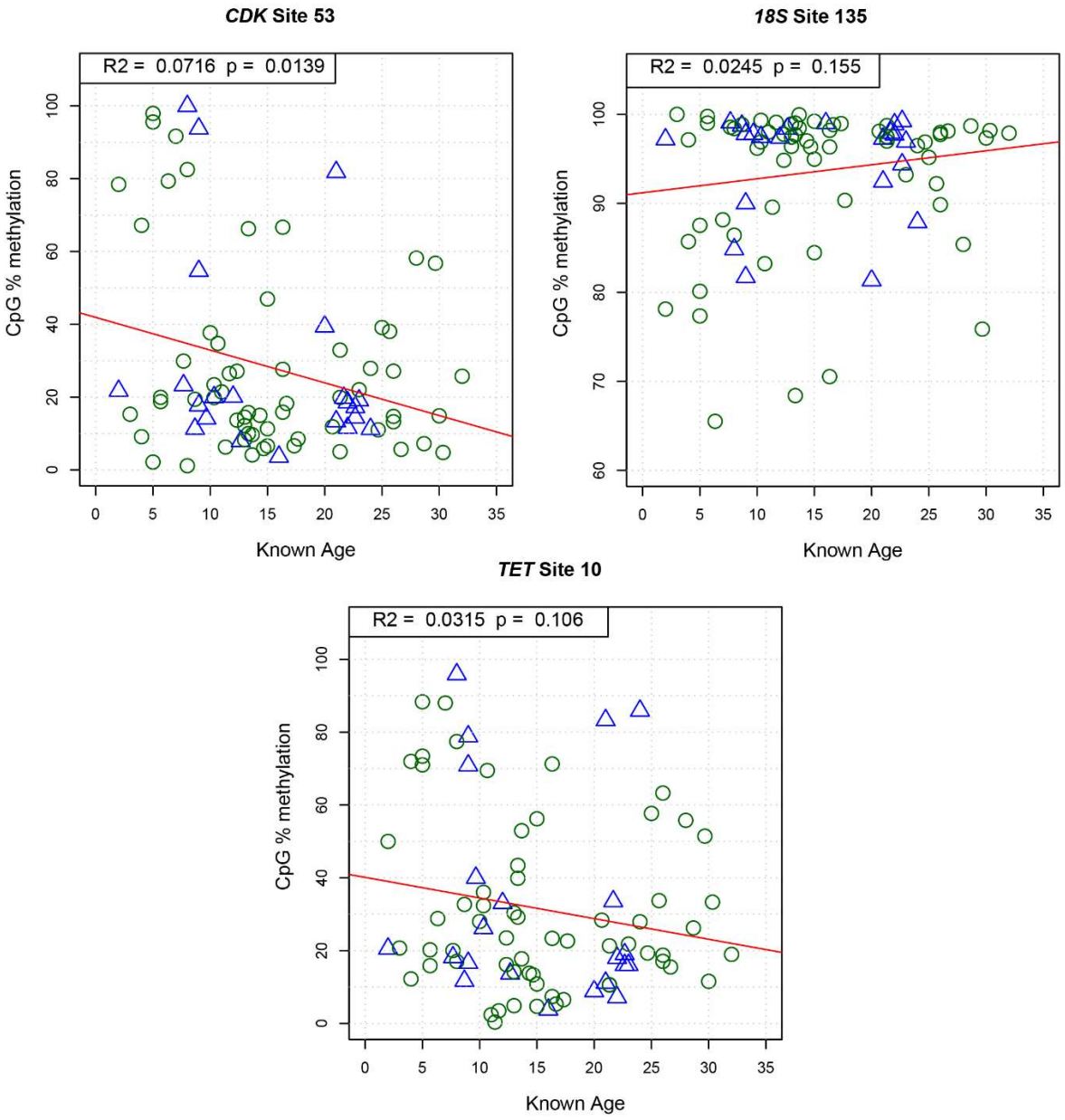

Figure 5.6 Pilot whale relationships between age and CpG site methylation of *CDKN2A* CpG site 53, *18S* CpG site 135, and *TET2* CpG site 10, the assays used in the multiple linear regression. Methylation was determined in n = 84 whales. Males are indicated by blue triangles, and females by green circles.

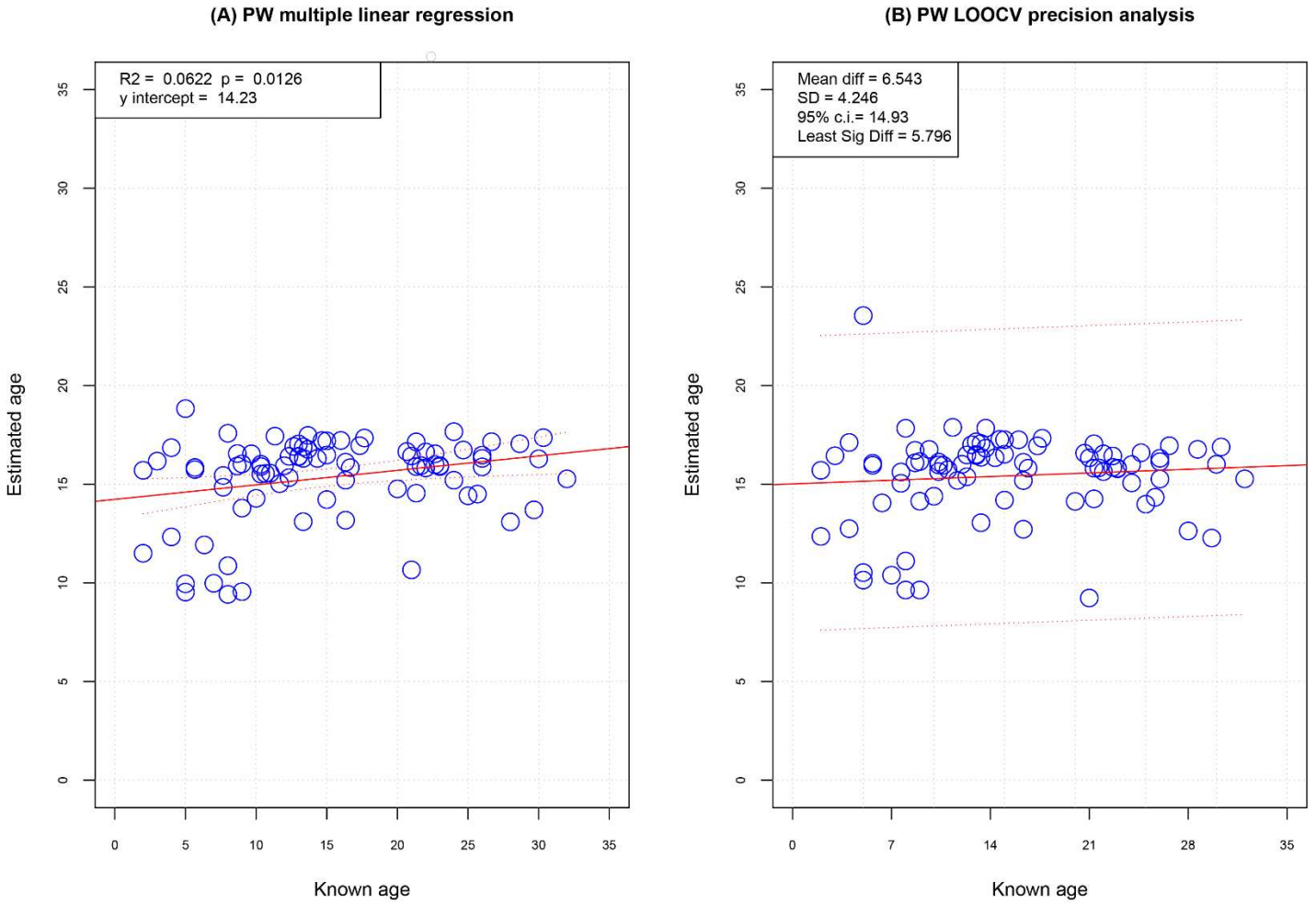

Figure 5.7 Multiple linear regression (A) and LOOCV (B) of known ages and estimated ages in pilot whales from methylation values at three CpG sites, *CDKN2A* site 53, *18S* site 135, and *TET2* site 10 in n = 84 samples. 95% confidence limits of the regression line are shown with dashed red lines.

The LOOCV analysis indicated the mean average difference between known and predicted age to be 6.543 years with a standard deviation of 4.246 (Figure 5.7 B). Based on the calibration from 84 whales, the 95% confidence interval for predictions was 14.93 years.

5.5 Discussion and future research perspective

5.5.1 Comparison of homologous loci age and DNAm relationships

For both southern right and sperm whales, I was able to develop an age estimation model with comparative accuracy to previous studies. This was achieved through the utilisation of a homology-based epigenetic analysis approach, using age-associated genes that have been assessed in other cetacean and model species. This research focused on loci homologous to those already identified as having age associations with DNAm within humpback whales, as well as those investigated in humans and mice (Table 4.1). In the *TET2* region in southern right whales, a negative correlation was found between age and DNAm at the site utilised in these models, a relationship supported by the hypomethylation observed with increasing age at the site utilised in age estimation in humpback whales (Polanowski et al., 2014). A similar relationship was also found within the *CDKN2A* site assessed in southern right whales. This relationship differed, however, from that identified within the humpback whale study as well as in human and mice studies, which showed relationships between hypermethylation and increasing age in *CDKN2A* (Gonzalez-Zulueta et al., 1995; Maegawa et al., 2010; Polanowski et al., 2014). This indicated potential variations in either the CpG site analysed or some variation of the epigenetic mechanisms between the *CDKN2A* gene between species. Relationships between hypermethylation and increasing age were found in the *KLF14* sites in both the sperm and southern right whale analyses, which have also been identified as having a positive correlation between DNAm and age in humans (Zbieć-Piekarska et al., 2015). In addition, a significant relationship was found between the *PDE4C* site hypermethylation and increasing age in both whale species, a genomic region that has also shown to have a positive relationship between DNAm and age in humans (Bekaert et al., 2015; Weidner et al., 2014). Finally, both the *CCDC105* and *18S* sites assessed in southern right whales, and the *CCDC105* site assessed in sperm whales, exhibited a positive

correlation between hypermethylation and age, a relationship which has also been supported by previous studies (Galamb et al., 2016; Wang & Lemos, 2019). This demonstrates that, for these species and potentially other cetaceans, age associations are likely conserved in homologous genomic regions. Therefore, for these conserved regions, age estimation analyses can be adapted rapidly for a new species without the need for novel determination of age-related genomic regions. The determination of exact age-related CpG sites within these regions will likely still be required, as the sites with significant age associations discovered in this study varied from those identified in previous cetacean age estimation studies (Beal et al., 2019; Polanowski et al., 2014; Tanabe et al., 2020). Further evidence for this is provided by the fact that different age associated CpG sites were identified in the same genomic regions between southern right and sperm whales, with the strongest correlated sites varying in the *PDE4C*, *CCDC105*, and *KLF14* assays.

5.5.2 DNAm variation within sites

With regards to total DNAm variation at specific sites, in sperm whales DNAm variation was between 0 and 40% for *KLF15* and between 0 to 80% for *PDE4C* and *CCDC105*. In southern right whales, *KLF15*, *PDE4C,* and *CCDC105* all showed large DNAm differences across the sample subsets used for the final model, with DNAm variation between samples of up to 100% at each gene's selected CpG site. The southern right whale *18S* site varied by about 30%, with CpG site methylation between 70 and 100%, and *CDKN2A* and *TET2* varied by roughly 40%, between 60 and 100% DNAm. These all indicated quite large total CpG site methylation variation. In comparison, two of the three markers used for the humpback whale age estimation study showed less than 3% total variation in DNAm across all samples (Polanowski et al., 2014). Repeat sequencing, increased sample numbers, and the reduction of methodological error to further investigate sites with limited total DNAm methylation variation could potentially lead to the identification of additional age associated CpG sites within the assessed loci.

