School of Pharmacy and Biomedical Sciences

The Investigation of Sex Specific DNA Regions in *Morelia spilota imbricata*

Ryan Peter Harris

This thesis is presented for the Degree of

Doctor of Philosophy - Biomedical Science

of

Curtin University

June 2018

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.

Animal Ethics

The research presented and reported in this thesis was conducted in compliance with the National Health and Medical Research Council Australian code for the care and use of animals for scientific purposes 8th edition (2013). The proposed research study received animal ethics approval from the Curtin University Animal Ethics Committee, Approval Number # AEC_2011_25.

Signature:

Date: December 12th, 2018

Acknowledgements

I would like to acknowledge and express my gratitude to the following people and organisations for the support and assistance they have provided to me while I conducted and completed my PhD thesis.

First and foremost, I would like to express my gratitude to my primary PhD supervisor Associate Professor David Groth. Thank you, David, for not only allowing me to undertake a PhD thesis that was obscure and outside of your normal research field, but also contributing a large proportion of the funding required to undertake the research. Your unyielding determination to pass on your knowledge of scientific research and molecular genetics and your mentorship has always had a positive impact on me and my life.

I would also like to thank my secondary PhD supervisor Dr Keith Gregg and my close friend Dr Saul Cowan for their contribution to my research and my thesis preparations. And all friends and colleagues that have contributed to my research and my thesis preparations.

I would like to express my gratitude to all individuals and organisations contributing tissue samples used to conduct research in this thesis. This includes, Ken Thompson from Rockingham Reptile & Grow. The Gaikhorst family, particularly Mieke and Kristy from Armadale Reptile and Wildlife Centre. And the West Australian Reptile Park. A special thank you to veterinarian Karen Payne for donating her time and expertise in collecting snake blood samples for the project.

I would like to express my deepest gratitude to my wife Alicia Harris. Thank you, Alicia, for the love and support you have provided to me throughout my studies, my thesis, and my life.

Lastly, I would like to thank Curtin University. Where I completed my undergraduate bachelor’s degree and honour’s degrees and now my PhD. Thank you, Curtin, for awarding me an Australian Postgraduate Award (APA) / Curtin University Postgraduate Scholarship (CUPS) scholarship, without which I would not have been able to afford to complete my PhD.

Abstract

It is a widely accepted theory, that all snake species share a common, female heterogametic, sex determination system, comprising ZZ males and ZW females. However, the research evidence supporting this theorem only exists for snake species belonging to the three snake families Colubridae, Viperidae and Elapidae. Snake species belonging to other evolutionary older snake families, such as Pythonidae and Boidae, have had no conclusive evidence presented that supports this theory. In this thesis, three independent molecular genetic techniques have been used; Random Amplified Polymorphic DNA (RAPD) analysis, Representational Difference Analysis (RDA) and a comparative genomics approach to identify sex specific differences in a non-model snake species *Morelia spilota*, belonging to the Pythonidae snake family.

This investigation has resulted in the identification and characterisation of a series of DNA regions that could be sex-linked in *M. spilota*. After investigating 102 RAPD primers, a 440bp RAPD fragment was identified that most likely represents a repetitive element and has the potential to be sex associated. A series of other possible sex-linked DNA regions were also identified using RDA. Five DNA regions identified have functions associated with sexual development, sex reversal, fertility, or sex cell maturation in vertebrate species. One DNA region representing the Mbt Domain Containing 1 (*MBTD1*) gene, is likely to be located on the Z chromosome of snakes. Additionally, two DNA regions that show DNA sequence discrepancies between sexes were identified. It is believed that these discrepancies are likely caused by the presence of an allele representing these two DNA regions, which are only present in female snakes.

A comparative genomics approach involving sequencing of DNA regions representing 13 of the 20 genes, previously mapped to the Z chromosome of *Elaphe quadrivirgata* was performed. Approximately 16,045bp of genomic DNA was sequenced from the proposed snake Z chromosome. However, no sex-linked differences within the DNA regions sequenced were identified.

In the final stages of this thesis, an attempt to expand our knowledge of the previously identified DNA regions and the surrounding DNA regions was performed using genomic libraries. Two large insert, cosmid, genomic libraries were prepared, representing one male and one female *M. spilota*. A comparative analysis of the DNA sequences produced from the male and female libraries using second-generation sequencing techniques was intended. However, difficulties in the cosmid clone screening protocol and difficulties with the relatively new second-generation sequencing techniques,

lead to additional sequencing information that most likely did not represent the DNA regions of interest.

In this thesis, three techniques were developed which has resulted in the identification of a series of DNA regions, which have the potential to be sex-linked in the snake species *M. spilota*. DNA sequences from multiple regions of genomic DNA representing the proposed Z chromosome of *M. spilota* were also generated. While two cosmid libraries representing the male and female genome of *M. spilota* were produced, identifying and sequencing clones from these libraries, using, at the time, a relatively new second-generational sequencing technique proved more difficult and time consuming than originally anticipated. As is commonly the case when performing fundamental research, the results produced within this thesis cannot be used to prove or disprove our original hypothesis. However, it is hoped that the DNA regions thus identified, the techniques developed, and resources created can be used to further our understanding of snake sex determination.

Glossary of Terms

GSD	Genetic Sex Determination
ESD	Environment Sex Determination
TSD	Temperature-dependent Sex Determination
bp / kb / Mb / Gb	basepair / kilo basepair / mega basepair / giga basepair
DNA	Deoxiribose Nucleaic Acid
RNA	Ribose Nucleaic Acid
cDNA	complementary DNA
mRNA	messanger RNA
SNP	Single Nucleotide Polymorphism
FISH	Fluorescent *in situ* Hybridization
MBN	Mung Bean Nuclease
EDTA	ethylenediaminetetraacetic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
X-gal	5-bromo-4-chloro-3-indolyl-D-galactopyranoside
Amp	Ampicillin
LB	Luria Bertani
LBA	Luria Bertani Agar
SOC	Super Optimal broth with Catabolite repression
SDS	Sodium Dodecyl Sulphate
CFU	Colony Forming Unit
$MgCl_2$	magnesium chloride salt
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
RDA	Representational Differences Analysis
CGH	Comparative Genomic Hybridization
HSU	Heterogametic Sex Unidentified group
HFI	Heterogametic Female Identified group
RE	Restriction Endonuclease
BSA	Bovine Serum Albumin
TSAP	Thermosensitive Alkaline Phosphatase
pH	fluid acidity/alkali indicator
g	gravitational force

TE	Tris EDTA buffer
TAE	Tris, Acetate and EDTA buffer
hpH_2O	high purity water
°C	Degrees Celsius
♀	female
♂	male
pg / ng / µg / mg / g	picogram / nanogram / microgram / milligram / gram
mM / M	millimolar / molar
µl / ml / L	microliter / millilitre / litre
V / kV	volt / kilovolt

Table of Contents

DECLARATION II

ANIMAL ETHICS II

ACKNOWLEDGEMENTS III

ABSTRACT IV

GLOSSARY OF TERMS VI

CHAPTER 1 - THE EVOLUTION OF SEX DETERMINATION MECHANISMS IN AMNIOTES 1

Sexual Reproduction 1

Sex Determination 1

Mammalian Sex Determination 5

Discovering the Mammalian Testis Determining Factor; SRY 6

Sex Determining Region Y (SRY) 8

Monotreme Sex Determination 9

Avian Sex Determination 11

Avian Sex Determining Genes 11

Reptile Sex Determination 15

Turtles 16

Lizards 17

Snakes 17

Snake Cytology 18

Snake Genetics 21

Conclusion 27

Thesis Hypothesis Aims and Objectives. 28

CHAPTER 2 - RANDOM AMPLIFIED POLYMORPHIC DNA ANALYSIS 32

Introduction 32

Materials and Methods 36

Snake Slough DNA Extraction 36

RAPD PCR and Visualisation 37

Investigation of Sex Associated RAPD Fragments 40

Results 40

RAPD Analysis 42

Discussion 49

Random Amplified Polymorphic DNA Analysis *50*
RAPD Primer OPA-17 *53*
CONCLUSION 56

CHAPTER 3 - REPRESENTATIONAL DIFFERENCE ANALYSIS **58**

INTRODUCTION 58
RDA Protocol Preparation *59*
RDA Hybridisation *59*
MATERIALS AND METHODS 61
Genomic DNA Preparation *62*
Tester and Driver Amplicon Preparation *64*
Hybridisation Round One *66*
Hybridisation Round Two to Six *67*
RDA Enriched DNA Cloning *67*
Genomic DNA Investigation *69*
RESULTS 72
Male and Female Enriched DNA Fragments *81*
Female Hind III - Insert 1 *82*
Male Bgl II - Insert 2 *84*
RDA Fragments DNA Secondary Structures *85*
DISCUSSION 87
Representational Difference Analysis *90*
Group 1 – Repetitive Elements *91*
Female Hind III – Insert 1 92
Male Bgl II – Insert 2 94
Group 2 - Coding and Noncoding Region of the Genome *96*
CONCLUSION 104

CHAPTER 4 - COMPARATIVE GENOMICS **106**

INTRODUCTION 106
MATERIALS AND METHODS 111
DNA Extraction *111*
Comparative Genomic Analysis *111*
Primer Design *114*
CTNNB1 Gene *116*

PCR optimisation *117*
DNA Sequencing *118*
Primer Re-design *119*
RESULTS 120
Amphiphysin (AMPH) Gene *124*
Phosphatidylinositol-5-Phosphate 4-Kinase Type 2 Alpha (PIP4K2A) Gene *124*
Tax1 Binding Protein (TAX1BP1) Gene *125*
Genomic Sequencing Results *126*
Catenin Beta 1 (CTNNB1) Gene *127*
RAB5A, Member RAS Oncogene Family (RAB5A) Gene *134*
DISCUSSION 138
Comparative Genomics *139*
CTNNB1, Catenin Beta 1 *140*
CONCLUSION 142

CHAPTER 5 - COSMID LIBRARY CONSTRUCTION AND SECOND GENERATION SEQUENCING 144

INTRODUCTION 144
MATERIAL AND METHODS 146
Genomic DNA preparation *147*
SuperCos1 Vector Preparation *148*
Ligation and Packaging of Genomic DNA *149*
Preparation of Host Strain *150*
Cosmid Library Titre, Amplification, and Storage *150*
Genomic Library Assessment *151*
Identifying CTNNB1, RAB5A and WAC Positive Cosmid Clones *153*
Roche 454 Sequencing *154*
Ion Torrent Sequencing *156*
Sequence Analysis *157*
RESULTS 157
Roche 454 Sequencing *158*
Cosmid Clone - C1 162
Cosmid Clone - C2 162
Cosmid Clone - C3 163
Cosmid Clone - C5 164
Cosmid Clone - W1 165

Cosmid Clone - W2 165
Cosmid Clone - R2 166
Ion Torrent Sequencing *166*
Cosmid Clone - 4 169
Cosmid Clone - 6 170
DISCUSSION 171
Cosmid Library Screen and Clone Selection *171*
Roche 454 Sequencing *173*
Ion Torrent Sequencing *176*
Conclusion *179*
CHAPTER 6 - SYNOPSIS **181**
RANDOM AMPLIFIED POLYMORPHIC DNA ANALYSIS 182
REPRESENTATIONAL DIFFERENCE ANALYSIS 183
COMPARATIVE GENOMICS 186
COSMID GENOMIC LIBRARY 187
LIMITATION OF THE THESIS 188
IMPACT OF FINDINGS AND RECOMMENDATIONS FOR FUTURE RESEARCH 189
REFERENCES **191**
APPENDIX A: REAGENT RECIPES **207**
APPENDIX B: DNA SEQUENCES DATA **210**

List of Figures

Figure 1-1 The Distribution of Vertebrate Sex Determination Systems 4
Figure 1-2 Evidence for Mechanisms of Snake Sex Determination 21
Figure 1-3 A comparative cytogenic map of chromosomes 1-7, Z and W chromosomes of E. quadrivirgata (Matsubara K. et al. 2006) 23
Figure 1-4 Morelia spilota imbricata (Southwest Carpet Python) 31
Figure 2-1 Morelia spilota imbricata after sloughing 35
Figure 2-2 Snake slough DNA extraction 42
Figure 2-3 RAPD fragment profile of OPE-05 and OPD-13 43
Figure 2-4 RAPD fragment profile OPA-17 45
Figure 2-5 Diagrammatic and sequence representation DNA fragment produced by OPA-17 47
Figure 2-6 RAPD OPA-17, 17 M. spilota templates of previously determined sex 49
Figure 3-1 A simplified representation of the RDA enrichment process 60
Figure 3-2 Comparison of RDA protocol after hybridisation round 3 73
Figure 3-3 RDA hybridisation round 3 with increased PCR cycles 74
Figure 3-4 Modified RDA protocol hybridisation rounds 1 to 6 76
Figure 3-5 DNA secondary structure Female Hind III - Insert 1 87
Figure 3-6 Tester DNA secondary structure self-priming and RDA enrichment 94
Figure 4-1 cDNA map of Z and W chromosomes of three snake species 107
Figure 4-2 cDNA map of the Z chromosome of E. quadrivirgata 108
Figure 4-3 Multiple sequence alignment of reference EIF1 gene sequence 113
Figure 4-4 CTNNB1 sequence contigs from M. spilota and P. bivittatus, using an A. carolinensis reference assembly 130
Figure 4-5 Multiple sequences alignment, contig 4 and CTNNB1 sequences 133
Figure 4-6 RAB5A mononucleotide microsatellite polymorphism, one male one female 135
Figure 4-7 RAB5A mononucleotide microsatellite polymorphism, 7 animals 136
Figure 4-8 Representation polymorphic microsatellite sequences composition 137
Figure 5-1 SuperCos 1 vector (Agilent Technologies) 145
Figure 5-2 - Cosmid Clone - C1 Reference Assembly 162
Figure 5-3 - Cosmid Clone - C2 Reference Assembly 163
Figure 5-4 - Cosmid Clone - C3 Reference Assembly 164
Figure 5-5 - Cosmid Clone - C5 Reference Assembly P. bivittatus scaffolds 7960 164
Figure 5-6 - Cosmid Clone - C5 Reference Assembly P. bivittatus scaffolds assembly 9586 165
Figure 5-7 - Cosmid Clone – W2 Reference Assembly 166