5.5.3 Final cetacean age predictor comparisons

The final sperm whale age estimation model demonstrated a relatively strong fit for the calibration dataset, comparable with that of the humpback age estimator (adj. R^2 value of 0.787), though lower than that of the beluga epigenetic clock, which achieved

a R^2 value of 0.92 (Bors et al., 2020; Polanowski et al., 2014). For the southern right whale model, the fit was poorer, an expected result with weaker individual site relationships than those in sperm whales. The LOOCV analyses assessed the mean absolute error, allowed for the comparison of confidence intervals to lifespan, and were used as a measure of utility of the model for age estimation. In these analyses, the sperm whale model performed comparatively poorly, while the southern right whale performed comparatively well. The mean absolute error for the sperm whale model was almost a year more than both the humpback and beluga whale studies, with 2.991 and 2.9 years respectively, while the southern right whale achieved a low mean absolute error of less than two years (Bors et al., 2020; Polanowski et al., 2014). This indicates that age estimates may be more accurate for southern right whales, while less accurate for sperm whales. When considering 95% confidence values, as an estimate for age estimation accuracy when considering species lifespan, the same trend is visible. For the southern right whale model, a CI/lifespan result of roughly 8.8% (5.7 CI/65 years) of the whale's conservatively estimated lifespan was observed, based on the longevity identified in the closely related Northern right whale (Hamilton et al., 1998). This CI/lifespan ratio is comparable to that of 9% reported for the humpback whale age estimation assay (Polanowski et al., 2014). While the southern right whale result is comparable to previous studies, the sperm whale had a 95% confidence interval of roughly 17% (12 CI/70 years) of a sperm whale's lifespan, which is nearly double the CI/lifespan ratio reported for the humpback whale epigenetic age estimator (Evans & Hindell, 2004; Ohsumi, 1966; Polanowski et al., 2014). This was an expected result given that this age predictor had less strongly correlated CpG sites and a limited number of samples in the calibration dataset (Figure 7.3). Such a large confidence interval means that the utility of the model would be limited to generalised age estimations. While the southern right whale model performed well regarding CI and mean absolute error, this is also likely due to the limited age range and positively skewed animal dataset. While this allows for an accurate model for ages up to 12, it may have reduced accuracy for estimates outside of this age range, which includes most of the adult southern right whale population (Best, Brandão, & Butterworth, 2020). Finally, the pilot whale model demonstrated a 95% confidence interval of 25% (15 CI/60 years) of a pilot whale's lifespan, which is a far greater CI compared to that of other models within this and previous research, an expected result given that my

age predictors had CpG sites with poor and non-significant correlations (Bloch et al., 1993; Polanowski et al., 2014).

5.5.4 Calibration set and CpG site number considerations

The accuracy and wide confidence intervals present in these models is likely a product of low samples sizes and limited age ranges of calibration samples. Issues relating to these calibration datasets are discussed in detail in chapter 7. Increasing the sample size, and especially the age range, is recommended to increase the accuracy and applicability of both age estimation models. In addition, the inclusion and identification of more informative markers present for a MLR model may increase these model's accuracy, with similar studies in humans and cetaceans having shown that increasing informative CpG sites can significantly increase the statistical power of age estimation models (Bekaert et al., 2015; Bors et al., 2020; Polanowski et al., 2014). The beluga epigenetic clock, while showing a comparatively strong relationship within its final model, utilised 23 different CpG sites (Bors et al., 2020). This is a far greater than the three sites used both in my sperm whale and the humpback epigenetic age estimators, and the six sites utilised by the southern right whale model (Polanowski et al., 2014). While the southern right whale model utilised six sites, these sites had considerably weaker correlations between age and methylation than those used in the other cetacean models, and a weaker final model fit was an expected result (Beal et al., 2019; Polanowski et al., 2014). These results highlight the utility of incorporating more CpG sites into age estimators. However, these benefits can be limited due to the increased number of sites that need assessment and the requirement for more costly, large scale age associated site determination approaches (Bors et al., 2020; Heiss et al., 2020). Future investigation into other age-related genes, such as *ELOVL2* and *GRIA2*, to potentially elucidate additional sites with age relationships could increase model accuracy and could lead to the development of more widely applicable age estimation assays.

5.5.5 Issues with homology-based epigenetic age estimation

Here, I developed a homology-based epigenetic analysis workflow that was able to successfully identify CpG site DNAm age-associations in sperm and southern right whales. However, this approach was not successful in all species. The model developed for pilot whales does not represent a viable tool for age estimation,

containing large error margins and poor regression coefficients. This is a result of weak correlations between age and DNAm of CpG sites utilised in the MLR model. The reason for this appears to be primarily due to the lack of age-related CpG sites in the homology-based loci assessed in this research. Potential inaccuracies within age records for the calibration set may also have been a source of error. As reported in other studies, often CpG sites and even genomic regions can lack age-association, even in closely related taxa (De Paoli-Iseppi, Polanowski, et al., 2017; Tanabe et al., 2020). The application of age estimation models to different species commonly utilises homology of the gene region being analysed between species, specifically within the first exon. The first exon of a target gene is an area shown to have strong relationships between methylation and transcriptional regulation, often being part of a CpG island, as well as being well conserved across taxa (Brenet et al., 2011). This makes these regions ideal targets for DNAm analysis. While vertebrates have tended to show conservation of age-associated DNAm in a various genes, such as *TET2* in whales, humans, and bats (Grönniger et al., 2010; Polanowski et al., 2014; Wright et al., 2018) or *CDKN2A* in mice, humans, and whales, these relationships do not exist in all vertebrates. For example, a recent study on birds found limited DNAm correlations with age in 13 genomic regions which presented strong correlations in mammalian model organisms, with the conclusion that orthologous mammalian genes in birds do not conserve age-related DNAm signatures (De Paoli-Iseppi, Polanowski, et al., 2017; Galamb et al., 2016; Maegawa et al., 2010; Polanowski et al., 2014). Similarly, age associated CpG sites identified in humpback whales were not conserved in Antarctic minke whales, though other age-associated CpG sites were still identified within the targeted genomic loci (Tanabe et al., 2020). These studies, and my results, indicate the potential issues arising from homology-based epigenetic analyses, including the workflow developed in chapter 4 and applied in this thesis. While useful in offering a rapid technique for applying previous assays to new species without the need for novel age-associated site determination, there is the potential that they are not applicable to all species. They are therefore best utilised with very closely related species where age associated sites have already been confirmed.

5.5.6 Future approach to a successful pilot whale epigenetic age estimator

With pilot whales being one of the most-commonly stranded cetacean species globally, further investigations into age demographics between stranded pods and healthy, live populations would increase our knowledge of this poorly understood phenomenon, as well as guiding conservation policy (Oremus et al., 2013). Novel identification of epigenetic age-associated sites and the development of a successful epigenetic age estimation tool for pilot whales is therefore proposed as a focus of further research. This would allow for the rapid assessment of wild and stranded populations of pilot whales and further investigation into the population demographics of the species. It is likely that multiple other loci contain unexplored age associated CpG sites, especially in regions such as *ELOVL2* and *GRIA2*. These genomic regions have been used as age estimation markers previously in other mammal species, but primers designed in this study from existing assays did not amplify in pilot whales (Koch & Wagner, 2011; Polanowski et al., 2014; Zbieć-Piekarska et al., 2015). Of particular interest would be *GRIA2*, a gene identified to have age associations with DNAm in minke whales, bottlenose dolphins, and humpback whales (Beal et al., 2019; Polanowski et al., 2014; Tanabe et al., 2020). While preliminary tests were attempted with *GRIA2* primers, initial amplifications failed, and the further development and sequencing of an assay was discontinued. I recommend further investigation of this gene, and others which have shown correlations between age and DNAm, which could potentially identify age associated CpG sites and feature in a successful epigenetic age estimator for pilot whales.

5.5.7 Recommendation for novel epigenetic age estimator development

For the development of an accurate age estimator, it may be more appropriate to utilise an alternative epigenetic analysis technique for the initial determination of age associated CpG sites, such as for those in pilot whales. While homology-based approaches have showed promise for southern right and sperm whales, the apparent lack of age associated sites identified in pilot whales highlights the major pitfall of a homology-based approach. Instead, the utilisation of a novel CpG site identifying technology may be more suitable here. In the beluga whale epigenetic age estimator, for example, age associated sites were discovered *de novo* through the use of a custom methylation array which measured methylation levels at more than 37,000 separate

CpG sites (Bors et al., 2020). From this, the 23 strongest age associated sites were utilised in the final robust epigenetic age estimator which achieved a regression coefficient of 0.92. Therefore, this approach may be more suitable for new species, or where homology-based approaches have previously failed. When these sites are identified, a homology-based approach such as the one developed here can then be utilised for rapid high-throughput analysis. I recommend that future research consider a combination of both homology-based applications and methylation array utilisation for *de novo* CpG site identification and sequencing.

Chapter 6 – Assessment of the novel Illumina-based high-throughput sequencing epigenetic analysis workflow

6.1 Adaptation to the Illumina high-throughput DNA sequencing workflow

This thesis makes an important contribution in the development of DNAm-based age estimation workflow utilising rapid Illumina high-throughput DNA sequencing and the bioinformatics pipelines able to interpret the data. The approach utilized PCR amplification of target regions from bisulfite-converted DNA with primers containing 5' adaptors for Illumina sequencing requirements and unique MIDs. This method, which is described in detail in chapter 4 and visually represented in Figure 6.1, was a novel conversion of the Illumina NGS system from previous eDNA metabarcoding work to age estimation using DNAm analysis (Koziol et al., 2019; Stat et al., 2017).

6.2 Advantages over the traditional epigenetic sequencing approach

The PyroMark pyrosequencing system (Qiagen) has been the primary method used for DNAm analysis in previous epigenetic age estimation studies (Beal et al., 2019; Ito et al., 2018; Polanowski et al., 2014). While this approach is generally able to achieve consistent results, there are a number of limitations: only one assay can be sequenced at a time; the cost per run is higher than Illumina systems; there are inherent PCR biases; and homopolymer and sequence accuracy issues exist (Beal et al., 2019; Crary-Dooley et al., 2017; Ito et al., 2018; Polanowski et al., 2014; Wright et al., 2018). By comparison, I was able to multiplex hundreds of samples over multiple assays onto one Illumina MiSeq sequencing run. This reduced time and sequencing costs, as well as the potential for experimental error, and led to high sequencing accuracy across all samples. This workflow could be further expanded into single-tube reaction assays, capable of sequencing bcDNA over multiple assays and allowing for rapid age estimation of many multiplexed samples. Full workflow completion, from sample DNA extraction to final CpG site DNAm results, could be realistically achieved in under three days across multiple assays for dozens of samples. While this approach is time and cost-effective, it is important that it is also accurate and precise. This chapter aims to assess the repeatability and accuracy of this novel workflow.

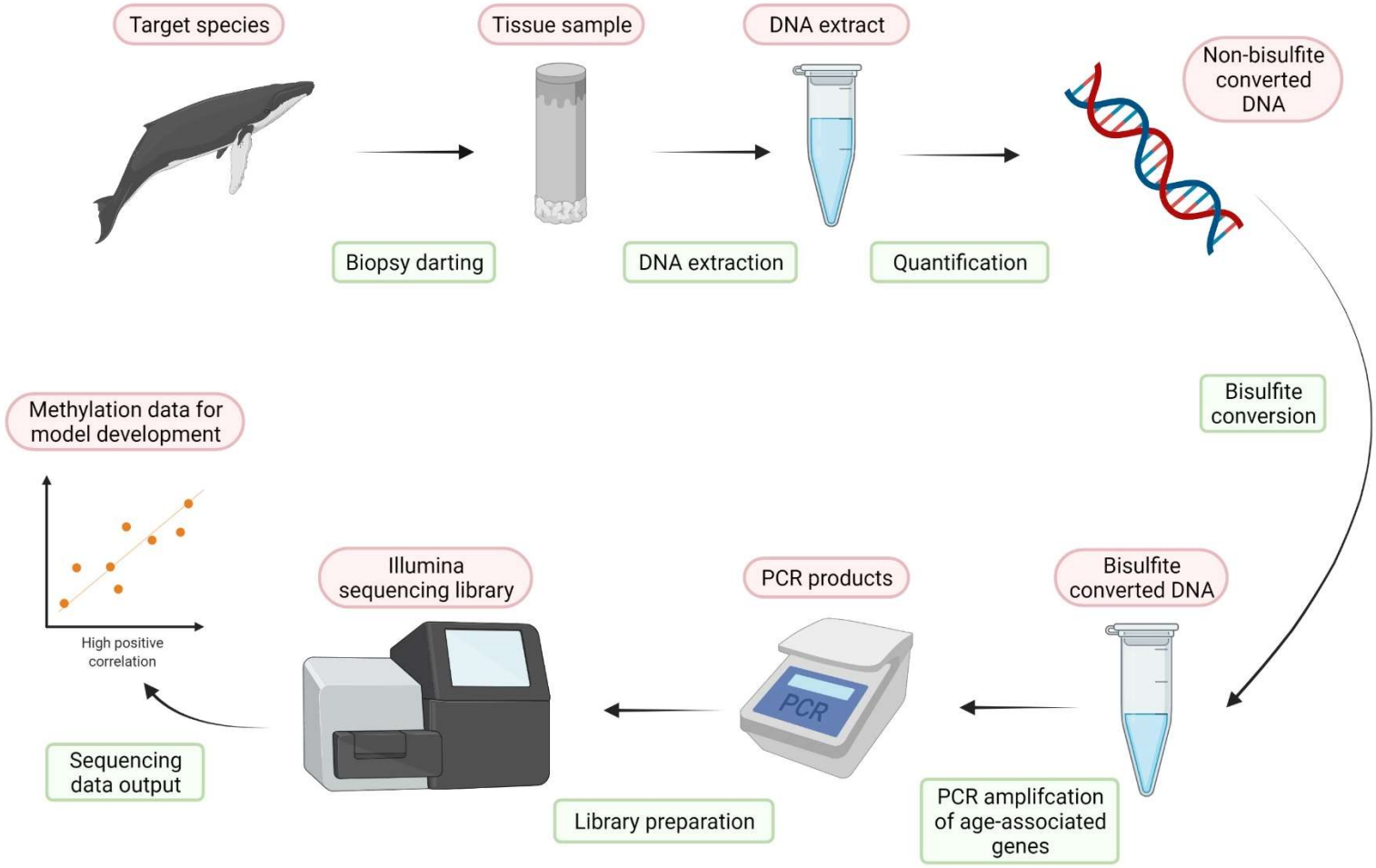

Figure 6.1 Illumina-based high-throughput epigenetic workflow diagram.

6.3 Assessment of Illumina platform sequencing

6.3.1 Methods

The Illumina NGS workflow, as detailed in chapter 4, was successfully implemented to sequence bcDNA, with all successfully analysed samples obtaining a sequencing depth of at least 500 reads and most samples sequenced with over 10,000 reads. To assess the repeatability of this workflow, a set of 31 pilot whale samples were re-sequenced on different runs and analysed at *CCDC105* site 14. A second assessment compared between-PCR replicate variation on a set of 32 separately sequenced PCR replicates of *KLF14* site 59. Finally, four DNA samples were bisulfite-converted separately and analysed over three sites within *KLF14* to determine whether variation existed between sample bcDNA replicates. A series of paired two-tailed t-tests compared these paired sample sets.

6.3.2 Results

Minimal variation between sequencing runs or replicates was found. There was no significant difference between runs. The two runs were statistically similar at *CCDC105* site 14 ($t_{30} = 0.4601$, $p = 0.6488$). There was also no significant difference between PCR replicates of *KLF14* site 59 ($n = 32$, $t_{31} = -0.5624$, $p = 0.5780$). There was also no significant difference between separate bisulfite conversions on DNA sample replicates of *KLF14* ($n = 12$, $t_{11} = 0.8538$, $p = 0.4114$).

6.4 Recommendations, considerations, and future applications

With appropriate NGS equipment available, and with previously identified age-associated regions, this study has shown that an age estimation assay can potentially be developed for a new species and applied within days. However, a consideration with this workflow are the limitations of homology-based approaches. These issues have been discussed in chapters 5 and 6, but as an amplicon focussed workflow, determination of novel age associated loci is difficult, and would require multiple primer sets for various loci to be designed. While it offers a viable and useful approach to CpG site methylation determination for known sites, in new species and for novel age-associated site identification, other approaches may be more appropriate. These include whole-genome bisulfite sequencing and custom methylation arrays, which have the ability to determine methylation levels at CpG sites throughout the genome

and identify novel sites for incorporation into epigenetic age estimators (Bors et al., 2020; Horvath & Raj, 2018; Olova et al., 2018). As discussed throughout this thesis, custom methylation array methodologies offer an effective tool for the identification of novel age associated CpG sites. However, these custom arrays are expensive to develop and require specialist equipment (Heiss et al., 2020; Lin et al., 2020). The novel workflow developed here offers a low-cost alternative that can be used to assess previously described CpG sites. It can be used by any molecular focused lab with minimal equipment requirements. The combination of both approaches may be most effective in the design and application of epigenetic age estimation to newly investigated species, as discussed in chapter 5. Regardless, this workflow offers a valuable tool for a range of ecological investigations, not only age estimation. DNAm can be used as a biomarker for many biological interactions, including pollutant exposure or environmental variations, and the utilisation of the NGS high-throughput amplicon workflow presented in this study could lead to significantly increased analysis capabilities for various ecological applications (Basu et al., 2013; Nilsen et al., 2016; Weyrich et al., 2016). I have successfully developed and adapted the epigenetic methodology used in previous, traditional pyrosequencing-based studies to the Illumina MiSeq NGS platform to sequence bcDNA accurately and at reduced comparative costs, in both time and monetary value (Ito et al., 2018; Polanowski et al., 2014; Wright et al., 2018; Zhou et al., 2006). This novel application is repeatable between runs, between PCR replicates, and between bisulfite conversions.

Chapter 7 – Discussion and future directions

7.1 Summary of findings

In this thesis, the growing potential of epigenetic analyses for use in ecological applications has been identified, especially for the estimation of animal age. This is important in cetacean studies, where age estimation has historically been difficult. Epigenetic analyses offer a viable and comparatively low-cost approach to the age estimation of live, wild animals, and represent a significant breakthrough for rapid biomonitoring of cetacean populations.

For the development of cetacean epigenetic age estimation assays, I have pioneered a novel, high-throughput epigenetic analysis workflow based upon the Illumina NGS platform. This workflow was able to rapidly assess homology-based loci that had shown associations between DNAm and age in previous research (Table 4.1). The precision and reproducibility of this workflow was assessed and found to be comparable to previous methodologies, while offering a reduced cost and more rapid amplicon focussed sequencing alternative.

Utilising this novel workflow, I developed a successful epigenetic age estimation assay for both southern right and sperm whales. The southern right whale age estimation assay can estimate the age of an individual with a mean absolute difference of 1.984 years within a 95% confidence interval of 5.667 years (8.8% of the animal's potential lifespan). The sperm whale age estimation assay can estimate the age of animals with a mean average difference of 3.875 years and within a 95% confidence interval of 11.91 years (17% of the animal's potential lifespan). These represent the first epigenetic age estimation models applied to either species. While achieving results comparable to published cetacean epigenetic age estimators, calibration dataset limitations reduce the full applicability of these models, and the addition of larger, more age diverse sample sets is recommended to completely validate this approach.

The inherent issues with homology-based epigenetic approaches, however, were highlighted by the failure to detect age-associated sites in long-finned pilot whales. This precluded the development of a successful age estimation model for that species. The lack of similarly age associated CpG sites and loci is a documented problem with epigenetic age estimation studies (De Paoli-Iseppi, Polanowski, et al., 2017; Tanabe et

al., 2020). I therefore conclude that the utilisation of alternative DNAm analyses, such as custom methylation arrays, may be best practice in the initial determination of age associated CpG sites in newly investigated species.

7.2 Limitations

The major limitations of the research conducted in this thesis relate to the calibration datasets utilised for southern right and sperm whale epigenetic analyses, and the number of loci assessed in sperm and pilot whales. While models were successfully developed for sperm and southern right whales, these limitations mean that they should be applied with caution. The southern right whale assay was designed for and tested on predominately young animals (< 12 years old), reducing its applicability to most adult animals. There was an appropriate age range for sperm whales, but only a small number of animals (n = 10) in the calibration dataset. This limits the practical confidence in application of this model, although statistically the model is sound. In the case of pilot whales, the lack of identified age associated CpG sites has precluded the development of an epigenetic age estimator.

7.2.1 Age estimation calibration datasets

A robust and extensive dataset is vital for the development of any statistical model. In epigenetic age estimation analysis, the key factors in developing a statistically reliable and precise assay are a sufficient number of samples, a range of ages within the samples, and accurate 'known' ages. The models developed in chapter 5 highlight the importance of high-fidelity, sufficiently sized calibration datasets.

Capturing an extensive range of ages is valuable in calibration datasets for age estimation models. The accuracy of the model can only be statistically assessed within the outer limits of an age range. Any age estimation or extrapolation of models outside of these limits is based on the assumption that the model will continue to be correct and accurate and that the correlations observed within those known age limits are conserved in older and younger individuals (Jaqaman & Danuser, 2006; Peters, Herrick, Urban, Gardner, & Breshears, 2004). As an organism ages, epigenetic relationships change, and changes in age and CpG site correlations in older and younger animals have been well documented, with age estimation assays often

reducing in accuracy towards the extremes of an animal's lifespan (Beal et al., 2019; Bors et al., 2020; Polanowski et al., 2014).

Calibration datasets that are based upon a limited number of calibration samples will increase the likelihood of spurious correlations between age and DNAm due to the reduction in statistical power and limitation of replication of results (Jaqaman & Danuser, 2006). Additionally, experimental noise, such as error associated with 'known age' assessments, and sex or age range biases and outliers will all be compounded by reduced sample numbers (Button et al., 2013; Van Selst & Jolicoeur, 1994; Varoquaux, 2018).

7.2.2 Pilot whale calibration dataset

The most extensive calibration dataset available in this research was the pilot whale set of 84 known age samples. As seen in Figure 7.1, sample ages ranged from two to 32 years old (see Table App A 1), with a relatively even spread in most age groups. This represents a strong dataset for the calibration of an epigenetic age estimation assay. It is comparable to existing published studies. The humpback whale assay used a sample set with similar age distribution (a few weeks to 30 years old) from 45 samples, and the beluga whale assay calibration dataset consisted of 67 samples from whales between 0 (foetus) and 49 years old (Bors et al., 2020; Polanowski et al., 2014). The pilot whale calibration data set presented in this thesis offers a strong dataset for future research. However, regarding the sites assessed here, no correlation between predicted age and known age was found, likely due to no strongly age associated CpG sites being present within the targeted genomic regions. Another potential confounding factor could be in the error produced by the known age estimates. Error here would compound error within the models, reducing the strength of the DNAm to age relationship.

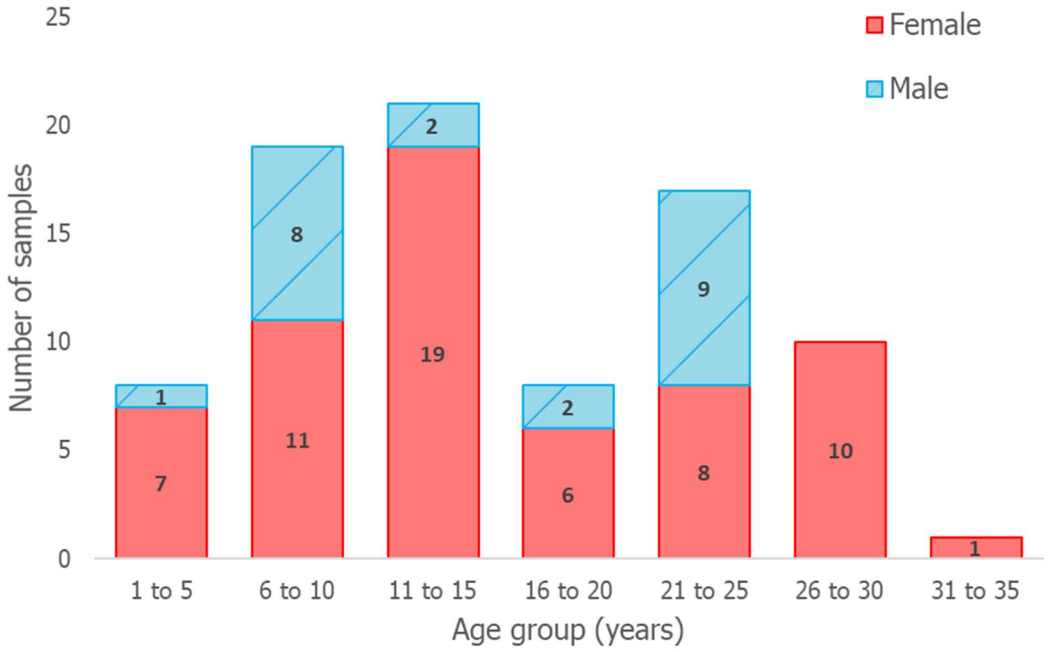

Figure 7.1 Age distributions of known age pilot whale samples used in final predictive models (n = 84). Females (n = 62) are represented by solid red bars and males (n = 22) by hatched blue bars. Numbers within bars represent number of individuals for that age group and sex.