Figure 5-8 - Cosmid Clone - R2 Reference Assembly *166*

Figure 5-9 ECOR1 digest of cosmid clones 1 through 8 *167*

Figure 5-10 - Ion Torrent Cosmid Clone 4 *170*

Figure 5-11 - Ion Torrent Cosmid Clone 6 *171*

List of Tables

Table 2-1 RAPD primers 38

Table 2-2 RAPD fragments identified 44

Table 2-3 RAPD primers investigated in published research 52

Table 3-1 RDA adaptors 64

Table 3-2 Reaction conditions for hybridisation rounds 67

Table 3-3 Male and female RDA insert primers 70

Table 3-4 Male and female RDA insert optimised PCR conditions 72

Table 3-5 RDA enriched DNA regions RepeatMasker and BLAST results 78

Table 3-6 RDA enriched DNA regions NCBI P. bivittatus annotation release 101 BLAST results 80

Table 3-7 Genomic PCR products sequenced 82

Table 3-8 Sequence similarity between Female Hind III - Insert 1 and TY3/Gypsy sequences 84

Table 3-9 BLAST results Male Bgl II - Insert 2 85

Table 3-10 DNA secondary structure results 86

Table 3-11 Human chromosome location of gene localised to snake Z and W chromosome 101

Table 3-12 Human chromosome location of RDA sequence gene homologues 102

Table 4-1 Genes mapped to E. quadrivirgata Z chromosome 109

Table 4-2 Primer pairs for gene targets 115

Table 4-3 Primer pairs targeting the CTNNB1 gene 117

Table 4-4 Primer pairs redesigned 119

Table 4-5 Primer pairs, optimised PCR conditions 121

Table 4-6 Gene target sequencing results 123

Table 4-7 P. bivittatus BLAST matches 2013 data 127

Table 4-8 CTNNB1 primer pairs, optimised PCR conditions 128

Table 4-9 CTNNB1 DNA sequencing results 129

Table 4-10 CTNNB1 M. spilota and P. bivittatus contig assembly 129

Table 4-11 CTNNB1 M. spilota and P. bivittatus genomic DNA sequence alignment 132

Table 5-1 PCR primers of three reference genes 152

Table 5-2 Roche 454 library oligonucleotide adaptors 155

Table 5-3 PCR results from cosmid amplification pools 158

Table 5-4 Cosmid clones T3 and T7 sequencing 159

Table 5-5 Roche 454 sequence data 161

Table 5-6 Ion Torrent sequence data 168

Table 5-7 Comparison after sequence trimming 169

Table 5-8 Gene with sexual function *184*

Chapter 1 - The Evolution of Sex Determination Mechanisms in Amniotes

"The work on genetics of reptilian sex determination has thus been predicated on discoveries of mammalian sex-determining genes. Extrapolations from those discoveries to look at gene expression in reptiles have not generalized well. Furthermore, if the set of genes involved in reptilian sex determination is not entirely a subset of genes involved in mammalian sex determination, then an approach that simply borrows discoveries from mammals will ultimately fail to discover the full reptilian mechanism." James J. Bull, 2004 (Valenzuela and Lance 2004).

Sexual Reproduction

Sexual reproduction is a complex biological process that involves the fusion of a male (sperm) and a female (ovum) gamete to produce a zygote. From a genetic perspective this process results in the rearrangement and combining of half the genetic information from each parent, creating a new, genetically different offspring (Barton and Charlesworth 1998). This genetic exchange between two individuals is known as genetic recombination and is a key adaptive advantage of sexual reproduction over other forms of reproduction (Barton and Charlesworth 1998).

In the majority of cases, we can assume that a phenotypic trait that gives a great or important adaptive advantage to an organism, will be highly conserved (Valenzuela and Lance 2004). Given the adaptive significance of sexual reproduction and its prevalence among vertebrates, it may well be assumed that the mechanism by which sex is determination is also highly conserved (Valenzuela and Lance 2004). However in reality we observe the opposite, with sex being determined by an enormous variety of different, seemingly unrelated mechanisms, which have evolved independently (Valenzuela and Lance 2004).

Sex Determination

Sex determination is a biological mechanism, which results in the differentiation of offspring into either a male or female phenotype (Hadany and Beker 2007). Amongst vertebrates many sex determination systems exists (Ezaz, Rami Stiglec, et al. 2006; Ferguson-Smith 2006; Matsubara et al.

2006; Vallender and Lahn 2006). Nevertheless they can be divided into two major groups: Genetic Sex Determination and Environmental Sex Determination (Bull 1980; Ezaz, Rami Stiglec, et al. 2006; Quinn et al. 2007). Although these two groups are fundamentally different, they are not mutually exclusive, with research showing that ESD and GSD are intimately linked and are affected by the same selection pressures (Bull 1983; Radder et al. 2007).

Genetic Sex Determination (GSD) relies on a genetic switch that directs sexual differentiation towards either a male or female phenotype (Bull 1980; Modi and Crews 2005; Ezaz, Rami Stiglec, et al. 2006; Matsubara et al. 2006; Vallender and Lahn 2006). Although multiple genes may be involved in the production of a male or female phenotype, only one gene, the primary sex determining gene, is the trigger that initiates the cascade of events, which ultimately lead to sexual differentiation. Therefore, under GSD the sex of an offspring is determined at the moment of fertilisation, though the primary sex determining gene may have little or no influence until later embryonic development.

Genetic Sex Determination (GSD) mechanisms, most commonly take the form of a single pair of sex chromosomes (Ezaz, Rami Stiglec, et al. 2006). Paired sex chromosome system can occur as male heterogametic (XY males and XX females) as seen in mammals or as female heterogametic (ZZ males and ZW females) as seen in birds (Ezaz, Rami Stiglec, et al. 2006). Among vertebrates male heterogamety (XX/XY) occurs at about twice the frequency of female heterogamety (ZZ/ZW) (Ezaz, Rami Stiglec, et al. 2006). Although both male and female heterogametic sex chromosome systems have arisen independently several times in vertebrates (Ezaz, Rami Stiglec, et al. 2006).

GSD can also manifest as a dose dependent single sex chromosome system (XX/XO), examples of which have been found in insects, fish and amphibians (Valenzuela and Lance 2004; Veyrunes et al. 2008). Alternatively, multiple pairs of sex chromosomes can be present, as observed in egg laying mammals; Monotremes (Bull 1980; Pask and Graves 1999; Grutzner et al. 2004; Veyrunes et al. 2008). It is important to note that the nomenclature of sex chromosomes, that is XX/XY, ZZ/ZW, XX/XO etc., does not refer to any orthology (or lack of) between different sex chromosome systems, but only indicates the sex chromosome arrangement.

The evolution of a new GSD system begins with one chromosome of an autosomal pair acquiring a male or a female sex determining allele (Muller 1914). Provided that this sex determining allele is favorable, chromosomal recombination between this autosomal pair will be suppressed at this new sex determining locus (Muller 1914; Charlesworth 1991). In the case of XX/XY sex chromosome

systems, the proto-Y chromosome acquired a male sex determining allele, resulting in a non-recombining region forming to encompass that allele. In the case of a ZZ/ZW sex chromosome system, the opposite has occurred, with the proto-W chromosome acquiring a female sex determining allele and the formation of the subsequent non-recombining region.

Over evolutionary time the sex chromosome pairs become increasingly differentiated. While the X and Z chromosomes remain relatively unchanged, as they can still undergo chromosomal recombination in at least one sex (Charlesworth 1991). Once a Y or W chromosome acquires a non-recombining region, any mutation acquired in that region cannot undergo normal recombination repair. Consequently, this region is subject to large deletions, sequence mutations and the proliferation of repetitive elements. Over time these mutations lead to the non-recombining region of the sex chromosome spreading along the length of the chromosome, introducing repetitive elements and deleting DNA sequences, interrupting functional genes as it grows (Charlesworth 1991). This process ultimately leads to a Y or W chromosome that is dramatically reduced in size, with very few functioning genes and abundant repeating elements (Charlesworth 1991). Furthermore it can eventually lead to the loss of the Y or W chromosome completely, provided a substitute can be found for the original sex determining allele (Charlesworth 1991).

The substitution of a primary sex determining gene is one way in which a sex determination transition can occur. A sex determination transition is an exchange of one sex determination system for another and the process has occurred many times, independently, throughout vertebrate evolution (Sarre, Ezaz, and Georges 2011). Although very interesting, this unusual phenomenon will not be discussed in detail here because it is outside the scope of this thesis, but it has been reviewed in depth recently by Sarre *et al.* (Sarre, Ezaz, and Georges 2011).

Environmental Sex Determination (ESD) uses a non-genetic environmental signal in order to initiate sexual differentiation (Bull 1980; Fredric J Janzen and Paukstis 1991; Pask and Graves 1999; Modi and Crews 2005; F. J. Janzen and Phillips 2006). This is usually experienced by the developing embryo within a specific time period after conception (Bull 1980; Fredric J Janzen and Paukstis 1991; Pask and Graves 1999; Modi and Crews 2005; F. J. Janzen and Phillips 2006). Examples of ESD systems have been documented in amphibians, fish and reptiles and can be influenced by a multitude of environmental triggers including; temperature, pheromones, photoperiod length and water pH (Bull 1983; Shine 1999). Temperature-dependent Sex Determination (TSD) is the most common form of ESD documented in vertebrates (Bull 1980; Ezaz, Rami Stiglec, et al. 2006; Quinn et al. 2007).

Temperature-dependent Sex Determination (TSD) is hypothesised to be the first sex determination mechanism from which all others have evolved (Valenzuela and Lance 2004). The sexes of animals utilizing this mechanism are genetically indistinguishable (Bull 1980; Ezaz, Rami Stiglec, et al. 2006; Shoemaker and Crews 2009). The process relies on specific temperature thresholds, during specific periods of embryonic development, to initiate sexual differentiation (Bull 1980; Ezaz, Rami Stiglec, et al. 2006; Shoemaker and Crews 2009). TSD systems have been found in the last two remaining species of tuatara, all crocodilian species studied and is found sporadically in species of fish, lizards and turtles (Bull 1980; Ezaz, Rami Stiglec, et al. 2006; Shoemaker and Crews 2009; Silber 2011).

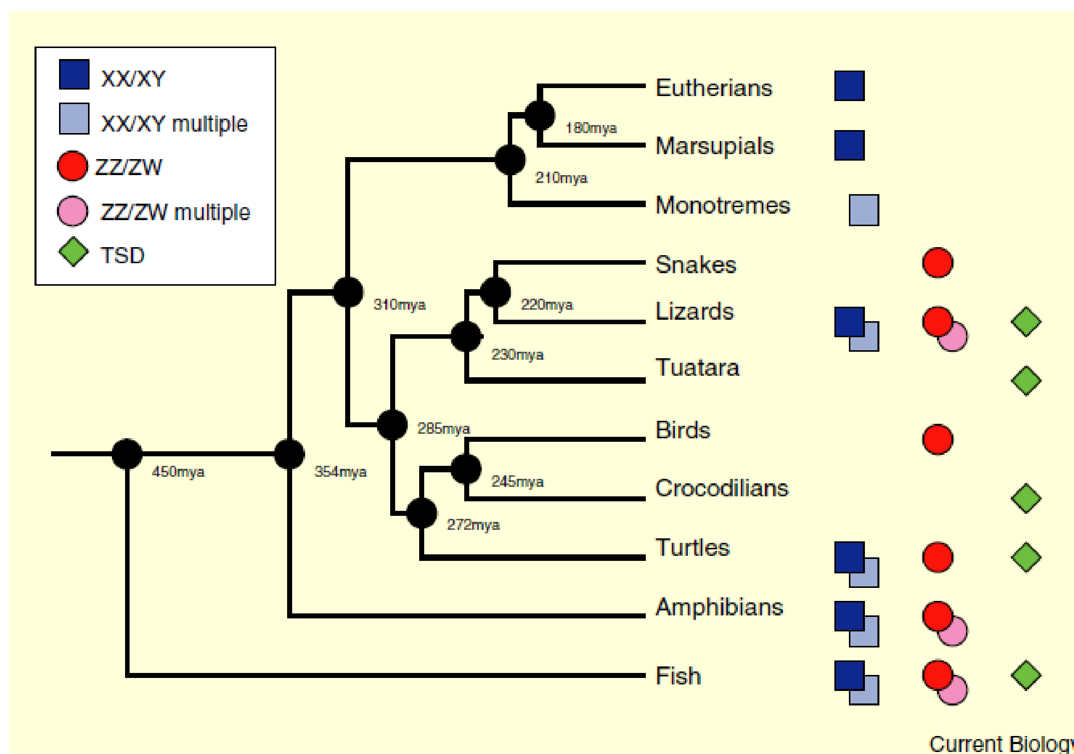

Figure 1-1 The Distribution of Vertebrate Sex Determination Systems (Ezaz, Rami Stiglec, et al. 2006)

Figure 1-1 shows the sex determination systems represented in the major vertebrate groups and the predicted times at which they diverged (Ezaz, Rami Stiglec, et al. 2006). These sex determination systems have been proven to occur in at least one member of the respective clade (Ezaz, Rami Stiglec, et al. 2006). XX/XY multiple ZZ/ZW multiple indicate the involvement of more than one pair of sex chromosomes in sex determination system (Ezaz, Rami Stiglec, et al. 2006).

Mammalian Sex Determination

Phylogenetically, living mammals can be split into two distinct subclasses; Prototheria, which comprises egg laying Monotremes and Theria, which comprises two infraclasses Metatherian (marsupial) and Eutherian (placental) mammals. Prototherian and Therian mammals use distinctly different sex determination systems and because of this they will be discussed separately.

Therian mammals have two sex chromosomes: X and Y. This GSD system is male heterogametic, with XX females and XY males. Evidence suggests that the X and Y chromosome of Therian mammals originated from a pair of autosomes approximately 180 million years ago (Wallis, Waters, and Graves 2008). However, over these 180 million years of evolution the X and Y chromosome have become highly differentiated.