7.2.3 Southern right and sperm whale calibration dataset

For sperm whales and southern right whales, where significant relationships between age and DNAm at a range of CpG sites were identified, the suitability of the calibration datasets is an important factor. However, as can be seen in both Figures 7.2 and 7.3, both datasets have limitations.

For southern right whales, the age of the calibration samples are heavily skewed to younger animals, with the oldest individual being 12 years old (Figure 7.2). This skewing of age was due to the sampling methodology, whereby replicate samples were collected from the same individuals, starting with their birth. This resulted in most individuals having a sample from their first year, but fewer recaptures occurring thereafter. Sampling was only started in 1995, so the maximum age of recaptures is also limited (Carroll et al., 2016). These samples, however, have true known ages, as they were tracked and sampled from birth, and as this collection grows, so does the potential for this to become a powerful dataset for future age estimation model calibration.

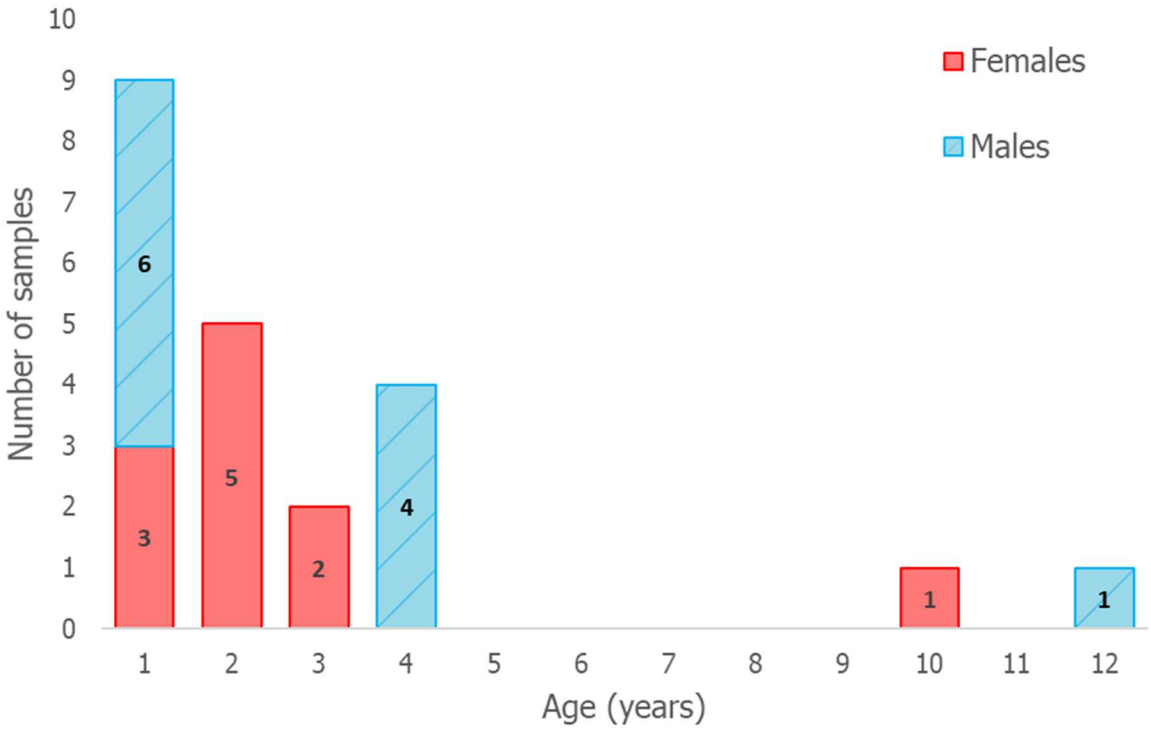

Figure 7.2 Age distributions of known age southern right whale samples used in final predictive models (n = 22). Females (n = 11) are represented by solid red bars and males (n = 11) by hatched blue bars. Numbers within bars represent number of individuals for that age and sex.

In sperm whales, the largest limitation in the calibration dataset was the small final sample number utilised for the predictive models. At only 10 individuals, this represents less than a third of most other epigenetic age predictive models (Beal et al., 2019). While a diverse age range of samples was present, there were no samples younger than 10 years old or older than 35 (Figure 7.3), although this is not necessarily a limitation. A similar age range has been used in previous cetacean epigenetic age estimation studies (Polanowski et al., 2014). Additionally, younger cetaceans can generally have their age estimated visually by size (Carroll et al., 2016; Polanowski et al., 2014). Therefore, this model is useful, but with acknowledgement of the small number of samples upon which it is based.

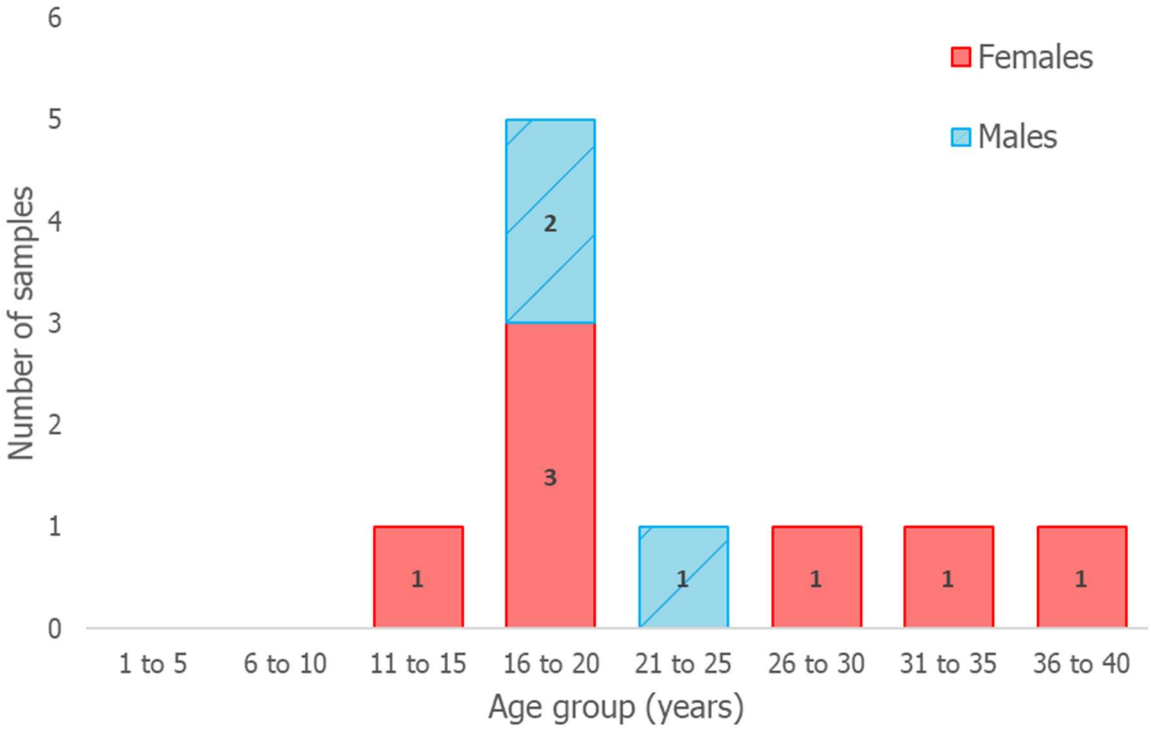

Figure 7.3 Age distributions of known age sperm whale samples used in final predictive models (n = 10). Females (n = 7) are represented by solid red bars and males (n = 3) by hatched blue bars. Numbers within bars represent number of individuals for that age group and sex.

7.2.4 Considerations and recommendations for calibration datasets

A statistical model relies upon its calibration dataset. In the development of models, a predictive assay's effectiveness or predictive accuracy is limited by the initial data upon which the model is built (Jaqaman & Danuser, 2006). The importance of calibration datasets is therefore paramount, and the issues with the datasets used in this thesis highlight this point. For an accurate, widely applicable epigenetic age estimation predictive model, both a large age range and a sufficient sample number are required.

In epigenetic age estimation analyses, this is especially relevant. Limited age ranges reducing a model's accuracy outside of the maximum and minimum age boundaries, requiring extrapolation for any older or younger samples, a process that reduces the suitability and applicability of an age estimation assay (Peters et al., 2004). Additionally, epigenetic age relationships have been shown to vary significantly with the extremes of age, and very young or old animals often have to be excluded from analyses due to their variable methylomes (Beal et al., 2019; Bors et al., 2020; Polanowski et al., 2014). For example, in the bottlenose dolphin epigenetic age

estimation model, the accuracy of the model was found to decrease with older individuals, a result concluded to be largely due to a lack of older individuals within the sample set (Beal et al., 2019). Additionally, it contributed to a skewing of ages towards younger individuals, a result similar to my southern right whale dataset. The oldest samples available in my datasets were: 32 years old for pilot whales, which is half the maximum pilot whale life expectancy of 60 years (Bloch et al., 1993; Lockyer, 1993; Martin et al., 1987); 35 years old for sperm whales, roughly half their expected lifespan (Evans & Hindell, 2004; Ohsumi, 1966); and only 12 years old for southern right whales, which is less than 20% of the estimated 65 year lifespan of right whales (Hamilton et al., 1998). But this may not necessarily be a critical limitation. For comparison, the published humpback whale age estimation assay had a maximum known age calibration sample of only a third of the potential life span of the species (Polanowski et al., 2014). The linear relationship observed in their study, as well as in other DNAm age-related investigations, indicated the model would retain accuracy outside of the calibration range (Bocklandt et al., 2011; Maegawa et al., 2010).

The sperm and pilot whale calibration datasets also contain potential for increased statistical noise within the final age estimation assay, as odontocete 'known' ages were estimated through GLG analysis with no exact chronological ages known. Therefore, the error around these age estimates is unknown. A perfect calibration dataset would consist of animals with known birth dates, such as those for the southern right whale dataset. Variation in age estimated through methods such as tooth GLG analysis will be compounded in molecular models developed using this information. While not often feasible, future development of DNAm-based age estimation assays should attempt to utilise animals with known chronological ages and across entire lifespans to limit the accumulation of statistical noise.

In future epigenetic age estimation research, I recommend diverse age profiles, with maximum age samples of at least two thirds of the target species' lifespan. I also recommend a dataset size of at least 40 individuals, based on previously published successful models, but increasing this number if possible is ideal. For calibration purposes, having true known sample ages is also preferable, and can be achieved by tracking animals from birth. This would reduce statistical noise and compounding error brought in by secondary age estimations such as GLG counts acting as a proxy for 'known' age.

7.3 Significance of thesis

This thesis represents the first application of epigenetic analyses to the determination of age in long-finned pilot whales and southern right whales, and the first successful epigenetic age estimator for southern right and sperm whales. While the age estimation of odontocetes through tooth cementum and dentine GLG analysis is the primary form of age estimation, and is accurate to within a year, it is only applicable to deceased animals, reduces in accuracy in older whales, and requires expert analysis (Bloch et al., 1993). For southern right whales, no practical method for age estimation of wild, live populations besides costly visual surveys exist (Carroll et al., 2016; Stamation et al., 2020). In comparison, my DNAm-based age estimation assay can be applied quickly to both live and stranded populations, is relatively inexpensive, and can be undertaken by any laboratory equipped with NGS equipment.

This research highlights the suitability of epigenetic approaches to age estimation as a low cost, low invasiveness alternative to traditional age estimation. In cetaceans, where age estimation of live animals is often impossible, this provides a powerful tool for conservation and resource managers. By utilising these tools, live, wild cetacean populations can be studied. This would increase understanding and information of population characteristics, and allow for informed resource and conservation management decisions to be made. By understanding age profiles, population recovery could be assessed, species reproduction rates could be monitored, and key populations for threatened species identified and protected.

Additionally, this thesis describes the development and testing of a novel and cost-effective high-throughput, homology-based epigenetic age estimation workflow based upon DNAm analysis and utilising the Illumina NGS sequencing platform. This workflow can be applied to both live whales, through minimally invasive biopsy darting, as well as allowing for rapid age estimation of deceased animals. The adaptability and cost-effectiveness of the NGS-based DNAm analysis workflow represents an innovative and effective tool for age estimation and other epigenetic studies that require high-throughput amplicon sequencing. The low cost and high speed of this approach are of great significance to conservation management groups, where budgets and timeframes are often restricted. This research also highlights the benefits and the potential pitfalls surrounding homology-based epigenetic age

analyses. Specifically, the failure to create a successful pilot whale age estimation model indicating the potential need for alternative epigenetic analyses for initial age associated marker determination.

7.4 Future directions

The refinement of the epigenetic age estimation assays developed for both southern right and sperm whales should be a future research focus. This could be achieved both through the addition of larger, age-diverse calibration datasets to produce more robust models, and the assessment of more loci that have been previously identified as age associated. Another viable approach to this would be using custom methylation arrays, an approach proven to identify age related CpG sites in cetaceans and able to build high quality epigenetic age estimators (Bors et al., 2020). This could also be applied to the pilot whale datasets, where homology-based approaches failed to identify age associated CpG sites. This may facilitate the development of an epigenetic age estimator for this species in the future.

While epigenetic age estimators offer clear benefits over traditional age estimation, the field of ecologically focused epigenetic analyses is still very much in its infancy. In this thesis, the potential for utilising epigenetic analyses for the determination of other key population biology characteristics besides age has been highlighted (chapter 2). The further development of epigenetic tests for sex and sexual maturity, which previous studies have indicated is possible, should be a key research focus. For animals that lack genomic sex determination this would offer an important tool for assessing sex ratios in populations, a valuable piece of population level information. This can be utilised in determining impacts of various stressors, including climate change, and assessing population health, especially with temperature dependent species such as reptiles. For sexual maturity, an epigenetic assay would represent the first molecular tool able to test for this key population biology parameter. This would provide a useful biological metric that could indicate the age of first reproduction of many species, indicating potential population growth, and also inform reproduction variability in response to stressors. This information would allow for more informed decisions around population management and conservation to be made.

7.5 Thesis conclusion

Age is an important metric when assessing animal populations. Age structure information allows for prediction of population growth and recovery rates and is useful in guiding conservation. For cetaceans, age estimation is especially difficult, and novel molecular methodologies offer great potential. This thesis describes an investigation into the application of epigenetic analyses for the assessment of age in cetaceans. Here, I have reviewed the growing body of research into ecologically applied epigenetics. I have identified a range of epigenetic age estimators, with many applied successfully to natural populations, including wild cetaceans. Further to this, I have successfully developed a novel approach to epigenetic analysis and validated its performance. This homology-based, high-throughput NGS workflow has been applied successfully for the first time in determining DNAm rates at age-associated loci in cetaceans. I was able to develop the first southern right and sperm whale epigenetic age estimation models, although with calibration dataset considerations that may limit their applicability. Additionally, I have highlighted key issues with homology-based epigenetic approaches, identifying no age-associated CpG sites in a large pilot whale dataset.

The results of this thesis provide a substantial contribution to the emerging field of ecological epigenetics. It provides future direction for the development of robust age estimation models, as well as a novel workflow for high-throughput epigenetic marker analysis. In a world where biodiversity is under increasing threat, tools for efficient and accurate biomonitoring are vital. The utilisation of the novel workflows and the further development and adaptation of epigenetic age estimators and other potential assays described throughout this thesis will all facilitate the development and growth of a valuable epigenetic biomonitoring toolbox.

References

Aird, D., Ross, M., Chen, W. S., Danielsson, M., Fennell, T., Russ, C., … Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(R18), 1–14.

Allsop, D. J., & West, S. A. (2003). Constant relative age and size at sex change for sequentially hermaphroditic fish. *Journal of Evolutionary Biology*, *16*(5), 921–929.

Amano, M., Yamada, T. K., Brownell, R. L., & Uni, Y. (2011). Age determination and reproductive traits of killer whales entrapped in ice off Aidomari, Hokkaido, Japan. *Journal of Mammalogy*, *92*(2), 275–282.

Amos, B., Schlötterer, C., & Tautz, D. (1993). Social structure of pilot whales revealed by analytical DNA profiling. *Science*, *260*(5108), 670–672.

Anastasiadi, D., Esteve-Codina, A., & Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin*, *11*(1), 37.

Anastasiadi, D., & Piferrer, F. (2020). A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Molecular Ecology Resources*, *20*(2), 387–397.

Angelini, F., & Ghiara, G. (1984). Reproductive modes and strategies in vertebrate evolution. *Bolletino Di Zoologia*, *51*(1–2), 121–203.

Angeloni, F., Wagemaker, N., Vergeer, P., & Ouborg, J. (2012). Genomic toolboxes for conservation biologists. *Evolutionary Applications*, *5*(2), 130–143.

Arnason, U., Gullberg, A., Gretarsdottir, S., Ursing, B., & Janke, A. (2000). The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. *Journal of Molecular Evolution*, *50*(6), 569–578.

Aylwin, C. F., Toro, C. A., Shirtcliff, E., & Lomniczi, A. (2019). Emerging Genetic and Epigenetic Mechanisms Underlying Pubertal Maturation in Adolescence. *Journal of Research on Adolescence*, *29*(1), 54–79.

Greig, A. B., Secchi, E. R., Zerbini, A. N., & Dalla Rosa, L. (2020). Stranding events of southern right whales, *Eubalaena australis*, in southern Brazil. *Journal of Cetacean Resource Management*, (1982), 157–160.

Baker, C. S., & Clapham, P. J. (2004). Modelling the past and future of whales and whaling. *Trends in Ecology and Evolution*, *19*(7), 365–371.

Baldanzi, S., Watson, R., McQuaid, C. D., Gouws, G., & Porri, F. (2017). Epigenetic variation among natural populations of the South African sandhopper *Talorchestia capensis*. *Evolutionary Ecology*, *31*(1), 77–91.

Basu, N., Head, J., Nam, D.-H., Pilsner, J. R., Carvan, M. J., Man, H., … Scheuhammer, A. M. (2013). Effects of methylmercury on epigenetic markers in three model species: Mink, chicken and yellow perch. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *157*(3), 322–327.

Beal, A. P., Kiszka, J. J., Wells, R. S., & Eirin-Lopez, J. M. (2019). The bottlenose dolphin epigenetic aging tool (BEAT): A molecular age estimation tool for small cetaceans. *Frontiers in Marine Science*, *6*(SEP), 1–10.

Beasley, I., Cherel, Y., Robinson, S., Betty, E., Hagihara, R., & Gales, R. (2019). Stomach contents of long-finned pilot whales, *Globicephala melas*, mass-stranded in Tasmania. *PloS One*, *14*(1), e0206747.

Bekaert, B., Kamalandua, A., Zapico, S. C., Van de Voorde, W., & Decorte, R. (2015). Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics*, *10*(10), 922–930.

Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodiversity and Conservation, 29*(7), 2089-2121.

Best, P. B., Brandão, A., & Butterworth, D. S. (2020). Demographic parameters of southern right whales off South Africa. *Journal of Cetacean Research and Management*, 161–169.

Bista, B., & Valenzuela, N. (2020). Turtle insights into the evolution of the reptilian karyotype and the genomic architecture of sex determination. *Genes*, *11*(4), 1–16.

Bloch, D., Lockyer, C., & Zachariassen, M. (1993). Age and Growth Parametres of the Long-Finned Pilot Whale off the Faroe Islands. *Report of the International Whaling Commission*, *15*(Special Issue 14), 163–208.

Bocklandt, S., Lin, W., Sehl, M. E., Sánchez, F. J., Sinsheimer, J. S., Horvath, S., & Vilain, E. (2011). Epigenetic predictor of age. *PloS One*, *6*(6), e14821.

Borrell, A., Bloch, D., & Desportes, G. (1995). Age trends and reproductive transfer of organochlorine compounds in long-finned pilot whales from the Faroe Islands. *Environmental Pollution*, *88*(3), 283–292.

Bors, E. K., Baker, C. S., Wade, P. R., ONeill, K., Shelden, K. E. W., Thompson, M. J., … Horvath, S. (2020). An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*, *14*(5), 1263-1273.

Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, *11*(2), 106–115.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: a unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182.

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PloS One*, *6*(1), e14524.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Camacho, C., Canal, D., & Potti, J. (2017). Lifelong effects of trapping experience lead to age-biased sampling: lessons from a wild bird population. *Animal Behaviour*, *130*, 133–139.

Campana, S. E. (2001). Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology*, *59*(2), 197–242.

Caracappa, S., Pisciotta, A., Persichetti, M. F., Caracappa, G., Alduina, R., & Arculeo, M. (2016). Nonmodal scutes patterns in the Loggerhead Sea Turtle (*Caretta caretta*): a possible epigenetic effect? *Canadian Journal of Zoology*, *94*(5), 379–383.

Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, *11*(7), 1094–1119.

Carroll, E. L., Fewster, R. M., Childerhouse, S. J., Patenaude, N. J., Boren, L., & Baker, C. S. (2016). First direct evidence for natal wintering ground fidelity and estimate of juvenile survival in the New Zealand southern right whale *Eubalaena australis*. *PloS One*, *11*(1), e0146590.

Casale, P., Mazaris, A., & Freggi, D. (2011). Estimation of age at maturity of loggerhead sea turtles *Caretta caretta* in the Mediterranean using length-frequency data. *Endangered Species Research*, *13*(2), 123–129.

Chaloupka, M., Osmond, M., & Kaufman, G. (1999). Estimating seasonal abundance trends and survival probabilities of humpback whales in Hervey Bay (east coast Australia). *Marine Ecology Progress Series*, *184*, 291–301.

Chatterjee, N., Gim, J., & Choi, J. (2018). Epigenetic profiling to environmental stressors in model and non-model organisms: Ecotoxicology perspective. *Environmental Health and Toxicology*, *33*(3), e2018015.

Chen, X., He, Y., Wang, Z., & Li, J. (2018). Expression and DNA methylation analysis of cyp19a1a in Chinese sea perch *Lateolabrax maculatus*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *226*(July), 85–90.

Chevallier, C., Gauthier, G., & Berteaux, D. (2017). Age Estimation of Live Arctic Foxes *Vulpes lagopus* Based on Teeth Condition. *Wildlife Biology*, *4*(1), wlb.00304.

Chittleborough, R. (1959). Determination of age in the humpback whale, *Megaptera nodosa* (Bonnaterre). *Marine and Freshwater Research*, *10*(2), 125.

Chittleborough, R. (1965). Dynamics of two populations of the humpback whale, *Megaptera novaeangliae* (Borowski). *Marine and Freshwater Research*, *16*(1), 33.

Clapham, P. J. (1992). Age at attainment of sexual maturity in humpback whales, *Megaptera novaeangliae*. *Canadian Journal of Zoology*, *70*(7), 1470–1472.

Cook, P. E., Hugo, L. E., Iturbe-Ormaetxe, I., Williams, C. R., Chenoweth, S. F., Ritchie, S. A., … O'Neill, S. L. (2006). The use of transcriptional profiles to predict adult mosquito age under field conditions. *Proceedings of the National Academy of Sciences*, *103*(48), 18060–18065.

Crary-Dooley, F. K., Tam, M. E., Dunaway, K. W., Hertz-Picciotto, I., Schmidt, R. J., & LaSalle, J. M. (2017). A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. *Epigenetics*, *12*(3), 206–214.

Darimont, C. T., Fox, C. H., Bryan, H. M., & Reimchen, T. E. (2015). The unique ecology of human predators. *Science*, *349*(6250), 858–860.

De Paoli-Iseppi, R., Deagle, B. E., McMahon, C. R., Hindell, M. A., Dickinson, J. L., & Jarman, S. N. (2017). Measuring Animal Age with DNA Methylation: From Humans to Wild Animals. *Frontiers in Genetics*, *8*(AUG), 2010–2017.

De Paoli-Iseppi, R., Polanowski, A. M., McMahon, C., Deagle, B. E., Dickinson, J. L., Hindell, M. A., & Jarman, S. N. (2017). DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. *PloS One*, *12*(12), e0189181.

De Paoli-Iseppi, R., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A., & Jarman, S. N. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular Ecology Resources*, *19*(2), 411–425.

De Silva, S., Elizabeth Webber, C., Weerathunga, U. S., Pushpakumara, T. V., Weerakoon, D. K., & Wittemyer, G. (2013). Demographic variables for wild Asian elephants using longitudinal observations. *PLoS ONE*, *8*(12).

Domingos, J. A., Budd, A. M., Banh, Q. Q., Goldsbury, J. A., Zenger, K. R., & Jerry, D. R. (2018). Sex-specific dmrt1 and cyp19a1 methylation and alternative splicing in gonads of the protandrous hermaphrodite barramundi. *PLOS ONE*, *13*(9), e0204182.

Donald, P. F. (2007). Adult sex ratios in wild bird populations. *Ibis*, *149*(4), 671–692.