The most extensively studied sex determination system of any Therian mammal, or vertebrate for that matter, is *Homo sapien*. The human X chromosome contains about 155Mb of DNA and at last count about 1300 genes, which include both housekeeping genes and genes with sex-related functions (Ross et al. 2005). The human Y chromosome contains about 60Mb of DNA, one third of its estimated original autosomal size (Skaletsky et al. 2003). It contains far fewer functioning genes than the X chromosome and has abundant repeating elements (Charlesworth 1991; Ezaz, Rami Stiglec, et al. 2006). Of the potentially 1300 genes, which were once located on the ancestral Y chromosome, only 45 have survived (Wallis et al. 2007; Wallis, Waters, and Graves 2008). Within the male-specific region (non-recombining region of chromosome Y) there are 156 transcribed sequences, but only 27 are translated into proteins (Skaletsky et al. 2003). Many of the 27 genes remaining have now evolved functions associated with male reproduction, causing them to undergo positive selection pressure, which has prevented their deletion from the Y chromosome (Wallis et al. 2007).

Comparisons of the X chromosome among different species of placental mammals show that it is highly conserved, with respect to both gene content and gene order, which is almost identical in distantly related species (Ohno 1966; O'Brien et al. 1999). The only exception is the rodent family Muridae, which still has conserved gene content, but has a different gene order (Soullier et al. 1998; Arakawa et al. 2002). In contrast, comparison of placental Y chromosomes shows that different mammals have retained a different subsets of genes, from those that existed in the ancestral autosome, from which the Y chromosome originated (Graves 2006).

The X and Y chromosome of marsupial mammals are generally smaller than that of placental mammals (Alsop et al. 2005). Comparative studies have shown that the X chromosome of the *Macropus eugenii* (Tammar wallaby), show it is homologous over a large region, with the human X chromosome, including the whole of the long arm and the pericentric region (Alsop et al. 2005). However the remainder of the short arm of the human X chromosome displays homology to chromosome 5 in *M. eugenii* (Alsop et al. 2005; Graves 2008). This indicates that most of the placental X and marsupial X is conserved between the two infraclasses. However, there has been a region added to the X chromosome of placental mammals, which was once a part of chromosome 5 and remains as such in marsupials (Graves 2008).

The Y chromosomes of humans and *M. eugenii* have five active genes in common (Murtagh, Waters, and Graves 2010). Homologs of thirteen other genes from the human Y chromosome have been identified on autosomal chromosomes of the *M. eugenii* (Murtagh, Waters, and Graves 2010). In addition, unique protein coding sequences have also been found on the Y chromosome of the *M. eugenii* (Murtagh, Waters, and Graves 2010). This evidence indicates that just like different species of placental mammals, marsupial mammals have retained a different subset of genes, which were once located on the ancestral Y chromosome. There is in fact only one gene that is conserved on the Y chromosome of almost all Therian mammals; Sex Determining Region Y (*SRY*).

Discovering the Mammalian Testis Determining Factor; *SRY*

In 1959 Welshons and Russell first hypothesised that the Y chromosome was responsible for producing a male phenotype in humans (Welshons and Russell 1959). The unknown determinant located on the Y chromosome was named the human Testis Determining Factor (TDF). We now know that the Y chromosome of Therian mammals contains a TDF known as the Sex Determining Region Y (*SRY*) gene. However, *SRY* wasn't discovered until 1990, 31 years after Welshons and Russell's hypothesis.

The very first mechanism hypothesised to be the mammalian TDF, responsible for sex determination in mammals, was the H-Y antigen. Researchers observed that female mice immunized against isogenic male mouse cells, rejected skin grafts from those same male mice (Muller U 1996). This observation was attributed to an undiscovered male-specific antigen, H-Y antigen, which therefore was predicted to be encoded on the Y chromosome (Muller U 1996). Mapping experiments indicated that the gene responsible for the cell surface H-Y antigen in humans was located on the Y chromosome, which supported this hypothesis (Muller U 1996). However, the H-Y antigen was

disproved to be the mammalian TDF after it was discovered that Turner syndrome patients (XO females), lacking a Y chromosome were still H-Y antigen positive (Haseltine FP et al. 1982). Furthermore, investigation in mice revealed that testis would differentiate with or without the presence of the H-Y antigen (Haseltine FP et al. 1982).

In 1975 Singh and Ray-Chaudhuri, discovered a repeating DNA sequence made-up of GATA and GACA tandem repeats from the W chromosome of *Bungarus caeruleus* (Common krait), which they named Banded krait minor-satellite (Bkm) (Singh and Ray-Chaudhuri 1975). Then in 1982 a repetitive Murine DNA sequence that showed similarity to Bkm was discovered to be specific to male mice. It was subsequently hypothesized that the Bkm-like repeat was the TDF in mammals (Epplen et al. 1982). However, no similar repeats could be found in humans and the hypothesis was discarded (Kiel-Metzger et al. 1985).

In 1987 a 280kb region of the human Y chromosome was discovered to be missing from a sex reversed XY female (Page et al. 1987). From this 280kb region the Zinc Finger Y-chromosomal protein (*ZFY*) gene was isolated (Page et al. 1987). Previous studies had shown that other zinc finger proteins bind and interact with nucleic acids, affecting transcription (Page et al. 1987). Therefore the *ZFY* gene was hypothesised to be the TDF of humans (Page et al. 1987). However, additional research showed that the *ZFY* gene is ubiquitously expressed and it was also mapped to an autosome in marsupials, which proved it could not be the TDF of all Therian mammals (Sinclair et al. 1988; Palmer et al. 1989).

Sex Determining Region Y (*SRY*)

It wasn't until 1990 that the true human TDF, Sex Determining Region Y (*SRY*), was discovered simultaneously by two different research groups (Jagar et al. 1990; Sinclair et al. 1990). This was achieved by analyzing Y chromosome deletions from sex reversed individuals, leading to the identification of a common 35kb region, located on the short arm of the Y chromosome (Jagar et al. 1990; Sinclair et al. 1990). Within this region a significant open reading frame was found, which encoded the *SRY* gene (Jagar et al. 1990; Sinclair et al. 1990). Subsequent studies included: investigating the presence of *SRY* in sex reversed mice, identifying point mutations within the *SRY* gene of human XY females, the expression of *SRY* in somatic cells only, and reversing the sex of XX mice by inserting a single copy of the *SRY* gene (Koopman et al. 1990, 1991; Harley et al. 1992; Hawkins et al. 1992; Rossi et al. 1993). All of these studies contributed to the conclusion that the *SRY* gene is the TDF of humans and mice (Koopman et al. 1990, 1991; Harley et al. 1992; Hawkins et al. 1992; Rossi et al. 1993). Since 1990, *SRY* gene homologues have been identified in almost all placental mammals studied and an *SRY* homologue has also been mapped to the Y chromosome of marsupial mammals, proving that the *SRY* gene is the primary sex determining gene in nearly all Therian mammals (Sinclair et al. 1990; Foster and Graves 1994; Wallis et al. 2007).

Ellobius lutescens, *Ellobius tancrei* and *Ellobius talpinus* are three species of small rodent commonly known as mole voles and *Tokudaia osimensis* a species of spiny country-rats, all of which do not use the common Therian XX/XY sex determination system and the *SRY* gene (Walter Just et al. 1995; Sutou, Mitsui, and Tsuchiya 2001; W. Just et al. 2007). These four Therian mammals all belong to the rodent family Muridae, but no longer have a Y chromosome or an *SRY* gene (Walter Just et al. 1995; Sutou, Mitsui, and Tsuchiya 2001; W. Just et al. 2007). It is believed that another gene, possibly located on an autosome has evolved a new sex determining function (Walter Just et al. 1995; Sutou, Mitsui, and Tsuchiya 2001; W. Just et al. 2007). This has allowed the deletion of the *SRY* gene and the eventual loss of the entire Y chromosome (Walter Just et al. 1995; Sutou, Mitsui, and Tsuchiya 2001; W. Just et al. 2007). The result is a new, XX female and XO male, sex determination system (Walter Just et al. 1995; Sutou, Mitsui, and Tsuchiya 2001; W. Just et al. 2007). No sex determining gene has yet been found for any of these four species of rodent.

The action of *SRY* has been attributed to an 80 amino acid domain called the HMG-box, which shows moderate similarity to high mobility group proteins. The HMG-box is the only region of the *SRY* gene that is conserved among mammals of different species (Foster et al. 1992). The HMG-box affects transcription of other genes by binding to a six base pair region (A/T)ACAA(T/A) of DNA and bending

it to an angle of between 85-90° (Ferrari et al. 1992; Harley et al. 1992). It is hypothesized that the *SRY* protein binds to the testis differentiating gene SRY-box 9 (*SOX9*), up-regulating its transcription, leading to the formation of testes and ultimately a male phenotype (Vidal et al. 2001). However, others have suggested that the non-conserved nature of the whole *SRY* gene, across different species of mammals, and the discovery of human XX males with no *SRY* gene, contradicts this hypothesis (McElreavy et al. 1993). Consequently an alternative hypothesis has been formulated, which suggests that *SRY* acts to inhibit a secondary unknown factor that represses male development and this secondary unknown factor could potentially be the SRY-box 3 (*SOX3*) gene, which is the gene from which *SRY* is thought to have originally evolved (McElreavy et al. 1993; Graves 1998).

Monotreme Sex Determination

There are only five species still extant today that belong to the subclass of mammal Prototheria: the last subclass of mammals yet to be discussed. This includes one species of platypus (*Ornithorhynchus anatinus*) and four species of echidna (*Tachyglossus aculeatus*, *Zaglossus attenboroughi*, *Z. bartoni* and *Z. bruijini*), which all belong to the order Monotremata (Wallis, Waters, and Graves 2008). Monotremes have an extremely unusual sex determination system made up of multiple pairs of sex chromosomes (Grutzner et al. 2004; Rens et al. 2004). In the platypus five pairs of X chromosomes ($X_1X_1X_2X_2X_3X_3X_4X_4X_5X_5$) produce a female phenotype, while five X chromosomes and five Y chromosomes ($X_1Y_1X_2Y_2X_3Y_3X_4Y_4X_5Y_5$) produce a male phenotype (Grutzner et al. 2004; Rens et al. 2004). In echidnas five pairs of X chromosomes ($X_1X_1X_2X_2X_3X_3X_4X_4X_5X_5$) produce a female phenotype, while five X chromosomes and four Y chromosomes ($X_1Y_1X_2Y_2X_3Y_3X_4Y_4X_5$) produce a male phenotype (Rens et al. 2007). The reason for this strange, overly complicated, multiple sex chromosome system is unknown and it has never been observed, to this extreme, in any other vertebrate. The only other examples of a sex determination system that has greater than four sex chromosome pairs, occurs in species of invertebrates and plants (Wallis, Waters, and Graves 2008).

It was originally thought that the sex chromosome system of monotremes diverged from a common system belonging to all mammals. Supporting evidence for this theory came from a study using *in situ* hybridization and showed DNA-probes specific for 8 genes, common to both placental and marsupial X chromosomes, could be mapped to the X1 chromosome of the platypus (Watson et al. 1990). However, this theory was disproved, when a more comprehensive study revealed that the X chromosome region, common to both placental and marsupial mammals maps to an autosome in the platypus (Waters et al. 2005). Furthermore, no orthologue of the *SRY* gene has been found in

monotremes (Grutzner et al. 2004). The ancestor of the *SRY* gene, *SOX3*, has been mapped to an autosomal chromosome in both the platypus and echidna (Wallis et al. 2007). This evidence strongly suggests that the Therian and Monotremata sex determination systems evolved independently of each other (Wallis et al. 2007). Therefore, the Therian sex determination system, including the X and Y chromosomes and the sex determining *SRY* gene, most probably originated after Monotremata and Therian mammals diverged 166 million years ago, but before placental and marsupial mammals diverged approximately 148 million years ago (Wallis, Waters, and Graves 2008).

Surprisingly, monotreme sex chromosomes have considerable homology with bird sex chromosomes (Grutzner et al. 2004; El-Mogharbel et al. 2007; Rens et al. 2007; Veyrunes et al. 2008; Warren et al. 2008). A number of studies have revealed that the Z chromosome of chickens shares homology with regions belonging to platypus chromosomes X1, Y1, X2, X3 and a large proportion of X5 (Grutzner et al. 2004; El-Mogharbel et al. 2007; Rens et al. 2007; Veyrunes et al. 2008; Warren et al. 2008). With the homologous region on the X5 chromosome including a homologue of the gene Doublesex and mab-3 related transcription factor 1 (*DMRT1*): the most promising candidate for the primary sex determining gene in birds (El-Mogharbel et al. 2007). Monotremes and birds are thought to have diverged approximately 310 million years ago (Wallis, Waters, and Graves 2008; Kumar and Hedges 1998). This information has led scientists to hypothesise that a sex determination system that is ancestral to all amniotes may possibly be conserved in both monotremes and birds (Wallis, Waters, and Graves 2008).

Although very little is known about the sex determining genes of monotremes, *DMRT1* and GATA binding protein 4 (*GATA4*) are two gene candidates identified as the possible primary sex determining gene (Wallis, Waters, and Graves 2008). In the platypus, *DMRT1* has been mapped to the X5 chromosome and *GATA4* has been mapped to both X1 and Y1 chromosomes (El-Mogharbel et al. 2007; Grafodatskaya et al. 2007). Both genes are known to bind to DNA and affect its transcription (El-Mogharbel et al. 2007; Grafodatskaya et al. 2007). However, there is currently no other information known about either monotreme gene. It is also important to note that because only the genome of a female platypus has ever been completely sequenced, the five Y chromosomes, which make up approximately 6% of the male genome, are only represented by studies involving gene mapping of the X chromosome homologues onto the five Y chromosomes. Furthermore, if a new undefined gene on any one of the Y chromosomes is responsible for sex determination in monotremes, as *SRY* is in Therian mammals, it will continue to be overlooked by studies adopting this approach.

Avian Sex Determination

All birds have two sex chromosomes; a Z and a W. In this sex determination system females are the heterogametic sex, containing one Z and one W chromosome and males are homogametic with two copies of the Z chromosome. These sex chromosomes are believed to have evolved from a pair of autosomes, as did the X and Y of mammals. Comparative gene mapping between human and chicken sex chromosomes shows no homology, indicating that the Therian and Avian sex chromosome systems have evolved independently of each other (Fridolfsson et al. 1998; I. Nanda et al. 2002).

Over time the Z and W chromosomes of birds have become differentiated, due to the phenomenon known as W chromosome degradation. However, the rate at which this divergence has occurred varies among different bird lineages (Takagi, Itoh, and Sasaki 1972; Takagi and Sasaki 1974). The sex chromosomes of the most modern bird species are heteromorphic, characterised by a large Z chromosome, and a relatively small heterochromatic W chromosome (Takagi and Sasaki 1974; Ezaz, Rami Stiglec, et al. 2006; Vallender and Lahn 2006). In contrast, the Z and W chromosome of Ratites have diverged very little, appearing almost identical (Takagi, Itoh, and Sasaki 1972). Despite the differences in W chromosome differentiation, the Ratite Z chromosome is homologous with the Z chromosome of the chicken over its entire length (Swathi Shetty, Griffin, and Graves 1999). This suggests that the bird sex chromosome system most probably evolved before Ratites diverged from modern birds over 80 million years ago (Swathi Shetty, Griffin, and Graves 1999).

Gallus gallus domesticus (domesticated chicken) has become the model bird species for investigating the Avian sex determination system. The chicken Z chromosome contains 75mb of DNA with 840 genes, according to the most recent construct of the chicken genome (Fulton et al. 2004). In comparison the W chromosome is less than half the size of the Z, with few functioning genes and is highly heterochromatic (Fulton et al. 2004; McQueen and Clinton 2009). However, the W chromosome is very poorly represented by the chicken genome construct, with only 260kb included, less than 1% of the annotated sequence, within which only 4 genes have been identified (Fulton et al. 2004; McQueen and Clinton 2009).

Avian Sex Determining Genes

The primary sex determining gene of birds is yet to be conclusively identified and hence the genetic sex determination mechanism is also unknown. Currently there are two hypothesised mechanisms for bird sex determination. The first is a single female dominant gene located on the W chromosome

(McQueen and Clinton 2009). This gene would act similarly to that of the mammalian *SRY* gene, but in reverse, producing a female (ZW) phenotype in its presence and a male (ZZ) phenotype in its absence (McQueen and Clinton 2009). The second hypothetical mechanism is a gene located on the Z chromosome, which would differentiate a male and female phenotype by a dose dependent mechanism (McQueen and Clinton 2009). Under this mechanism one copy of the gene would produce a female (ZW) phenotype, while two copies of the gene would produce a male (ZZ) phenotype (McQueen and Clinton 2009).

The very first primary sex determining gene discovered was the *SRY* gene of humans (Jagar et al. 1990; Sinclair et al. 1990). This gene was identified by investigating human individuals with unusual sex chromosome combinations and deletions (Jagar et al. 1990; Sinclair et al. 1990). Consequently, this approach was originally adopted in the hope of identifying a similar sex determining factor in birds. However after intensive karyotype investigation, very few birds with sex chromosome aberrations could be identified (Clinton 1998). No ZO bird has ever been identified, indicating that this sex chromosome combination is most probably lethal to embryos (Graves 2003). Only one ZZW bird has ever been reported in the literature: a chicken in 1933 (Crew 1933; Ellegren 2001). Therefore, this approach was abandoned. Nowadays molecular genetics techniques have been adopted, primarily focusing on the differences observed in gene expression within the gonads of chicken embryos beginning sexual differentiation.

Doublesex and mab-3 related transcription factor 1 (*DMRT1*) is the top candidate for being the primary sex determining gene in birds. *DMRT1* was first implicated in sex determination in 1999, when a sex reversed XY human was discovered with a single deletion from the p-arm of chromosome 9 (Raymond, Parker, et al. 1999). This deletion included the *DMRT1* gene and for the first time implied its crucial role in testis differentiation in humans (Raymond, Parker, et al. 1999; Raymond, Kettlewell, et al. 1999). Research had also shown that human chromosome 9 is an orthologue of the chicken Z chromosome and so it was suggested that *DMRT1* might play a role in bird sex determination (Indrajit Nanda et al. 1999). Investigation into this hypothesis showed that *DMRT1* was not only located on the Z chromosome of chicken, but it was conserved among birds of distantly related lineages including one ratite; *Dromaius novaehollandiae* (Emu), but was absent from the W chromosome in all bird species studied (Indrajit Nanda et al. 1999; S. Shetty et al. 2002).

The *DMRT1* gene codes for a novel nuclear transcription factor, with a zinc-finger like DNA-binding motif, called the DM domain (Ellegren 2001). *DMRT1* is not related to the mammalian *SRY* gene, but

it is thought to work in a similar way, by binding and subsequently regulating expression of a downstream gene or genes involved in sexual differentiation (Ellegren 2001). Homologues of the *DMRT1* gene have been found in many vertebrates including mammals, fish and reptiles; that exhibit both GSD and TSD mechanisms (Raymond et al. 2000). It has been found to be expressed in the embryonic urogenital system of all vertebrates studied (Raymond et al. 2000). Furthermore *DMRT1* homologues, involved in sex determination, have also been discovered in a *Caenorhabditis elegans* (nematode) and *Drosophila melanogaster* (common fruit fly) (Raymond et al. 2000). This shows that the *DMRT1* gene and its association with sex determination has been conserved not only among vertebrates but also among animals belonging to other phyla (Raymond et al. 2000). No other gene associated with sex determination has ever been found to be conserved among such distantly related organisms.

Reverse transcriptase-PCR (RT-PCR) experiments have shown that *DMRT1* is expressed in the genital ridge of chicken embryos, prior to gonadal differentiation and is expressed at a higher level in males than in females (Raymond, Parker, et al. 1999; Craig A. Smith et al. 1999). This gives strong evidence to suggest the *DMRT1* gene is the primary sex determining gene in birds and acts through a dosage dependent sex determination mechanism (Raymond, Parker, et al. 1999; Craig A. Smith et al. 1999). However, *DMRT1* is expressed in the gonads as early as 3.5 days after the beginning of embryonic development, which is long before gonadal differentiation begins at day 6.5 (C.A. Smith et al. 2007). This evidence has led to the hypothesis that *DMRT1* acts indirectly on sex differentiation and a secondary factor may be involved (C.A. Smith et al. 2007). An alternative hypothesis proposes that the expression of *DMRT1* has a cumulative affect and must reach a threshold before it activates genes further downstream of the sex differentiation cascade (C.A. Smith et al. 2007). One study has shown that *DMRT1* expression in male chickens is at its peak at day 6, which fits well with the beginning of gonadal differentiation at day 6.5 (Craig A. Smith, Katz, and Sinclair 2003).

Histidine triad nucleotide binding protein, W-linked (*HINTW*), was discovered simultaneously by two independent groups working to identify novel genes on the chicken W chromosome (Hori et al. 2000; O'Neill et al. 2000). One group identified a gene which they believed encoded an altered form of PKC inhibitor/interacting protein (*PKCI*) on the W chromosome of the chicken and hence it was given the name *WPKCI* (Hori et al. 2000). The gene was isolated using a subtractive hybridization technique, comparatively screening a cDNA library, made from pooled undifferentiated chicken gonads (Hori et al. 2000). Southern blotting revealed 40 tandem copies of the gene and fluorescent *in situ* hybridization (FISH) located it at the non-heterochromatic end of the chicken W chromosome

(Hori et al. 2000). The other group identified the same gene, which they named Avian Sex specific W linked (*ASW*) gene (O'Neill et al. 2000). In this study they used the technique of representational difference analysis (RDA), using DNA from chicken embryos (O'Neill et al. 2000). More recent research has shown that the gene actually encodes a histidine triad nucleotide binding protein and hence is now called *HINTW* (Ceplitis and Ellegren 2004).

HINT proteins are a sub-group of proteins from the Histidine triad (HIT) family and are characterised as nucleotide hydrolase enzymes (C.A. Smith et al. 2007). However the *HINTW* gene has lost the DNA sequence that encodes the region of the protein responsible for its catalytic function (C.A. Smith et al. 2007). Whereas *HINTZ*, *HINTW*'s homologue, located on the Z chromosome of chickens still retains its catalytic region and therefore it is assumed also its catalytic function (C.A. Smith et al. 2007).

HINTZ is expressed in the gonads of both male and female chickens (Hori et al. 2000). *HINTW* is only expressed in female gonads, as it is located on the W chromosome (Hori et al. 2000). *HINTW* starts to be expressed at day 2-3 after the beginning of embryonic development in females, long before the initiation of gonadal differentiation at day 6.5 (Hori et al. 2000). However just like *DMRT1* the *HINTW* gene reaches its highest expression level between days 5 and 6 (Hori et al. 2000). This evidence supports either *HINTW* acting on a secondary factor or the HINTW protein having a cumulative effect, which must reach a threshold before differentiation begins (Hori et al. 2000).

The function of HINTW is not yet known. However, it is hypothesised that HINTW functions in a W chromosome dominant fashion to control sex determination, producing a female in its presence and a male in its absence. It is thought to do this by binding to HINTZ protein, which in this hypothesis is required for male testis development (Pace and Brenner 2003). HINTW and HINTZ have been shown to interact *in vitro* and this interaction reduces HINTZ catalytic activity (Pace and Brenner 2003). Furthermore, the *HINTW* gene has been shown to be under positive selection pressure and is also conserved among carinate birds (all birds not including ratites) (Pace and Brenner 2003). However, no homologue of *HINTW* has ever been found in a ratite. This indicates that either *HINTW* in not the primary sex determining gene of birds generally, or that sex determination in ratites is initiated by a different genetic mechanisms (C.A. Smith et al. 2007).

Reptile Sex Determination

Reptiles are the third sub-group of vertebrates, along with mammals and birds that make up the group amniotes. However, they are by far the least studied and therefore the least well understood. This becomes evident when investigating reptile sex determination and the mechanisms that they use to differentiate sex. There are many reasons for this imbalance in experimental data, which could be due to reptiles not being as charismatic as mammals or birds, but there have also been some fundamental technical difficulties when it comes to investigating the sex determination mechanisms of reptiles. First, unlike mammals or birds the captive breeding of reptiles is less widespread and more complicated to achieve successfully. This has meant that there is no foundation on which to conduct breeding experiments or allow the identification of sex chromosome abnormalities to study reptile sex determination mechanisms. Secondly, the detection of sex-linked markers has proven extremely difficult. However, reptiles are easily studied cytologically and as a result most information known about reptile sex determination mechanisms has been obtained by cytological research only.

Fossil evidence indicates that reptiles and birds shared a common evolutionary ancestor approximately 285 million years ago. Both groups share the cytological similarity of a relatively small number of macrochromosomes and many microchromosomes. However, unlike birds, reptiles exhibit a staggering variety of different sex determination mechanisms. This includes multiple independent examples of every GSD system ever observed, as well as the use of TSD, which makes reptiles the only terrestrial vertebrate identified to use an ESD system.

The very first discovery of an animal using temperature as a sex determination mechanism was a small, brightly coloured lizard; *Agama agama* (Rainbow Agama), discovered by Charnier in 1966. Since then, many other TSD reptiles have been identified. These include two species of Tuatara, which are the last living remnants from the order Sphenodontia (Peter Uetz, Jirí Hošek, and Jakob Hallerman 2012). The majority of turtle species studied, 69 from a total of 79 (Valenzuela and Lance 2004). There are data implicating 19 species of lizard, as well as all Crocodilian species studied (Valenzuela and Lance 2004).

Most publications on this topic since 1994 state that all crocodilian species use a TSD mechanism to differentiate sex, usually without citation. However, published data on TSD exists for only 12 of the 25 extant crocodilian species (Lang and Andrews 1994). This assumption occurred after a 1994 publication by Lang and Andrews, who produced a consolidated study of the 12 crocodilian species

that had been investigated for TSD (Lang and Andrews 1994). In this publication, the author hypothesised that all crocodilian species use a TSD system to differentiate sex (Lang and Andrews 1994). Although all data obtained so far support this hypothesis, TSD has not been investigated in 13 of the total 25 Crocodilian species and hence the claim remains hypothetical.

Before the 1960s all species of reptile that had been karyotyped were observed to lack cytologically distinguishable sex chromosomes and it was therefore concluded that sex chromosomes were absent from the entire Reptilian class (Mattey and Brick 1957). The first discovery of heteromorphic sex chromosomes in a reptile species was met with disbelief (Bull 1980). This was partially due to the sex chromosome systems that had thus far been discovered; those of mammals, birds and *Drosophila*, which had been shown to be ubiquitous across many related taxa (Bull 1980). Unlike other vertebrate groups, reptiles have a diverse array of GSD mechanisms. Although cytologically distinguishable sex chromosomes are quite rare in reptiles, sex chromosomes and therefore GSD mechanisms have been identified in species of turtle, lizard and snake (Ezaz, Rami Stiglec, et al. 2006).

Turtles

The majority of turtle species have been found to use TSD, suggesting that either TSD is ancestral to this order of reptiles or it has an inherent selective advantage (Ewert, Ewert, and Nelson 1991). One hundred and four turtle species have been karyotyped from a total of 327 extant species (Peter Uetz, Jirí Hošek, and Jakob Hallerman 2012; Olmo and Signorino 2010). Five were found to have discernible sex chromosomes (Olmo and Signorino 2010). Two of these species exhibit a male heterogametic XX/XY system, while the third has a female heterogametic ZZ/ZW system (Olmo and Signorino 2010). The fourth species appears to have a multiple-paired, male heterogametic system with XX females and XXY males (Sites, Bickham, and Haiduk 1979). While the fifth species, is the most recent to be confirmed as using a GSD system, but in this species sex is determined not by macrochromosomes, but by microchromosomes (Ezaz, Valenzuela, et al. 2006). Microchromosomes, as the name suggests, are very much smaller than their macro counterparts. This makes them extremely hard to identify using normal cytological techniques and even harder to distinguish from one another (Ezaz, Valenzuela, et al. 2006). Other turtle species may use microchromosomes to determine sex, but their micro sex chromosomes may simply be too small to be identified using karyotyping techniques, causing them to be overlooked (Ezaz, Valenzuela, et al. 2006).