Dunshea, G., Duffield, D., Gales, N., Hindell, M., Wells, R. S., & Jarman, S. N. (2011). Telomeres as age markers in vertebrate molecular ecology: invited technical review. *Molecular Ecology Resources*, *11*(2), 225–235.

Ellis, S., Franks, D. W., Nattrass, S., Cant, M. A., Bradley, D. L., Giles, D., … Croft, D. P. (2018). Postreproductive lifespans are rare in mammals. *Ecology and Evolution*, *8*(5), 2482–2494.

Ellis, S., Franks, D. W., Nattrass, S., Currie, T. E., Cant, M. A., Giles, D., … Croft, D. P. (2018). Analyses of ovarian activity reveal repeated evolution of post-reproductive lifespans in toothed whales. *Scientific Reports*, *8*(1), 1–11.

Elwen, S. H., Tonachella, N., Barendse, J., Collins, T., Best, P. B., Rosenbaum, H. C., … Gridley, T. (2014). Humpback whales off Namibia: occurrence, seasonality, and a regional comparison of photographic catalogs and scarring. *Journal of Mammalogy*, *95*(5), 1064–1076.

Evans, K., & Hindell, M. A. (2004). The age structure and growth of female sperm whales (*Physeter macrocephalus*) in southern Australian waters. *Journal of Zoology*, *263*(3), 237–250.

Fox, C. J., Benjamins, S., Masden, E. A., & Miller, R. (2018). Challenges and opportunities in monitoring the impacts of tidal-stream energy devices on marine vertebrates. *Renewable and Sustainable Energy Reviews*, *81*(February 2016), 1926–1938.

Freire-Aradas, A., Phillips, C., Mosquera-Miguel, A., Girón-Santamaría, L., Gómez-Tato, A., Casares de Cal, M., … Lareu, M. V. V. (2016). Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system. *Forensic Science International. Genetics*, *24*, 65–74.

Fretwell, P. T., Staniland, I. J., & Forcada, J. (2014). Whales from space: Counting southern right whales by satellite. *PLoS ONE*, *9*(2), 1–9.

Fullard, K. J., Early, G., Heide-Jorgensen, M. P., Bloch, D., Rosing-Asvid, A., & Amos, W. (2000). Population structure of long-finned pilot whales in the North Atlantic: a correlation with sea surface temperature? *Molecular Ecology*, *9*(7), 949–958.

Gabriele, C. M., Lockyer, C., Straley, J. M., Jurasz, C. M., & Kato, H. (2010). Sighting history of a naturally marked humpback whale (*Megaptera novaeangliae*) suggests ear plug growth layer groups are deposited annually. *Marine Mammal Science*, *26*(2), 443–450.

Galamb, O., Kalmár, A., Barták, B. K., Patai, Á. V, Leiszter, K., Péterfia, B., … Molnár, B. (2016). Aging related methylation influences the gene expression of key control genes in colorectal cancer and adenoma. *World Journal of Gastroenterology*, *22*(47), 10325–10340.

Giorli, G., & Goetz, K. T. (2019). Foraging activity of sperm whales (*Physeter macrocephalus*) off the east coast of New Zealand. *Scientific Reports*, *9*(1), 1–9.

Gonzalez-Zulueta, M., Bender, C. M., Yang, A. S., Nguyen, T., Beart, R. W., Van Tornout, J. M., & Jones, P. A. (1995). Methylation of the 5' CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Research*, *55*(20), 4531–4535.

Gopalan, S., Carja, O., Fagny, M., Patin, E., Myrick, J. W., McEwen, L. M., … Henn, B. M. (2017). Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. *Genetics*, *206*(3), 1659–1674.

Griffiths, R., & Tiwari, B. (1993). Primers for the differential amplification of the sex-determining region Y gene in a range of mammal species. *Molecular Ecology*, *2*(6), 405–406.

Grönniger, E., Weber, B., Heil, O., Peters, N., Stäb, F., Wenck, H., … Lyko, F. (2010). Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genetics*, *6*(5), e1000971.

Hamilton, P. K., Knowlton, A. R., Marx, M. K., & Kraus, S. D. (1998). Age structure and longevity in North Atlantic right whales *Eubalaena glacialis* and their relation to reproduction. *Marine Ecology Progress Series*, *171*, 285–292.

Hammerschlag, N., & Gallagher, A. J. (2017). Extinction Risk and Conservation of the Earth's National Animal Symbols. *BioScience*, *67*(8), 744–749.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., … Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*(2), 359–367.

Heiss, J. A., Brennan, K. J., Baccarelli, A. A., Téllez-Rojo, M. M., Estrada-Gutiérrez, G., Wright, R. O., & Just, A. C. (2020). Battle of epigenetic proportions: comparing Illumina's EPIC methylation microarrays and TruSeq targeted bisulfite sequencing. *Epigenetics*, *15*(1–2), 174–182.

Herman, D. P., Ylitalo, G. M., Robbins, J., Straley, J. M., Gabriele, C. M., Clapham, P. J., … Krahn, M. M. (2009). Age determination of humpback whales (*Megaptera novaeangliae*) through blubber fatty acid compositions of biopsy samples. *Marine Ecology Progress Series*, *392*, 277–293.

Herrel, A., Joly, D., & Danchin, E. (2020). Epigenetics in ecology and evolution. *Functional Ecology*, *34*(2), 381–384.

Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H. M., … Stuart, S. N. (2010). The impact of conservation on the status of the world's vertebrates. *Science (New York, N.Y.)*, *330*(6010), 1503–1509.

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., … O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105–108.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115.

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews. Genetics*, *19*(6), 371–384.

Huynen, L., Millar, C. D., & Lambert, D. M. (2002). A DNA test to sex ratite birds. *Molecular Ecology*, *11*(4), 851–856.

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports*, *8*(1), 9998.

Jackson, J. A., Patenaude, N. J., Carroll, E. L., & Baker, C. S. (2008). How few whales were there after whaling? Inference from contemporary mtDNA diversity. *Molecular Ecology*, *17*(1), 236–251.

Jaqaman, K., & Danuser, G. (2006). Linking data to models: Data regression. *Nature Reviews Molecular Cell Biology*, *7*(11), 813–819.

Jarman, S. N., Polanowski, A. M., Faux, C. E., Robbins, J., De Paoli-Iseppi, R., Bravington, M., & Deagle, B. E. (2015). Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology*, *24*(19), 4826–4847.

Jin, Z., & Liu, Y. (2018). DNA methylation in human diseases. *Genes and Diseases*, *5*(1), 1–8.

Jones, M. J., Goodman, S. J., & Kobor, M. S. (2015). DNA methylation and healthy human aging. *Aging Cell*, *14*(6), 924–932.

Jung, S. E., Lim, S. M., Hong, S. R., Lee, E. H., Shin, K. J., & Lee, H. Y. (2019). DNA methylation of the ELOVL2, FHL2, KLF14, C1orf132/MIR29B2C, and TRIM59 genes for age prediction from blood, saliva, and buccal swab samples. *Forensic Science International: Genetics*, *38*, 1–8.

Karlsson, A. O., Svensson, A., Marklund, A., & Holmlund, G. (2008). Estimating human age in forensic samples by analysis of telomere repeats. *Forensic Science International: Genetics Supplement Series*, *1*(1), 569–571.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

Khor, Y. M., Soga, T., & Parhar, I. S. (2016). Early-life stress changes expression of GnRH and kisspeptin genes and DNA methylation of GnRH3 promoter in the adult zebrafish brain. *General and Comparative Endocrinology*, *227*, 84–93.

Kniest, E., Burns, D., & Harrison, P. (2009). Fluke Matcher: A computer-aided matching system for humpback whale (*Megaptera novaeangliae*) flukes. *Marine Mammal Science*, *26*(3), 744–756.

Kobayashi, H., Hiura, H., John, R. M., Sato, A., Otsu, E., Kobayashi, N., … Arima, T. (2009). DNA methylation errors at imprinted loci after assisted conception originate in the parental sperm. *European Journal of Human Genetics*, *17*(12), 1582–1591.

Koch, C. M., Suschek, C. V., Lin, Q., Bork, S., Goergens, M., Joussen, S., … Wagner, W. (2011). Specific age-associated DNA methylation changes in human dermal fibroblasts. *PloS One*, *6*(2), e16679.

Koch, C. M., & Wagner, W. (2011). Epigenetic-aging-signature to determine age in different tissues. *Aging*, *3*(10), 1018–1027.

Koziol, A., Stat, M., Simpson, T., Jarman, S., DiBattista, J. D., Harvey, E. S., … Bunce, M. (2019). Environmental DNA metabarcoding studies are critically affected by substrate selection. *Molecular Ecology Resources*, *19*(2), 366–376.

Kukalová, M., Gazárková, A., & Adamík, P. (2013). Should I Stay or Should I go? The Influence of Handling by Researchers on Den use in an Arboreal Nocturnal Rodent. *Ethology*, *119*(10), 848–859.

Kurian, J. R., & Terasawa, E. (2013). Epigenetic control of gonadotropin releasing hormone neurons. *Frontiers in Endocrinology*, *4*(May), 1–7.

Kuroki, S., & Tachibana, M. (2018). Epigenetic regulation of mammalian sex determination. *Molecular and Cellular Endocrinology*, *468*, 31–38.

Kurose, N., Masuda, R., & Tatara, M. (2005). Fecal DNA analysis for identifying species and sex of sympatric carnivores: a noninvasive method for conservation on the Tsushima Islands, Japan. *The Journal of Heredity*, *96*(6), 688–697.

Laing, L. V., Viana, J., Dempster, E. L., Uren Webster, T. M., van Aerle, R., Mill, J., & Santos, E. M. (2018). Sex-specific transcription and DNA methylation profiles of reproductive and epigenetic associated genes in the gonads and livers of breeding zebrafish. *Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology*, *222*(April), 16–25.

Lambert, S., Blondin, P., Vigneault, C., Labrecque, R., Dufort, I., & Sirard, M. A. (2018). Spermatozoa DNA methylation patterns differ due to peripubertal age in bulls. *Theriogenology*, *106*, 21–29.

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2016). Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular Ecology*, *25*(8), 1681–1696.

Lebreton, J. D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, *62*(1), 67–118.

Lin, N., Liu, J., Castle, J., Wan, J., Shendre, A., Liu, Y., … He, C. (2020). Genome-wide DNA methylation profiling in human breast tissue by illumina TruSeq methyl capture EPIC sequencing and infinium methylationEPIC beadchip microarray. *Epigenetics*, *00*(00), 1–16.

Lindsay, A. R., & Belant, J. L. (2008). A simple and improved PCR-based technique for white-tailed deer (*Odocoileus virginianus*) sex identification. *Conservation Genetics*, *9*(2), 443–447.

Liu, S., Sun, K., Jiang, T., Ho, J. P., Liu, B., & Feng, J. (2012). Natural Epigenetic variation in the female great roundleaf bat (*Hipposideros armiger*) populations. *Molecular Genetics and Genomics*, *287*(8), 643–650.

Lockyer, C. (1993). A report on patterns of deposition of dentine and cement in teeth of pilot whales, genus *Globicephala*. *Reports of the International Whaling Commission, Special Issue 14*, (January 1993), 137–161.

Lockyer, C., Hohn, A. A., Doidge, W. D., Heide-Jørgensen, M. P., & Suydam, R. (2007). Age Determination in Belugas (*Delphinapterus leucas*): A Quest for Validation of Dentinal Layering. *Aquatic Mammals*, *33*(3), 293–304.

Lomniczi, A., Loche, A., Castellano, J. M., Ronnekleiv, O. K., Bosch, M., Kaidar, G., … Ojeda, S. R. (2013). Epigenetic control of female puberty. *Nature Neuroscience*, *16*(3), 281–289.

Low, M., & Pärt, T. (2009). Patterns of mortality for each life-history stage in a population of the endangered New Zealand stitchbird. *Journal of Animal Ecology*, *78*(4), 761–771.

Lowe, R., Danson, A. F., Rakyan, V. K., Yildizoglu, S., Saldmann, F., Viltard, M., … Faulkes, C. G. (2020). DNA methylation clocks as a predictor for ageing and age estimation in naked mole-rats, *Heterocephalus glaber*. *Aging*, *12*(5), 4394–4406.

Lowe, T. M., & Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964.

Maegawa, S., Hinkal, G., Kim, H. S., Shen, L., Zhang, L., Zhang, J., … Issa, J.-P. J. (2010). Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Research*, *20*(3), 332–340.