Lizards

Lizards (sub-order Lacertilia) are the largest and most diverse sub-order of reptiles. Diversity is also evident in the sex determination mechanisms this group exhibits. Of the 5,634 extant lizard species, 953 have been karyotyped. Cytologically distinguishable sex chromosome have been identified in 172 of those species (Peter Uetz, Jirí Hošek, and Jakob Hallerman 2012; Olmo and Signorino 2010). The Lacertilia include examples of male (XX/XY) and female (ZZ/ZW) heterogametic sex chromosome systems, with examples of both single-pair and multiple-pair systems. There are also examples of lizards using micro sex chromosomes (Ezaz et al. 2005). Furthermore, some lizard species have been shown to use a combination of both GSD and TSD (Radder et al. 2007). In the case of these animals it appears the GSD system is dominant at intermediate incubation temperatures, producing an even ratio of male and female offspring (Radder et al. 2007). But when the incubation temperature reaches a certain threshold, their TSD system is initiated, shifting the sex ratio towards one particular sex (Radder et al. 2007).

While the diversity in lizard sex determination mechanisms is abundant there are also examples of conserved sex determination systems. For example species belonging to the family Agamidae appear to have a conserved female heterogametic system (Ezaz et al. 2009). The Agamid ZZ/ZW sex chromosome system does not share any homology with the ZZ/ZW sex chromosome systems of birds or snakes, indicating that the system evolved independently (Ezaz et al. 2009).

Snakes

Snakes are widely believed to have a conserved GSD system, sharing evolutionary parallels with the GSD systems of birds and mammals rather than those of other reptiles. Susumu Ohno was the first to identify cytologically distinguishable sex chromosomes in snakes (Matsubara et al. 2006; Ohno 1966). Ohno's investigation showed that some snake species have a female heterogametic sex chromosomes system with ZZ males and ZW females (Ohno 1966). He also observed that some snake and some bird karyotypes closely resembled each other (Becak et al. 1964). Supporting his investigation, snakes have been shown to have a smaller number of macrochromosomes, most commonly 8 pairs, together with a larger number of microchromosomes, most commonly ten (Becak et al. 1964). This arrangement is also observed in many bird karyotypes (Becak et al. 1964). The genomes of birds and snakes also contain an equivalent amount of DNA, of which the observable Z chromosomes in some species of snake and bird, make up about 10% of the genome (Becak et al. 1964). The cytological similarities observed by Ohno led him to hypothesise that snakes and birds share a common, female heterogametic sex determination system, which evolved before

their ancestors diverged approximately 180 million years ago (Becak et al. 1964). However, in 2006, Ohno's hypothesis was overturned when molecular genetics research revealed that, despite karyotypic similarities, the observable sex chromosome of snakes and birds are unrelated, implying these two systems evolved independently (Matsubara et al. 2006).

Snake Cytology

At last count (February 2013) there were 3,432 extant species of snake (Peter Uetz, Jirí Hošek, and Jakob Hallerman 2012). Three-hundred species have been karyotyped and 137 of these have been observed to have cytologically distinguishable sex chromosomes (Olmo and Signorino 2010; Peter Uetz, Jirí Hošek, and Jakob Hallerman 2012). All 137 species have been reported to have a female heterogametic sex chromosome system, suggesting that a ZZ/ZW sex chromosomes system might be conserved among these snake species (Olmo and Signorino 2010). However, in the remaining 163 species of snake that have been karyotyped, no sex chromosomes have been identified (Olmo and Signorino 2010). Three thousand one hundred and thirty-two species of snake are yet to be karyotyped, leaving 3,295 snake species, approximately 96% of extant snakes with undetermined sex chromosomes. Despite this fact, it is a widely accepted theory that all snakes share a common, female heterogametic sex chromosome system.

This hypothesis has been reinforced by the concept of the stepwise differentiation of snake sex chromosomes. Snake species that belong to the evolutionarily youngest families, Elapidae and Viperidae, have the most differentiated sex chromosomes (Matsubara et al. 2006; Ohno 1966). In these families, the W chromosome differs from the Z both in size and overall banding pattern (Matsubara et al. 2006). Species belonging to the family Colubridae, which is evolutionarily older, have less differentiated sex chromosomes with a W chromosome that is most commonly equal in size to the Z, but differs in centromeric position (Matsubara et al. 2006). In evolutionarily older snake families such as Pythonidae and Boidae, the difference between the hypothesised Z and W chromosome cannot be distinguished using cytological techniques (Matsubara et al. 2006; Ohno 1966). The cause of this increase in differentiation among more recently evolved groups is not known, but the phenomenon has also been observed in birds (Matsubara et al. 2006; Ohno 1966).

When comparing the cytology of snakes, it becomes clear that they can be divided into two distinct groups, as illustrated in Figure 1-2. Firstly, the vast majority of snake species that have been karyotyped belong to three snake families, Viperidae, Colubridae and Elapidae, which will hereby be referred to collectively as the Heterogametic Female Identified (HFI) group (Olmo and Signorino

2010). Cytologically distinguishable sex chromosomes have been identified in just over half the species karyotyped in this group, with all examples presenting as heterogametic female (Olmo and Signorino 2010). In these examples, the Z and W chromosomes are the fourth or fifth largest chromosome pair (Matsubara et al. 2006). The Z chromosome is generally well conserved both in size and banding pattern (Mengden and Stock 1980; Matsubara et al. 2006). While the W chromosome varies greatly in size and banding pattern, both between the families and, in some cases, between closely related species (Matsubara et al. 2006; Mengden and Stock 1980; Mengden 1981).

The second group includes all remaining snake families, (18 in total) and henceforth will be referred to collectively as the Heterogametic Sex Unidentified (HSU) group. Cytological evidence from this group is much sparser, with nine families lacking any karyotype examples at all. However, the two largest families, Pythonidae and Boidae are represented relatively well (Olmo and Signorino 2010). Cytologically distinguishable sex chromosomes have never been identified in any species belonging to the HSU group. With the exception of one species, *Acrantophis dumerili* (Dumeril's boa), which belongs to the family Boidae (Olmo and Signorino 2010; Mengden and Stock 1980). When comparing the cytology of the HFI and HSU groups, non-differentially stained karyotypes show close similarity. However, comparative staining studies have shown major differences between species belonging to the families Colubridae and Boidae (Mengden and Stock 1980).

In 1980, Mengden and Stock published the karyotype of *A. dumerili* which, unlike other Boa species, was observed to have cytologically distinguishable sex chromosomes. Mengden and Stock observed a Z chromosome estimated to be the fourth largest chromosome and a W chromosome of a similar size but with a different centromeric position (Mengden and Stock 1980). However, Mengden and Stock reported: “most of our samples were from unsexed live specimens” (Mengden and Stock 1980). Therefore, significant doubt exists as to whether these cytologically distinguishable sex chromosomes constitute a ZZ/ZW sex chromosome pair. It is equally possible that these chromosomes could constitute a XX/XY sex chromosome system in this species. This is the first and only published karyotype from this species and, due to the limited information, neither hypothesis can be regarded as sound.

In the same report, Mengden and Stock also published a karyotype of *Elaphe subocularis* (Trans-Pecos rat snake), which belongs to the HFI group (Mengden and Stock 1980). The authors identified that this species has a distinguishable female heterogametic sex chromosome system (Mengden and

Stock 1980). However, the sex chromosomes appear remarkably different from those of other Colubrids, including closely related species from the same subfamily. The Z and the W chromosome of *E. subocularis* differs both in size and banding pattern from those of the closely related species *Elaphe obsoleta* (black rat snake), as do most of the autosomes (Mengden and Stock 1980). While the Z chromosome is approximately the fourth largest chromosome, the W chromosome appears to be almost twice the size of the Z (Mengden and Stock 1980). This makes the W chromosome the third largest chromosome and the largest W chromosome ever observed in any snake species. These data differ markedly from two *E. subocularis* karyotypes published in 1971 by Baker, Bull and Mengden and in 1995 by Camper and Hanks, which in both publications reported that no cytologically distinguishable sex chromosomes could be identified (Mengden and Stock 1980; Camper and Hanks 1995; Baker, Bull, and Mengden 1971). This conflicting information may reflect differences in the interpretation of the karyotype or could indicate an individual snake with a distinctly different sex chromosome system, or perhaps the phenotypic similarities may have proven deceptive and the snake may have been a different species entirely. As interesting as this observation is, the sex determination system of this species has never been studied in detail and again little conclusion can be made from the limited cytological data provided by Mengden and Stock and others (Mengden and Stock 1980; Camper and Hanks 1995; Baker, Bull, and Mengden 1971).

In most observable sex chromosome systems, the differentiation of the sex determining chromosome (that is the Y or W chromosome) results in a decrease in the active gene content, a reduction in total size, increase in repetitive elements and the accumulation of heterochromatin. Heterochromatin is tightly packaged DNA that usually is made up of repeating DNA elements. A total of 20 snakes species have been investigated for their heterochromatin content, using G-banding karyotypes (Olmo and Signorino 2010). All snake species belonging to the HFI group show a highly heterochromatic W chromosome, with the exception of *Ptyas mucosus* (oriental rat snake) which belongs to the family Colubridae (Olmo and Signorino 2010). In comparison, two species belonging to the HSU group have been investigated for heterochromatin, *Eryx johnii* (Indian sand boa) and *Gongylophis conicus* (rough-scaled sand boa), both are members of the sub-family Erycinae and both showed no detectable heterochromatin on the putative W chromosome (Olmo and Signorino 2010).

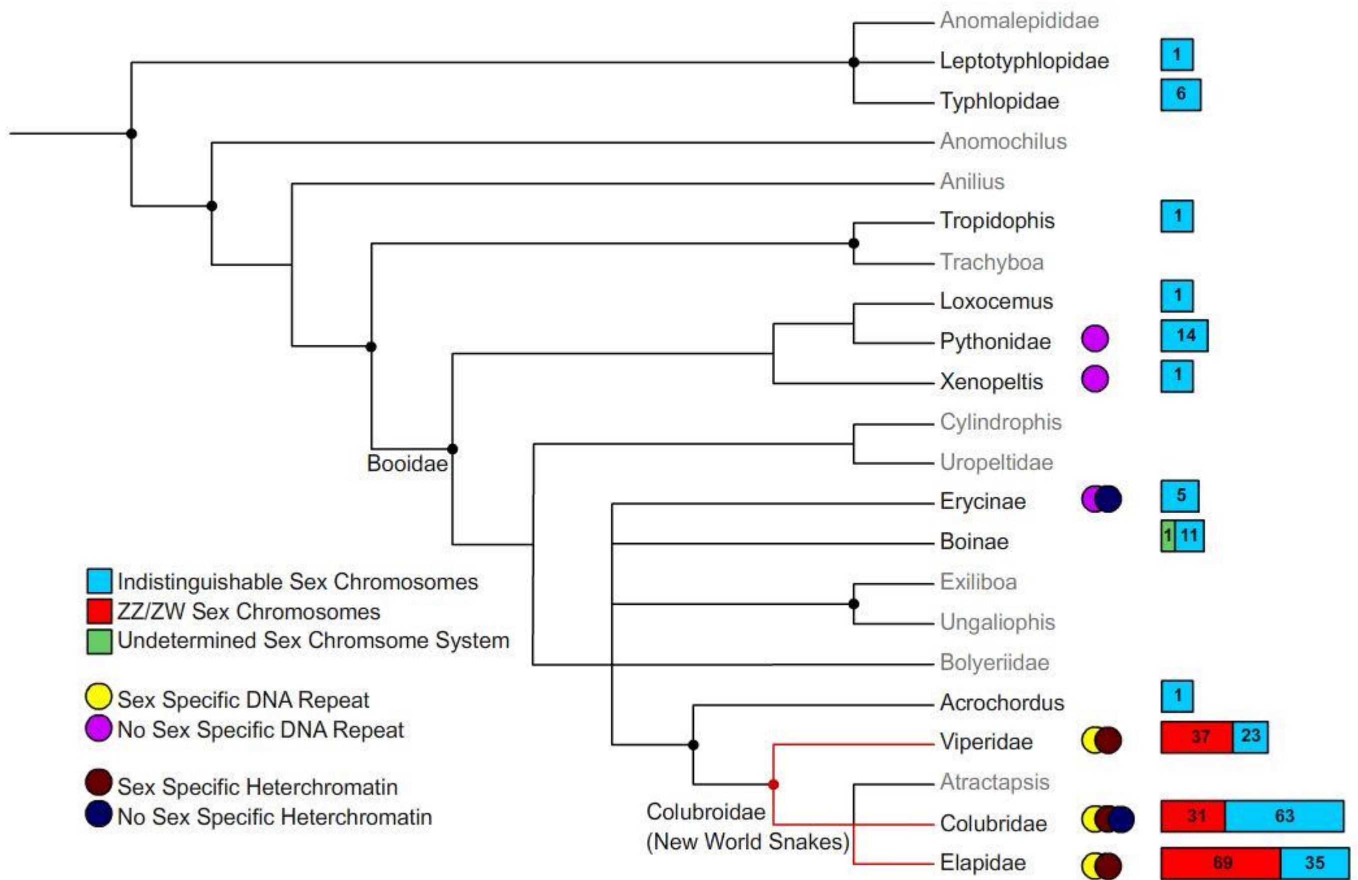

Figure 1-2 Evidence for Mechanisms of Snake Sex Determination

Figure 1-2 is a phylogeny of all major snake families showing the sex determination systems observed and investigated. ZZ/ZW represents a female heterogametic sex chromosomes system, BKM; banded krait minor satellite repetitive element found to be specific to the W chromosome and CTNNB1 repeat represents the repetitive elements discovered within the terminal intron of the gene CTNNB1 also found to be specific to the W chromosome. Note: this phylogenetic tree only indicates relationships, not evolutionary branching times.