Martín-del-campo, R., Bárcenas-ibarra, A., Sifuentes-romero, I., Llera-Herrera, R., & García-Gasca, A. (2018). Methylation status of the putative Pax6 promoter in olive ridley sea turtle embryos with eye defects: An initial approach. *Mechanisms of Development*, (March), 0–1.

Martin, A. R., Reynolds, P., & Richardson, M. G. (1987). Aspects of the biology of Pilot whales (*Globicephala melaena*) in recent mass strandings on the British coast. *Journal of Zoology*, *211*(1), 11–23.

Matsumoto, Y., Buemio, A., Chu, R., Vafaee, M., & Crews, D. (2013). Epigenetic Control of Gonadal Aromatase (cyp19a1) in Temperature-Dependent Sex Determination of Red-Eared Slider Turtles. *PLOS ONE*, *8*(6), e63599.

Matsumoto, Y., & Crews, D. (2012). Molecular mechanisms of temperature-dependent sex determination in the context of ecological developmental biology. *Molecular and Cellular Endocrinology*, *354*(1), 103–110.

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*, *536*(7615), 143–145.

Mayne, B., Berry, O., Davies, C., Farley, J., & Jarman, S. (2019). A genomic predictor of lifespan in vertebrates. *Scientific Reports*, *9*(1), 17866.

McNab, B. K. (1980). Food Habits, Energetics, and the Population Biology of Mammals. *The American Naturalist*, *116*(1), 106–124.

McRae, L., Böhm, M., Deinet, S., Gill, M., & Collen, B. (2012). The Arctic Species Trend Index: using vertebrate population trends to monitor the health of a rapidly changing ecosystem. *Biodiversity*, *13*(3–4), 144–156.

Metzger, D. C. H., & Schulte, P. M. (2018). The DNA methylation landscape of stickleback reveals patterns of sex chromosome evolution and effects of environmental salinity. *Genome Biology and Evolution*, *10*(3), 775–785.

Miller, R. A. (2001). Biomarkers of aging: prediction of longevity by using age-sensitive T-cell subset determinations in a middle-aged, genetically heterogeneous mouse population. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *56*(4), 180-6.

Morán, P., & Pérez-figueroa, A. (2011). Methylation changes associated with early maturation stages in the Atlantic salmon. *BMC Genetics*, *12*(1), 86.

Morin, P. A., Nestler, A., Rubio-Cisneros, N. T., Robertson, K. M., & Mesnick, S. L. (2005). Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. *Molecular Ecology*, *14*(10), 3275–3286.

Mourocq, E., Bize, P., Bouwhuis, S., Bradley, R., Charmantier, A., de la Cruz, C., … Griesser, M. (2016). Life span and reproductive cost explain interspecific variation in the optimal onset of reproduction. *Evolution; International Journal of Organic Evolution*, *70*(2), 296–313.

Nabi, G., McLaughlin, R. W., Hao, Y., Wang, K., Zeng, X., Khan, S., & Wang, D. (2018). The possible effects of anthropogenic acoustic pollution on marine mammals' reproduction: an emerging threat to animal extinction. *Environmental Science and Pollution Research*, *25*(20), 19338–19345.

Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L., & Piferrer, F. (2011). DNA Methylation of the Gonadal Aromatase (cyp19a) Promoter Is Involved in Temperature-Dependent Sex Ratio Shifts in the European Sea Bass. *PLoS Genetics*, *7*(12), e1002447.

Nilsen, F. M., Parrott, B. B., Bowden, J. A., Kassim, B. L., Somerville, S. E., Bryan, T. A., … Guillette, L. J. (2016). Global DNA methylation loss associated with mercury contamination and aging in the American alligator (*Alligator mississippiensis*). *Science of the Total Environment*, *545–546*, 389–397.

Nussey, D. H., Baird, D., Barrett, E., Boner, W., Fairlie, J., Gemmell, N., … Monaghan, P. (2014). Measuring telomere length and telomere dynamics in evolutionary biology and ecology. *Methods in Ecology and Evolution*, *5*(4), 299–310.

Ohsumi, S. (1966). Sexual segregation of the sperm whale in the North Pacific. *Sci Rep Whales Res Inst*, *20*(1958), 1–16.

Ojeda, S. R., Lomniczi, A., Loche, A., Matagne, V., Kaidar, G., Sandau, U. S., & Dissen, G. A. (2010). The transcriptional control of female puberty. *Brain Research*, *1364*, 164–174.

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, *19*(1), 1–19.

Olsen, M., Bérubé, M., Robbins, J., & Palsbøll, P. J. (2012). Empirical evaluation of humpback whale telomere length estimates; quality control and factors causing variability in the singleplex and multiplex qPCR methods. *BMC Genetics*, *13*(1), 77.

Oremus, M., Gales, R., Kettles, H., & Baker, C. S. (2013). Genetic evidence of multiple matrilines and spatial disruption of kinship bonds in mass strandings of long-finned pilot whales, *Globicephala melas*. *Journal of Heredity*, *104*(3), 301–311.

Ortega-Recalde, O., Day, R. C., Gemmell, N. J., & Hore, T. A. (2019). Zebrafish preserve global germline DNA methylation while sex-linked rDNA is amplified and demethylated during feminisation. *Nature Communications*, *10*(1), 1–10.

Palsbøll, P. J., Allen, J., Bérube´, M., Clapham, P. J., Feddersen, T. P., Hammond, P. S., … Øien, N. (1997). Genetic tagging of humpback whales. *Nature*, *388*(6644), 767–769.

Palsbøll, P. J., Clapham, P., Mattila, D., Larsen, F., Sears, R., Siegismund, H., … Arctander, P. (1995). Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Marine Ecology Progress Series*, *116*(1–3), 1–10.

Parrott, B. B., & Bertucci, E. M. (2019). Epigenetic Aging Clocks in Ecology and Evolution. *Trends in Ecology and Evolution*, *34*(9), 767–770.

Parrott, B. B., Bowden, J. A., Kohno, S., Cloy-McCoy, J. A., Hale, M. D., Bangma, J. T., … Guillette, L. J. (2014). Influence of tissue, age, and environmental quality on DNA methylation in *Alligator mississippiensis*. *Reproduction*, *147*(4), 503–513.

Peters, D. P. C., Herrick, J. E., Urban, D. L., Gardner, R. H., & Breshears, D. D. (2004). Strategies for ecological extrapolation. *Oikos*, *106*(3), 627–636.

Pierron, F., Bureau du Colombier, S., Moffett, A., Caron, A., Peluhet, L., Daffe, G., … Baudrimont, M. (2014). Abnormal Ovarian DNA Methylation Programming during Gonad Maturation in Wild Contaminated Fish. *Environmental Science & Technology*, *48*(19), 11688–11695.

Pike, D. G., Gunnlaugsson, T., Desportes, G., Mikkelsen, B., Vikingsson, G. A., & Bloch, D. (2019). Estimates of the relative abundance of long-finned pilot whales (*Globicephala melas*) in the Northeast Atlantic from 1987 to 2015 indicate no long-term trends. *NAMMCO Scientific Publications*, *11*(2019).

Pilsner, J. R., Lazarus, A. L., Nam, D.-H., Letcher, R. J., Sonne, C., Dietz, R., & Basu, N. (2010). Mercury-associated DNA hypomethylation in polar bear brains via the LUminometric Methylation Assay: a sensitive method to study epigenetics in wildlife. *Molecular Ecology*, *19*(2), 307–314.

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*, *14*(5), 976–987.

Pomilla, C., & Rosenbaum, H. C. (2005). Against the current: An inter-oceanic whale migration event. *Biology Letters*, *1*(4), 476–479.

Python Core Team. (2019). *Python: A dynamic, open source programming language*. Wilmington: Python Software Foundation. Retrieved from https://www.python.org/

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Retrieved from www.r-project.org

Radhakrishnan, S., Literman, R., Neuwald, J. L., & Valenzuela, N. (2018). Thermal Response of Epigenetic Genes Informs Turtle Sex Determination with and without Sex Chromosomes. *Sexual Development*, 308–319.

Reed, D. H., O'Grady, J. J., Brook, B. W., Ballou, J. D., & Frankham, R. (2003). Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biological Conservation*, *113*(1), 23–34.

Rey, O., Eizaguirre, C., Angers, B., Baltazar-Soares, M., Sagonas, K., Prunier, J. G., & Blanchet, S. (2020). Linking epigenetics and biological conservation: Towards a conservation epigenetics perspective. *Functional Ecology*, *34*(2), 414–427.

Richards, R. (2009). Past and present distributions of southern right whales (*Eubalaena australis*). *New Zealand Journal of Zoology*, *36*(4), 447–459.

Ricklefs, R. E. (2010). Life-history connections to rates of aging in terrestrial vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(22), 10314–10319.

Riekkola, L., Zerbini, A. N., Andrews, O., Andrews-Goff, V., Baker, C. S., Chandler, D., … Constantine, R. (2018). Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. *Ecological Indicators*, *89*(December 2017), 455–465.

Road, U., Nagabhushana, A., & Mishra, R. K. (2016). Finding clues to the riddle of sex determination in zebrafish. *Journal of Biosciences*, *41*(1), 145–155.

Robert, A., & Withers, P. (2007). *StatistiXL, Version 1.8, a powerful statistics and statistical analysis add-in for Microsoft Excel*. Washington DC: Microsoft. Retrieved from https://www.statistixl.com/

Rzeczkowska, P. A., Hou, H., Wilson, M. D., & Palmert, M. R. (2014). Epigenetics: A new player in the regulation of mammalian puberty. *Neuroendocrinology*, *99*(3–4), 139–155.

Sargsyan, A., Simonyan, A., Hovhannisyan, G., Arakelyan, M., & Aroutiounian, R. (2019). Application of the comet assay, micronucleus test and global DNA methylation analysis in *Darevskia* lizards as a sentinel organism for genotoxic monitoring of soil pollution. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*, *842*(October 2018), 117–124.

Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., … Wordsworth, S. (2020). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *22*(1), 85–94.

Seersholm, F. V., Werndly, D. J., Grealy, A., Johnson, T., Keenan Early, E. M., Lundelius, E. L., … Bunce, M. (2020). Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. *Nature Communications*, *11*(1), 1–10.

Seguin-Orlando, A., Schubert, M., Clary, J., Stagegaard, J., Alberdi, M. T., Prado, J. L., … Orlando, L. (2013). Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. *PloS One*, *8*(10), 19–23.

Serrano, E., Gallego, L., & Perez, J. M. (2004). Ossification of the Appendicular Skeleton in the Spanish Ibex *Capra pyrenaica* Schinz, 1838 (Artiodactyla: Bovidae), with Regard to Determination of Age. *Anatomia, Histologia, Embryologia: Journal of Veterinary Medicine Series C*, *33*(1), 33–37.

Shaw, C. N., Wilson, P. J., & White, B. N. (2003). A reliable molecular method of gender determination for mammals. *Journal of Mammalogy*, *84*(1), 123–128.

Smith, G., Smith, C., Kenny, J. G., Chaudhuri, R. R., & Ritchie, M. G. (2015). Genome-Wide DNA Methylation Patterns in Wild Samples of Two Morphotypes of Threespine Stickleback (*Gasterosteus aculeatus*). *Molecular Biology and Evolution*, *32*(4), 888–895.

Soto, F. A., Grandi, M. F., Garcia, N. A., Crespo, E. A., & Dans Silvana, L. (2017). Reproductive Parameters of Female Long-finned Pilot Whales (*Globicephala melas edwardii*) from the Southwestern Atlantic. *Zoological Studies*, *56*(39), 1–12.

Soulsbury, C. D., Lipponen, A., Wood, K., Mein, C. A., Hoffman, J. I., & Lebigre, C. (2018). Age- and quality-dependent DNA methylation correlate with melanin-based coloration in a wild bird. *Ecology and Evolution*, *8*(13), 6547–6557.

Spólnicka, M., Pośpiech, E., Pepłońska, B., Zbieć-Piekarska, R., Makowska, Ż., Pięta, A., … Branicki, W. (2018). DNA methylation in ELOVL2 and C1orf132 correctly predicted chronological age of individuals from three disease groups. *International Journal of Legal Medicine*, *132*(1), 1–11.

Stamation, K., Watson, M., Moloney, P., Charlton, C., & Bannister, J. (2020). Population estimate and rate of increase of southern right whales *Eubalaena australis* in Southeastern Australia. *Endangered Species Research*, *41*, 373–383.

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., … Bunce, M. (2017). Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*(1), 12240.