Snake Genetics

In general, the genetics of snakes has been investigated very little and snake sex determination is no exception. The very first DNA sequence associated with sex in snakes was the banded krait minor-satellite (Bkm) repeat described in 1975 (Singh and Ray-Chaudhuri 1975). This sequence is composed of GATA and GACA tandem repeats and was discovered to be specific to the W chromosome in *Bungarus caeruleus* (common krait), belonging to the snake family Colubridae (Singh and Ray-Chaudhuri 1975).

The Bkm repeat has been investigated in nine different snake species (Olmo and Signorino 2010). Six snake species belonging to the HFI group have been shown to contain a Bkm repeat specific to their observable W chromosomes, with the frequency of the Bkm repeat increasing as the degree of observable differentiation increases between the Z and the W chromosomes (Olmo and Signorino 2010). However, the remaining three species, belonging to the HSU group; *E. johnii*, *Liasis fuscus* (brown water python) and *Xenopeltis unicolor* (sunbeam snake), have been shown to completely lack any chromosome specific Bkm repeat within their genome (Olmo and Signorino 2010).

The most comprehensive genetic study of snake sex chromosomes was published in 2006 by Matsubara and colleagues (Matsubara et al. 2006). In their publication they re-investigated S. Ohno's original hypothesis that snakes and birds share a common sex chromosomes system, by creating a high resolution cytological map of the snake species *Elaphe quadrivirgata* (Japanese four-lined rat snake) (Matsubara et al. 2006). They identified the chromosomal location of 109 genes in *E. quadrivirgata* and then compared these gene locations with those of chicken and human as seen in Figure 1-3 (Matsubara et al. 2006). Eleven genes were located to the observable Z chromosome of *E. quadrivirgata*, whilst in the chicken genome the corresponding genes were located to autosomes 2 and 27 and, in the human genome, to autosomes 3, 7, 10 and 17 (Matsubara et al. 2006). This group also found that cDNA fragments of 4 genes; *CHD1*, *DMRT1*, *ACO1* and *SOX9* that are associated with bird and mammal sex determination, all mapped to autosomes in snakes (Matsubara et al. 2006). These results revealed that despite the cytological similarities between snake and bird sex chromosomes, they are genetically very different (Matsubara et al. 2006). Indicating that the two systems have evolved independently of each other (Matsubara et al. 2006).

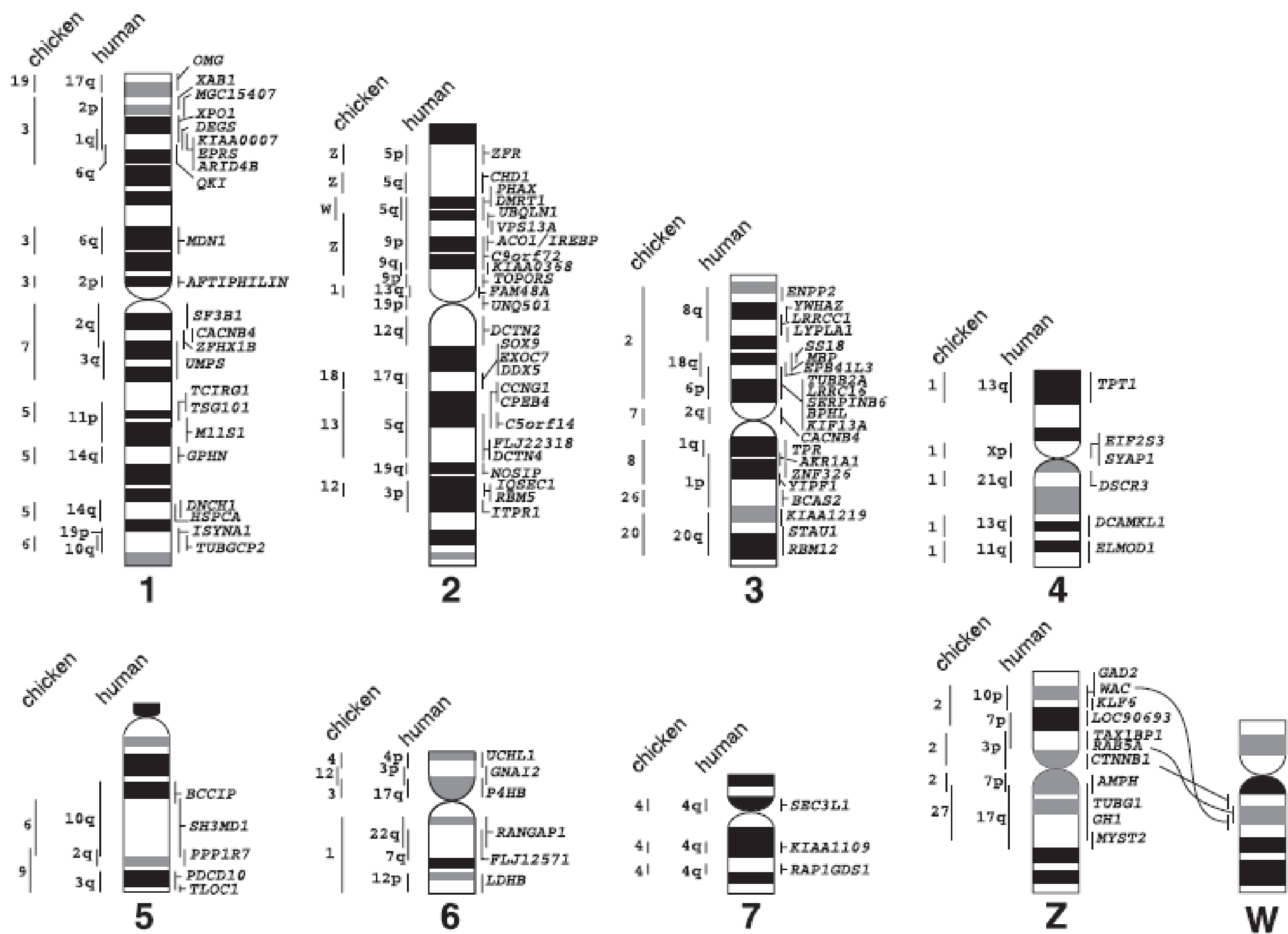

Figure 1-3 A comparative cytogenic map of chromosomes 1-7, Z and W chromosomes of E. quadrivirgata (Matsubara et al. 2006)

Figure 1-3 is a comparative cytogenetic map of chromosomes 1–7 and the Z and W chromosomes of E. quadrivirgata taken from Matsubara K. et al., 2006 publication. The chromosomal location of each gene is shown on the right of each chromosome (Matsubara et al. 2006)*. The ideogram of G-banded chromosomes was constructed in a previous study (Matsuda et al. 2005). The human and chicken chromosome segments with homology to the snake chromosomes and their chromosome numbers are indicated to the left of the snake chromosomes* (Matsubara et al. 2006)*. Gene symbols are according to the human nomenclature* (Matsubara et al. 2006)*.*

The same study also investigated the different degrees of sex chromosome differentiation seen among three snake species; *Python molurus* (Indian python), *E. quadrivirgata* and *Trimeresurus flavovirdis* (habu snake). When comparing the observable Z chromosomes of *E. quadrivirgata* and *T. flavovirdis* (HFI group), subtle differences were seen in gene order, but all 11 genes mapped to the same chromosome of these two snake species (Matsubara et al. 2006). However, when comparing the observable W chromosomes of these snakes only three genes could be mapped to the W of *E.*

quadrivirgata and not a single gene could be mapped to the W chromosome of *T. flavovirdis* (Matsubara et al. 2006). This evidence shows that while the Z chromosome of these two species is conserved the W chromosomes have differentiated to different degrees, reinforcing the cytological evidence for the step-wise differentiation of snake sex chromosomes. Alternatively, when the potential Z and W chromosomes of *P. molurus* (HSU group) were compared, all 11 genes could be mapped to both the proposed Z and W chromosome, with no difference in location or gene order seen between them (Matsubara et al. 2006).

In 2012, Matsubara and colleagues published new data that focused on identifying the GC heterogeneity in Sauropsida, which they used to infer details of their evolution (Matsubara et al. 2012). In this study, they added to the cDNA map of *E. quadrivirgata*, including 183 gene locations, of which 144 where on macrochromosomes and the remaining 39 on microchromosomes (Matsubara et al. 2012). This brought the total number of genes mapped to the Z chromosomes of *E. quadrivirgata* to 20 (Matsubara et al. 2012). The revised gene map of the Z chromosome again resembles the patterns found on autosome 2 and 27 in the chicken (Matsubara et al. 2012).

In 2010 a study investigating the repetitive elements of snake sex chromosomes, as well as performing a Comparative Genome Hybridization (CGH) for three snake species *Notechis scutatus* (Tiger snake), *Stegonotus cucullatus* (slaty-grey snake), and *L. fuscus*. The authors investigated Bkm repeats, 18S ribosomal related repeats and chicken Z chromosome sequences, as well as two new repetitive elements they discovered, shown to be associated with snake sex chromosomes (O'Meally et al. 2010). The two DNA repeats, which the authors refer to as DMRT1-related and CTNNB1-related repeat, have sequence homology to the gene *DMRT1*, which is located on the Z chromosome of chicken and is hypothesised to be the primary sex determining gene in birds, and to gene *CTNNB1*, which in 2006 was mapped to both the Z and W chromosome of *E. quadrivirgata* and *P. molurus* (Matsubara et al. 2006; O'Meally et al. 2010).

In this study they showed the W chromosome of the Colubridae species *S. cucullatus* to have female-specific sequences, composed of Bkm repeats and sequences similar to those found on chicken chromosome W (O'Meally et al. 2010). These repeats were dotted along the entire length of the W chromosome and became more concentrated at the distal region of the long arm (O'Meally et al. 2010). In comparison, in the Elapid species *N. scutatus*, the majority of the W chromosomes contained concentrated female-specific repetitive sequences originating from Bkm, DMRT1-related, CTNNB1-related, 18S-related , and chicken W chromosome similar sequences, (O'Meally et al. 2010).

Investigation into the snake species *L. fuscus*, which belongs to the HSU group, revealed no sex specific sequences within the entire genome (O'Meally et al. 2010).

Comparative Genomic Hybridization (CGH) also showed female-specific sequences along almost the entire length of the W chromosome in both HFI species *S. cucullatus* and *N. scutatus* (O'Meally et al. 2010). While CGH showed no female-specific regions, on the putative W chromosome or any other chromosome of the HSU species *L. fuscus* (O'Meally et al. 2010). As stated by the authors, the CGH technique used is sensitive enough to identify a 2-3Mb DNA region of sex specific DNA, indicating that no sex specific regions larger than 3Mb exist in *L. fuscus* (O'Meally et al. 2010).

Investigation using second generation sequencing to investigate the transcriptome of *Thamnophis elegans* (Western terrestrial garter snake), *Sistrurus miliarius* (pygmy rattlesnake) and *B. constrictor* (common boa), compared them to the genome of the lizard *Anolis carolinensis* (Green anole) (Hoff, M. 2013). The study investigated the total frequency of RNA molecules expressed, comparing the RNA expression of the sex chromosome of male and female snakes. It showed that RNA sequences representing the sex chromosomes of the female *T. elegans* and *S. miliarius*, belonging to Colubridae and Viperidae respectively, two HFI group species, are present in half quantities when compared to males and to the *A. carolinensis* homologue, chromosome 6, representing the divergent W chromosome (Hoff, M. 2013). However the putative sex chromosomes of *B. constrictor*, which belongs to the Boidae snake family and is a HSU snake species, showed no difference in expression between sexes, similarly to all other autosomes (Hoff, M. 2013).

In 2013 Vicoso et al. compare the genomic DNA sequences of the same three snake species; *T. elegans*, *S. miliarius* and *B. constrictor*. The authors DNA sequenced and assembled the genomes of one male and one female *T. elegans* and *S. miliarius* and used an assembled female genome of *B. constrictor* from another source to make comparisons. They also performed RNA-seq on one male and one female *S. miliarius* and *B. constrictor* to compare the patterns on gene expression of these two snake species and the two sexes. The study found that the proposed Z and W sex chromosomes of *B. constrictor* were homomorphic with no detectable sex specific DNA regions (Vicoso et al. 2013). The authors suggest that this result may be attributed to the sex identifying allele on the W chromosome of Boas being extremely small and was overlooked due to a lack of DNA sequence coverage, or alternatively that the sex identifying region present on W chromosome has come from a fusions between the W chromosomes and another chromosomes and the genome sequence assembly arranged the sequence data incorrectly on the chromosome from where the fused DNA

originated (Vicoso et al. 2013). The study also showed that the Z and W chromosomes of *T. elegans* and *S. miliarius* were heteromorphic across their entire length with no presents of a pseudoautosomal region like that found in mammals (Vicoso et al. 2013). This indicates that even in *T. elegans*, which only shows slight cytological differentiation between the Z and W chromosomes, these chromosomes have diverged greatly and no longer undergo recombination, which is not what was expected (Vicoso et al. 2013).

In a more recent study, the authors compared the DNA sequences produced from targeting the *CTNNB1* gene in males and females of 12 different snake species (Laopichienpong et al. 2017). The *CTNNB1* gene is one of the 20 genes, which has been previously mapped to the Z chromosomes of *E. quadrivirgata* (Matsubara et al. 2006). Eight of the 12 snake species; *Daboia siamensis* (Eastern Russell's viper), *Naja kaouthia* (monocellate cobra), *Naja siamensis* (Indochinese spitting cobra), *Ophiophagus hannah* (King cobra), *Bungarus candidus* (Malayan krait), *Oligodon fasciolatus* (small-banded kukri), *Boiga dendrophila* (mangrove snake) and *Coelognathus flavolineatus* (black copper rat snake) were able to be identified by sex using PCR, amplifying between exon 10 and 12 of the *CTNNB1* gene (Laopichienpong et al. 2017). Evidence to show the divergence of the *CTNNB1Z* and *CTNNB1W* genes in the 8 species, all belonging to the HFI group. The remaining 4 snake species; *Cylindrophis ruffus* (red-tailed pipe snake), *X. unicolor* and *Python bivittatus* (Burmese python), belong to the HSU group and *Xenochrophis flavipunctatus* (yellow-spotted keelback), which belong to the family Colubridae and the HFI group, could not be identified by sex (Laopichienpong et al. 2017). The authors suggest that *CTNNB1* on the Z and W chromosomes of the three snake species belonging to the HFI group, have not differentiated and as such produce a comparable DNA fragment profile (Laopichienpong et al. 2017). While, *X. flavipunctatus* is an example of the opposite extreme and has lost the *CTNNB1* gene from its highly differentiated W chromosome (Laopichienpong et al. 2017). Phylogenetic analysis of the DNA sequences produced showed the *CTNNB1Z* and *CTNNB1W* DNA sequences from the three HSU species were conserved (Laopichienpong et al. 2017). While the *CTNNB1W* sequences from the species belonging to the HFI group, were most closely related to the *CTNNB1Z* sequences from the HSU group (Laopichienpong et al. 2017). The authors suggest this is because the *CTNNB1Z* and *CTNNB1W* genes in the HFI group have diverged independently in each species, sometime after the HSU and HFI groups diverged from one another (Laopichienpong et al. 2017).

In the most recent study investigation of sex specific DNA markers in three snake species; *Boa imperator* (northern boa), *P. bivittatus* and *Crotalus atrox* (western diamondback rattlesnake), using

restriction-site-associated DNA sequencing (RAD-seq) was performed (Gamble et al. 2017). The RAD-seq method, involves identifying and investigating polymorphisms between two complex genomes by comparing the genetic markers produced after the two genomes were sampled (Miller et al. 2007; Baird et al. 2008). The technique was originally coupled with microarray analysis, but high throughput sequencing was quickly employed, allowing for each marker's sequence to be investigated separately (Miller et al. 2007; Baird et al. 2008). The results of the 2017 study revealed that the two snake species, *B. imperator* and *P. bivittatus* belonging to the HSU group, most likely use a male heterogametic sex determination system. While *C. atrox*, which belongs to the family Viperidae and the HFI group, most likely uses a female heterogametic sex determination system (Gamble et al. 2017). The authors of the study supported this finding by using PCR to prove some markers identified were male or female-linked for their respective snake species (Gamble et al. 2017). This is the first genetic evidence to suggest some snake species may in fact use a male heterogametic sex chromosomes system with XY males and XX females, which is in opposition to the theory that all snake species use a female heterogametic sex determination system.

Conclusion

Cytological and genetic information about snakes and their sex determination systems is limited at best. The information we do have indicates that about half of the snake species belonging to the families Viperidae, Elapidae and Colubridae have a conserved female heterogametic sex chromosome system, with ZZ males and ZW females, comprising the Heterogametic Sex Identified (HFI) group. This group shows high levels of conservation between their Z chromosomes, both cytologically and genetically, while the W chromosomes of these snake families, genera within these families and even some closely related species have diverged to different degrees. It has been shown that this snake sex determination system has evolved independently from other amniotes but has followed the same evolutionary patterns as the sex determination system of birds. This has resulted in the two sex chromosome systems presenting parallel karyotypic features.

In comparison within the remaining eighteen families, making up the Heterogametic Sex Unidentified group, no snake species has ever been shown to have cytologically distinguishable sex chromosomes with the exception of one single snakes, which had not had its sex identified. No genetic evidence exists that supports the hypothesis that the conserved sex determination system within the HFI group has ever been observed in any species belonging to the HSU group. Moreover, cytogenetic mapping of *P. molurus* has shown no difference between the genes located on the proposed Z and W chromosomes (Matsubara et al. 2006). Furthermore, transcriptome analysis of *B.*

constrictor has shown no detectable difference in the level of RNA expression from suspected Z and W chromosomes (Hoff, M. 2013). In addition, CGH analysis has shown no sex specific region larger than 3MB can be identified within the species *L. fuscus* (O'Meally et al. 2010). Most recently it has been discovered that *B. imperator* and *P. bivittatus* both belonging to the HSU group, most likely use a male heterogametic sex determination system (Gamble et al. 2017). The genetic evidence is now challenging the well-accepted theory that all snake species, share a common, female heterogametic sex chromosome system.

Thesis Hypothesis Aims and Objectives.

The aim of this study was to test the hypothesis: *Morelia spilota* (carpet python) has a female heterochromatic sex chromosomes system with ZZ males and ZW females. And that this genetic sex determination mechanism is initiated by a female-specific DNA region or regions that have evolved independently of other amniote sex determination systems.

To test this hypothesis, four different technical approaches were evaluated, each comprising a separate chapter within this thesis:

Chapter 2 - Randomly Amplified Polymorphic DNA Analysis

This chapter investigates sex-linked DNA regions within both male and female snake genomes, through PCR amplification of random polymorphic DNA regions using a short arbitrary oligonucleotide primer and comparing the DNA fragment profiles produced. It is expected that any sex-linked DNA regions will present in all male or female animals tested but be absent from the opposite sex.

Chapter 3 - Representational Difference Analysis

This chapter describes identification of sex-linked DNA regions within both male and female snake genomes, using a PCR-mediated subtractive hybridisation technique to selectively target DNA regions unique to both male and female snake genomes.

Chapter 4 - Comparative Genomics

This chapter aims to investigate the DNA regions of genes previously mapped to the Z and W chromosome of *E. quadrivirgata*, in our target species (Matsubara et al. 2012, 2006). Using a comparative genomic approach; designing primers specific for genomic regions within these gene, followed by PCR amplification and DNA sequencing the genomic DNA of these genes for a snake

species. It is planned to investigate the DNA sequences produced from both male and female snakes and identify any sex associated DNA regions.

Chapter 5 - Cosmid Library Construction and Second-Generation Sequencing
This chapter describes the construction of a large insert cosmid library, representing the genomes of one male and one female snake. This resource will then be used to investigate any sex-linked DNA regions in greater depth and to investigate genes already mapped to the snake Z and W chromosome in greater detail.

Model Species
The initial ideas that developed into this thesis project arose from request of local reptile keepers and breeders for a genetics-based sex test for their captive bred pythons. These requests were asked of Dr David Groth, who is the primary supervisor of this thesis project, who has provided a commercial sex test for birds for over 20 years. These questions from the local reptile keeping community started conversations and the first relationships that lead to the aims of this thesis project; to investigate sex-linked DNA regions within the most popular python species kept as a pet in Western Australia *Morelia spilota* (Figure 1-4).

All snakes sampled in this thesis belonged to the sub-species *Morelia spilota imbricata* (Southwest Carpet Python), which is endemic to the southwest of Western Australia (Macdonald 2017). This subspecies is listed by IUCN criteria Near Threatened, due to the destruction of natural habitat due to urbanisation. Within the species *Morelia spilota* six sub-species; *M. spilota imbricata*, *Morelia spilota cheynei* (jungle carpet python), *Morelia spilota mcdowelli* (Coastal carpet python), *Morelia spilota metcalfei* (inland carpet python), *Morelia spilota spilota* (diamond carpet python) and *Morelia spilota variegata* (carpet python) exist (Macdonald 2017; IUCN 2010). Two of these subspecies; *M. spilota spilota* and *M. spilota metcalfei* are listed as endangered in Victoria (IUCN 2010).

M. spilota is one of the most common snake species kept and bred as pets in Western Australia. This is because the snake species is a relatively easy snake to keep as a pet, with a wiliness to eat frozen food, a non-aggressive behaviour towards being handled by humans, a relatively impressive but manageable size and they exists in a variety of colours and patterns. This popularity in tern made the snake species an excellent model for this research project because samples, both sloughs and blood, were able to be obtained from live specimens, which had their sex confidently determined by both

documented mating or egg laying behaviour, which was confirmed by probing. Using DNA samples of snakes that had their sex confidently determined was crucial to the use of the molecular genetic techniques detailed in this thesis.

Snake species belonging to the Morelia genus are commonly referenced to as "tree pythons" with all 7 species belonging to the genus having an arboreal habitat. The Morelia genus is the equal largest genus belonging to the family Pythonidae, which is made up of 6 genera in total. The Pythonidae family is one of the major clades making up the snake superfamily Henophidia, which we have defined in this thesis as the Heterogametic Sex Unidentified (HSU) group of snakes. It is hoped that finding from this thesis research can be inferred to other snakes belonging to the Morelia genus and potentially the Pythonidae snake family. Giving insight into the sex determination of a snake family that belongs to the HSU snake group and has despite being investigated has not been shown to have a female heterogametic sex chromosomes system as is present in other snake species.

The objective of the four experimental thesis chapters is to use the DNA obtained from a non-model snake species *M. spilota* and, using three different molecular genetics approaches, investigate, and identifying DNA regions which are sex-linked. Once obtained, these sex-linked regions would be investigated in greater detail, using cosmid libraries combined with second generation sequencing. It was proposed to use this information to define these DNA regions, their length, chromosomal location, and their possible function in determining sex in snakes.

Figure 1-4 Morelia spilota imbricata (Southwest Carpet Python)

Chapter 2 - Random Amplified Polymorphic DNA Analysis

Introduction

It is currently hypothesised that all snake species share a common genetic sex determination system where the female is heterogametic (Barker and Barker 2006; Ezaz, Rami Stiglec, et al. 2006; Becak et al. 1964). This theory is based upon cytogenetic research that has been used to infer broad chromosomal similarities between snake and birds species, originally thought to be homologous with those of snakes (Barker and Barker 2006; Ezaz, Rami Stiglec, et al. 2006; Becak et al. 1964). To date, approximately 300 snake species have been karyotyped, however only 137 of these karyotypes appear to have cytologically distinguishable sex chromosomes (Olmo and Signorino 2010). One hundred and thirty six of these belong to three relatively closely related snake families, Elapidae, Viperidae and Colubridae, leaving approximately 96% of all snake species yet to have their sex chromosomes identified (Olmo and Signorino 2010). Other than the broad karyotype similarities observed between different snake species, the cytological evidence does not support the existence of a common sex chromosome system in any snake species not belonging to one of these three families.

In more recent times, studies into snake sex determination have employed several molecular genetic methodologies, which have largely replaced traditional karyotyping approaches. These techniques have been successfully used to identify the chromosomal location of a variety of different genes, including 11 genes which have been mapped to the Z chromosome in three species of snake (Matsubara et al. 2006, 2012). Molecular genetic techniques have also identified repetitive elements within the snake genome, showing sex-linkage in some snake species, but do not appear to be conserved among snake species generally and no function has been attributed to them (O'Meally et al. 2010; Olmo and Signorino 2010).

At the commencement of this research there was no published studies investigating the snake sex chromosomes as a unique system that has evolved independently of other vertebrate species. All previous research focussed on identifying similarities, either cytological or genetic, between the sex chromosome systems of snakes with that seen in birds and/or mammals. However, because of the limited knowledge about snake genetics, the identification of potentially novel or conserved sex determination systems requires techniques that do not depend upon previous generated DNA sequence data. In this chapter, this problem is addressed through the application of a molecular genetic technique called Random Amplified Polymorphic DNA (RAPD) analysis. This work described

in this chapter was performed prior to the introduction of affordable large-scale massively parallel sequencing and should be considered in that light.

Random amplified polymorphic DNA analysis is a PCR-based method, which uses a single primer, most commonly 10 base pairs in length, to amplify unknown DNA sequences within a genome. The technique was described in 1990, simultaneously by two independent groups, Williams *et al*. and Welsh and McClelland (Williams et al. 1990; Welsh and McClelland 1990). Since 1990, the use of RAPDs has become increasingly popular method because it is cheap, simple, and more importantly does not require any *a priori* sequence knowledge of the genome being investigated. It has been used for a considerable variety of different applications, including investigating genetic diversity, pedigrees, construction of genetic maps, identification of cultivars, pest resistance genes, and the identification sex-linked markers (Handrys, Balick, and Schierwater 1992; Bahy A. Ali et al. 2004; Franck A. Atienzar and Jha 2006; Leon Huynen, Craig D. Millar, and David M. Lambert 2002).

The RAPD technique is an adaptation of Polymerase Chain Reaction (PCR). However, unlike standard PCR, the short, single primer acts for both a forward and reverse priming of the DNA synthesis. This means that a DNA fragment will only be amplified if the following criteria are met: the primer must bind to complementary sequences in the forward and reverse orientation and must occur close enough to each other to be amplified under normal PCR conditions. The RAPD primer's short length and arbitrary, non-specific nature means that there is a high probability that these conditions will be adequately fulfilled multiple times within each genome, resulting in a RAPD fragment profile, which can subsequently be visualised using agarose or acrylamide gel electrophoresis.

Natural genetic mutation dictates that a RAPD fragment profile produced using two different genomic DNA templates is unlikely to be identical. Theoretically, a single nucleotide mutation at one RAPD priming site is all that is required to alter the RAPD fragment profile produced. RAPD fragments that are associated with a specific trait and/or a chromosomal region of interest, can be identified by comparing the RAPD fragment profiles produced from two different genomic templates. For example, comparing the RAPD fragment profiles produced from two different sexes could be used to identify DNA sequences that are potentially sex-linked. This technique has been previously used to successfully identify sex specific DNA markers in *Piper longum* (Long Pepper), *Salix viminalis* (Basket Willow), *Gracilaria gracilis* (Red Seaweed), *Pistacia vera* (Pistachio), *Simmondsia chinensis* (Jojoba), *Carica papaya* (Papaya), *Struthio camelus* (Ostrich), *Oncorhynchus mykiss* (Rainbow Trout), *Actinidia chinensis* (Kiwi fruit) and *Humulus lupulus* (Hop) just to name a few

(Banerjee, Manoj, and Das 1999; Semerikov et al. 2003; Martinez E. A. et al. 2003; Hormaza, Dollo, and Polito 1994; Agrawal et al. 2007; Urasaki et al. 2002; Bello and Sanchez 1999; Iturra et al. 1997; Gill, Harvey, and Fraser 1998; Polley, Ganal, and Seigner 1997).