Stat, M., John, J., DiBattista, J. D., Newman, S. J., Bunce, M., & Harvey, E. S. (2019). Combined use of eDNA metabarcoding and video surveillance for the assessment of fish biodiversity. *Conservation Biology : The Journal of the Society for Conservation Biology*, *33*(1), 196–205.

Strah, R., & Kunej, T. (2019). Molecular sexing assays in 114 mammalian species: In silico sequence reanalysis and a unified graphical visualization of diagnostic tests. *Ecology and Evolution*, *9*(8), 5018–5028.

Tachibana, M. (2016). Epigenetics of sex determination in mammals. *Reproductive Medicine and Biology*, *15*(2), 59–67.

Tanabe, A., Shimizu, R., Osawa, Y., Suzuki, M., Ito, S., Goto, M., … Sahara, H. (2020). Age estimation by DNA methylation in the Antarctic minke whale. *Fisheries Science*, *86*(1), 35–41.

Teschendorff, A. E., West, J., & Beck, S. (2013). Age-associated epigenetic drift: Implications, and a case of epigenetic thrift? *Human Molecular Genetics*, *22*(R1), 7–15.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., … Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148.

Thompson, G. G. (2007). Terrestrial vertebrate fauna surveys for the preparation of environmental impact assessments; how can we do it better? A Western Australian example. *Environmental Impact Assessment Review*, *27*(1), 41–61.

Thompson, M. J., VonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068.

Toïgo, C., Gaillard, J. M., Festa-Bianchet, M., Largo, E., Michallet, J., & Maillard, D. (2007). Sex- and age-specific survival of the highly dimorphic Alpine ibex: Evidence for a conservative life-history tactic. *Journal of Animal Ecology*, *76*(4), 679–686.

Trukhina, A. V., Lukina, N. A., Wackerow-Kouzova, N. D., & Smirnov, A. F. (2013). The variety of vertebrate mechanisms of sex determination. *BioMed Research International*, *2013*, 587460.

Valsecchi, E., Hale, P., Corkeron, P., & Amos, W. (2002). Social structure in migrating humpback whales (*Megaptera novaeangliae*). *Molecular Ecology*, *11*(3), 507–518.

Van Selst, M., & Jolicoeur, P. (1994). A Solution to the Effect of Sample Size on Outlier Elimination. *The Quarterly Journal of Experimental Psychology Section A*, *47*(3), 631–650.

Varoquaux, G. (2018). Cross-validation failure: Small sample sizes lead to large error bars. *NeuroImage*, *180*(June 2017), 68–77.

Wang, M., & Lemos, B. (2019). Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Research*, *29*(3), 325–333.

Wang, W. Y., Srivathsan, A., Foo, M., Yamane, S. K., & Meier, R. (2018). Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. *Molecular Ecology Resources*, *18*(3), 490–501.

Wang, Y., Wang, C., Zhang, J., Chen, Y., & Zuo, Z. (2009). DNA hypomethylation induced by tributyltin, triphenyltin, and a mixture of these in *Sebastiscus marmoratus* liver. *Aquatic Toxicology*, *95*(2), 93–98.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191.

Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., … Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology*, *15*(2), R24.

Wen, A. Y., You, F., Sun, P., Li, J., Xu, D. D., Wu, Z. H., … Zhang, P. J. (2014). CpG methylation of dmrt1 and cyp19a promoters in relation to their sexual dimorphic expression in the Japanese flounder *Paralichthys olivaceus*. *Journal of Fish Biology*, *84*(1), 193–205.

Weyrich, A., Benz, S., Karl, S., Jeschek, M., Jewgenow, K., & Fickel, J. (2016). Paternal heat exposure causes DNA methylation and gene expression changes of Stat3 in Wild guinea pig sons. *Ecology and Evolution*, *6*(9), 2657–2666.

White, P. A., Ikanda, D., Ferrante, L., Chardonnet, P., Mesochina, P., & Cameriere, R. (2016). Age Estimation of African Lions *Panthera leo* by Ratio of Tooth Areas. *PloS One*, *11*(4), e0153648.

Whitehead, H. (2002). Estimates of the current global population size and historical trajectory for sperm whales. *Marine Ecology Progress Series*, *242*, 295–304.

Wright, P. G. R., Mathews, F., Schofield, H., Morris, C., Burrage, J., Smith, A., … Hamilton, P. B. (2018). Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*, *18*(6), 1374–1380.

Yang, C., Ye, J., Liu, Y., Ding, J., Liu, H., Gao, X., … Ling, Y. (2018). Methylation pattern variation between goats and rats during the onset of puberty. *Reproduction in Domestic Animals*, *53*(3), 793–800.

Yang, C., Zhang, Y., Liu, W., Lu, X., & Li, C. (2018). Genome-wide analysis of DNA methylation in five tissues of sika deer (*Cervus nippon*). *Gene*, *645*(August 2017), 48–54.

Young, J. M., & Linacre, A. (2020). Massively parallel sequencing is unlocking the potential of environmental trace evidence. *Forensic Science International: Genetics*, *50*(July 2020), 102393.

Yuan, X., Ye, S., Chen, Z., Pan, X., Huang, S., Li, Z., … Zhang, Z. (2019). Dynamic DNA methylation of ovaries during pubertal transition in gilts. *BMC Genomics*, *20*(1), 1–13.

Zabel, F., Delzeit, R., Schneider, J. M., Seppelt, R., Mauser, M., & Václavík, T. (2019). Global impacts of future cropland expansion and intensification on agricultural markets and biodiversity. *Nature Communications*, *10*(2844), 1–10.

Zbieć-Piekarska, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Makowska, Ż., Pałeczka, A., … Branicki, W. (2015). Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics*, *17*, 173–179.

Zhou, Z., Poe, A. C., Limor, J., Grady, K. K., Goldman, I., McCollum, A. M., … Udhayakumar, V. (2006). Pyrosequencing, a high-throughput method for detecting single nucleotide polymorphisms in the dihydrofolate reductase and dihydropteroate synthetase genes of *Plasmodium falciparum*. *Journal of Clinical Microbiology*, *44*(11), 3900–3910.

Zubakov, D., Liu, F., van Zelm, M. C., Vermeulen, J., Oostra, B. A., van Duijn, C. M., … Langerak, A. W. (2010). Estimating human age from T-cell DNA rearrangements. *Current Biology, 20*(22), R970-971.

Appendix A

Appendix A contains tables in support of chapters 4, 5, 6, and 7.

Table App A 1 Age and sex information for cetacean samples used in epigenetic age analysis workflows. SRW indicates southern right whale; PW indicates pilot whale; SW indicates sperm whale.

Species	Name	Sex	Age	Species	Name	Sex	Age
SRW	10B	M	4	PW	1718	F	26.66667
SRW	11B	F	3	PW	1719	F	26
SRW	12A	F	1	PW	1721	M	9.666667
SRW	12B	F	2	PW	1722	M	21
SRW	13B	F	3	PW	1723	M	12.66667
SRW	14B	M	12	PW	1725	F	30.33333
SRW	15B	F	10	PW	1726	F	8
SRW	16A	F	1	PW	1728	F	21.33333
SRW	2B	F	2	PW	1729	F	12.33333
SRW	2C	F	2	PW	1730	F	16.33333
SRW	3A	F	1	PW	1733	F	11
SRW	3B	F	2	PW	1734	F	21.33333
SRW	3C	F	2	PW	1735	F	24.66667
SRW	4A	M	1	PW	1736	F	26
SRW	4D	M	4	PW	1737	F	13
SRW	5A	M	1	PW	1738	F	14.66667
SRW	5C	M	4	PW	1739	M	9
SRW	6A	M	1	PW	1740	F	14.33333
SRW	6B	M	4	PW	1741	F	12.33333
SRW	7A	M	1	PW	1742	F	13.66667
SRW	8A	M	1	PW	1743	F	10.33333
SRW	9A	M	1	PW	1745	F	15
SW	1659	F	17	PW	1746	M	22.66667
SW	1660	F	28	PW	1749	F	4
SW	1661	M	20	PW	1750	M	21.66667
SW	1662	M	23	PW	1751	F	30

SW	1663	M	18	PW	1752	F	10
SW	1664	F	35	PW	1753	F	13.33333
SW	1665	F	16	PW	1754	M	12
SW	1666	F	18	PW	1755	F	13
SW	1667	F	38	PW	1756	M	10.33333
SW	1668	F	15	PW	1757	F	16.33333
PW	1686	F	25.66667	PW	1760	M	2
PW	1688	F	24	PW	1761	F	23
PW	1689	F	25	PW	1762	F	20.66667
PW	1690	M	22	PW	1763	F	5.666667
PW	1691	F	11.33333	PW	1764	F	3
PW	1692	M	16	PW	1765	M	21
PW	1694	M	8.666667	PW	1766a	F	15
PW	1695	F	13.33333	PW	1767a	F	28
PW	1696	F	7.666667	PW	1768	F	6.333333
PW	1697	M	22	PW	1769	M	20
PW	1698	F	13	PW	1772	F	26
PW	1699	F	15	PW	1774	F	29.66667
PW	1703	F	28.66667	PW	1775	F	16.33333
PW	1704	F	17.33333	PW	1776	F	10.66667
PW	1705	F	21.33333	PW	1779	M	9
PW	1706	F	16.66667	PW	1780	F	2
PW	1708_1	F	13.66667	PW	1781	F	13.33333
PW	1709	M	7.666667	PW	37	M	9
PW	1710	F	10.33333	PW	38	F	4
PW	1711	F	17.66667	PW	39	F	5
PW	1712	F	8.666667	PW	40	F	8
PW	1713	F	5.666667	PW	41	F	7
PW	1714	F	11.66667	PW	42	M	24
PW	1715	M	22.66667	PW	43	F	5
PW	1716	M	23	PW	44	F	5
PW	1717	F	32	PW	45	M	8

Table App A 2 AIC and BIC results for multiple linear regressions used in predictive models. Combinations in bold represent the chosen multilinear regression, based on AIC and BIC scores and the strength of the model's regression. Only a subset of the best scoring southern right whale AIC and BIC results is shown.

Sperm whales		
Model	AIC	BIC
CCDC	74.25737	75.16513
KLF	74.41268	75.32043
PDE4C	65.62143	66.52918
PDE4CKLFCCDC	63.73773	65.25066
CCDCKLF	76.02689	77.23723
PDE4CKLF	67.60609	68.81643
PDE4CCCDC	63.98648	65.19682
Pilot whales		
Model	AIC	BIC
CDK	580.5147	587.8071
TET	584.07	591.3625
TET18S	584.6767	591.9691
18STETCDK	584.3397	596.4938
CDKTET	582.4663	592.1896
18STET	585.7271	595.4504
18SCDK	582.3615	592.0848
Southern right whales		
Model	AIC	BIC
TET	110.0547	113.3279
CDK	105.0479	108.321
CCDC	106.9775	110.2506
KLF	106.2687	109.5418
PDE4C	105.2513	108.5244
18S	109.6963	112.9694
TETCCDC	105.903	110.2672

TETKLF	106.6764	111.0406
TETCDK	105.0219	109.3861
TETCCDCKLF	104.1253	109.5805
TETCDKKLF	104.6709	110.1261
TETCCDCCDK	103.9509	109.4061
TETCCDCKLFCDK	103.9138	110.4601
CDKCCDC	103.8781	108.2422
CDKKLF	103.7823	108.1465
CDKCCDCKLF	103.015	108.4702
CCDCKLFCDK18S	104.7927	111.339
CCDCKLFCDKTETPDE4C18S	103.1472	111.8755
CCDCKLFCDKTETPDE4C	103.0432	110.6805
CCDCKLFCDKPDE4C18S	101.2506	108.8879
CCDCKLFCDKTET18S	105.845	113.4823
CCDCKLFTETPDE4C18S	102.8335	110.4708
CCDCTETPDE4C18S	102.9573	109.5036
KLFTETPDE4C18S	107.3525	113.8988
CCDC18S	107.6196	111.9838
TETPDE4C18S	107.9314	113.3866

Appendix B

To Whom It May Concern,

I, Matthew James Heydenrych, wrote and edited the manuscript, researched literature, developed the figures and tables, and contributed to the conceptual design of the review titled "Epigenetic measurement of key vertebrate population biology parameters" which is submitted to the journal *Frontiers in Ecology and Evolution*.

_____

I, as a Co-Author, endorse that this level of contribution by the candidate indicated above is appropriate.

Simon Neil Jarman _____

Benjamin John Saunders _____

Michael Bunce _____