In this present study, RAPD fragment profiles of male and female *Morelia spilota* (Figure 2-1) were generated using 102 different RAPD primers and the fragment profiles used to identify potential sex-linked DNA sequences. The snake species *M. spilota* was chosen for the experimental species because it is the most popular pet snake species in Western Australia and multiple individuals were readily available for DNA sampling. This species belongs to the Pythonidae family, and has been shown to have cytologically indistinguishable sex chromosomes (Olmo and Signorino 2010). This study uses a non-invasive DNA extraction technique, using sloughed skin as the primary DNA source, providing a less invasive DNA sampling method than techniques such as mouth swabs, blood sampling or scale clipping. An example of a *M. spilota* with after sloughing is shown in Figure 2-1.

Figure 2-1 Morelia spilota imbricata after sloughing

Materials and Methods

Snake sloughs were donated from Armadale Reptile and Wildlife Centre (304 - 308 South Western Hwy, Wungong WA 6112), West Australian Reptile Park (92 Henley Street, Henley Brook WA 6055) and Rockingham Reptile Traders (Unit 7, 117 Dixon Rd, Rockingham WA 6168). A total of 28 sloughs from 8 different species of snake were donated. For use within this project only sloughs from *Morelia spilota imbricata* were used. A total of 17 *M. spilota* sloughs were obtained, with five from known male and nine from known female specimens. The sex of these snakes was determined by probing and through observation of the sexual behaviour of the donor i.e. either visualisation of a hemipenis during mating (males) or egg production (females). The remaining 3 sloughs were from snakes of unknown sex.

To verify that the sloughs collected were from the identified individuals, only freshly shed sloughs were used. Any foreign material, such as faecal matter, within or on the slough was removed with a margin of 5 cm of the shed skin around the debris removed, with a pair of scissors. Some sloughs that were very fresh were still moist and these were dried at room temperature for 24 hours. After preparation, all sloughs were packaged into Zip-lock bags and stored in the dark away from insects. The use of snake sloughs as a source of DNA instead of blood sampling abrogated the need for animal ethics approval. Because a snake sloughs its skin naturally and once it has done so, it is considered no longer a part of the living animal, no animal ethics approval was required for slough collection or use in this project.

Snake Slough DNA Extraction

DNA extraction was performed on each of the *M. spilota* sloughs, using a modified version of the method detailed by James W. Fetzner Jr, 1999 (Fetzner 1999). A sample of snake slough, 1 cm^2 or smaller in size, was cut into 4 to 6 strips using scissors and pressed into a 1.5 ml microfuge tube using forceps. Cell lysis buffer (900 µl) and 20 µl of proteinase K (20 mg/ml) was added to each sample tube. The sample was then mixed gently by inverting the tube repeatedly for 10 seconds and placed within a heat block set to 55°C and allowed to digest over 12-24 hours. The digestion was gently mixed by inversion approximately six times during the incubation period. During the digestion process it was observed visually that the snake slough did not completely dissolve. After 24 hours, the sample was centrifuged at 15,000 g for 15 minutes. The sample was mixed by gently inverting the tube and centrifuged again at 15,000 g for a further 15 minutes. The supernatant phase of the sample was removed from the undigested slough remnants and placed into a clean 1.5 ml microfuge tube. RNase A (4 µl, 10 mg/ml) was added to the sample and incubated for 1 hour at 37°C.

After RNA digestion, 300 µl of 7.5 M ammonium acetate was added and the sample mixed through repeated inversion for 10 seconds, then placed on ice for 30 minutes. The sample was centrifuged at 15,000 g for 15 minutes, pelleting the protein, cell debris and salt. The supernatant was removed and placed into a fresh 1.5 ml microfuge tube. To ensure all the precipitate had been removed, the sample was centrifuged again at 15,000g for 15 minutes. The second supernatant was removed and placed in a fresh 1.5 ml microfuge tube.

The supernatant was divided evenly across two microfuge tubes, with approximately 600 µl in each and 900 µl of ice-cold isopropanol was added to each tube. The contents were mixed gently by inverting for 10 seconds. At this point a DNA precipitate was visible in the isopropanol solution. The sample was chilled at -20°C for at least 2 hours. The sample was then mixed gently and centrifuged at 15,000 g for 15 minutes. The precipitated DNA formed a visible pellet at the bottom of the tube.

Isopropanol was decanted from the sample, without disturbing the DNA pellet and the sample tubes were left open and inverted on a clean paper towel to allow any remaining isopropanol to evaporate. When there was no visible isopropanol left in the tubes, 500 µl of 70% ethanol was added to each tube gently to avoid disturbing the DNA pellet. The tubes were left to sit at room temperature for 15 minutes before again being centrifuged at 15,000 g for 15 minutes. The ethanol was removed making sure not to disrupt the DNA pellet. The 70% ethanol wash was repeated, and the tube left open to allow any remaining ethanol to evaporate. The DNA pellet was then resuspended in 100 µl of TE buffer.

RAPD PCR and Visualisation

Random Amplified Polymorphic DNA (RAPD) 10-mer primer sequences were obtained from the official Eurofins MWG Operon website. Using these sequences, 102 RAPD primers were constructed by Geneworks (Table 2-1).

Table 2-1 RAPD primers

KIT A

PRIMER NAME	Primer Sequence
OPA-01	CAGGCCCTTC
OPA-02	TGCCGAGCTG
OPA-03	AGTCAGCCAC
OPA-04	AATCGGGCTG
OPA-05	AGGGGTCTTG
OPA-06	GGTCCCTGAC
OPA-07	GAAACGGGTG
OPA-08	GTGACGTAGG
OPA-09	GGGTAACGCC
OPA-10	GTGATCGCAG
OPA-11	CAATCGCCGT
OPA-12	TCGGCGATAG
OPA-13	CAGCACCCAC
OPA-14	TCTGTGCTGG
OPA-15	TTCCGAACCC
OPA-16	AGCCAGCGAA
OPA-17	GACCGCTTGT
OPA-18	AGGTGACCGT
OPA-19	CAAACGTCGG
OPA-20	GTTGCGATCC

KIT B

PRIMER NAME	Primer Sequence
OPB-01	GTTTCGCTCC
OPB-02	TGATCCCTGG
OPB-03	CATCCCCCTG
OPB-04	GGACTGGAGT
OPB-05	TGCGCCCTTC
OPB-06	TGCTCTGCCC
OPB-07	GGTGACGCAG
OPB-08	GTCCACACGG
OPB-09	TGGGGGACTC
OPB-10	CTGCTGGGAC
OPB-11	GTAGACCCGT
OPB-12	CCTTGACGCA
OPB-13	TTCCCCCGCT
OPB-14	TCCGCTCTGG
OPB-15	GGAGGGTGTT
OPB-16	TTTGCCCGGA
OPB-17	AGGGAACGAG
OPB-18	CCACAGCAGT
OPB-19	ACCCCCGAAG
OPB-20	GGACCCTTAC

KIT C

PRIMER NAME	Primer Sequence
OPC-01	TTCGAGCCAG
OPC-02	GTGAGGCGTC
OPC-03	GGGGGTCTTT
OPC-04	CCGCATCTAC
OPC-05	GATGACCGCC
OPC-06	GAACGGACTC
OPC-07	GTCCCGACGA
OPC-08	TGGACCGGTG
OPC-09	CTCACCGTCC
OPC-10	TGTCTGGGTG
OPC-11	AAAGCTGCGG
OPC-12	TGTCATCCCC
OPC-13	AAGCCTCGTC
OPC-14	TGCGTGCTTG
OPC-15	GACGGATCAG
OPC-16	CACACTCCAG
OPC-17	TTCCCCCCAG
OPC-18	TGAGTGGGTG
OPC-19	GTTGCCAGCC
OPC-20	ACTTCGCCAC

KIT D

PRIMER NAME	Primer Sequence
OPD-01	ACCGCGAAGG
OPD-02	GGACCCAACC
OPD-03	GTCGCCGTCA
OPD-04	TCTGGTGAGG
OPD-05	TGAGCGGACA
OPD-06	ACCTGAACGG
OPD-07	TTGGCACGGG
OPD-08	GTGTGCCCCA
OPD-09	CTCTGGAGAC

KIT E

Primer Name	Primer Sequence
OPE-01	CCCAAGGTCC
OPE-02	GGTGCGGGAA
OPE-03	CCAGATGCAC
OPE-04	GTGACATGCC
OPE-05	TCAGGGAGGT
OPE-06	AAGACCCCTC
OPE-07	AGATGCAGCC
OPE-08	TCACCACGGT
OPE-09	CTTCACCCGA

KIT AB

Primer Name	Primer Sequence
OPAB-09	GGGCGACTAC
OPAB-18	CTGGCGTGTC

OPD-10	GGTCTACACC		**OPE-10**	CACCAGGTGA
OPD-11	AGCGCCATTG		**OPE-11**	GAGTCTCAGG
OPD-12	CACCGTATCC		**OPE-12**	TTATCGCCCC
OPD-13	GGGGTGACGA		**OPE-13**	CCCGATTCGG
OPD-14	CTTCCCCAAG		**OPE-14**	TGCGGCTGAG
OPD-15	CATCCGTGCT		**OPE-15**	ACGCACAACC
OPD-16	AGGGCGTAAG		**OPE-16**	GGTGACTGTG
OPD-17	TTTCCCACGG		**OPE-17**	CTACTGCCGT
OPD-18	GAGAGCCAAC		**OPE-18**	GGACTGCAGA
OPD-19	CTGGGGACTT		**OPE-19**	ACGGCGTATG
OPD-20	ACCCGGTCAC		**OPE-20**	AACGGTGACC

Polymerase chain reaction was performed in a total volume of 15 µl containing 0.625 units of hot-start Taq DNA polymerase (Bioline), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1.5 µl of 10× PCR buffer (Bioline), 0.2 mg/ml BSA, 0.4 µM of primer and 30 ng of template DNA. The thermal profile was 95°C for 10 min, 5 cycles of 94°C for 1 min, 42°C for 1 min and 72°C for 2 min, followed by 40 cycles of 94°C for 30 sec, 36°C for 40 sec, 72°C for 2 min and a final extension at 72°C for 5 min.

Bulked Segregation Analysis (BSA) was investigated by creating two pools of genomic DNA. Pool one contained 1 µg of genomic DNA from four different male samples. Pool two contained 1 µg of genomic DNA from four different female samples. Each pool was PCR amplified, using 30 ng of the pooled DNA as the template DNA, under the same PCR conditions detailed above. All 8 individuals that constituted both DNA pools were also PCR amplified individually, under the same PCR conditions at the same time, as well as one negative control (no template DNA). The results from all 11 PCR samples were size separated by electrophoresis on a 1% w/v agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide and visualized using ultraviolet transillumination. This process was repeated for the RAPD primers OPA-01, OPA-02, OPA-03, OPA-04 and OPA-05 to assess the viability of the method.

The final RAPD analysis used the genomic DNA from seven snakes (4 female and 3 male), for PCR amplification individually using each RAPD primer. With the addition of a negative control, the 8 PCR products were size separated by electrophoresis on a 1% w/v agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide and visualized using ultraviolet transillumination.

Investigation of Sex Associated RAPD Fragments

The RAPD fragments observed within the RAPD fragment profiles that showed sex association, that is they were present in all individuals of the same sex but absent from individuals of the opposite sex, were investigated further. The first step was to confirm the possible sex-linked nature of the amplification products. This involved repeating the RAPD PCR analysis with DNA extracted from all 14 individual *M. spilota* sloughs of a known sex.

Only one RAPD fragment was found to be sex-linked. Produced when amplifying genomic DNA using the RAPD primer OPA-17. All sex-associated RAPD fragments were excised from the agarose gel and placed into individual 1.5 ml microfuge tubes. The DNA was extracted from gel slices using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. The DNA extracted was DNA sequenced by Macrogen Korea, 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511.

The DNA extracted from the agarose gel slices from one female individual was inserted into the bacterial vector pCR™4-TOPO™ using the TOPO™ TA Cloning™ Kit for Sequencing (Thermofisher Scientific), with One Shot™ TOP10 Chemically Competent *Escherichia coli* (Thermofisher Scientific), according to the manufacturer's instructions. The transformed *E. coli* was suspended in 1 ml SOC media, incubated at 37°C for 1 hr and then 50 µl was plated onto LBA plates containing 100 µg/ml ampicillin and lastly incubated at 37°C for 24 hours. Ten white colonies representing each RAPD fragment isolate were picked and grown overnight at 37°C in 50 ml LB broth containing 100 µg/ml ampicillin. Plasmid DNA was extracted using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's protocol. The extracted plasmid DNA was visualised on 0.7% agarose gels, to confirm that the plasmids contained inserted DNA. DNA sequencing was performed on the plasmid DNA by Macrogen Korea, 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511.

Results

DNA was successfully extracted from all the 17 *M. spilota* sloughs. The original reptile skin DNA extraction method described by Fetzner (Fetzner 1999) resulted in DNA extractions that appeared to be very high in protein content (Figure 2-2). Protein contamination resulted in inconsistent RAPD PCR results, which was attributed to the DNA degrading quickly and not surviving freeze/thawing process more than 5 times. It was observed that a 1 $inch^2$ piece of snake slough for the DNA extraction, suggested in the original protocol, was too large to produce a DNA extraction free from protein contamination. The protein concentration obtained with the original technique was reduced by decreasing the size of the starting material from 1 $inch^2$ (6.45 cm^2) to 1 cm^2. The protein

contamination was further reduced by removing the snake slough remnants immediately after the 24-hour digestion and a more effective separation of protein in the protein precipitation step. Although this resulted in a decrease in DNA yield, the final DNA extraction purity was improved and resulted in better performance and more consistent results in the subsequent RAPD experiments despite the DNA extracted being frozen and thawed multiple times. The degree of contamination, from protein and other contaminants, was assessed by measuring the absorbance ratio of all DNA extractions at 260/280 nm and 260/230 nm using a NanoDrop spectrophotometer. Using this modified technique, samples had a 260/280 nm absorbance ratio of 1.8 or higher and a 260/230 nm absorbance ratio of 2.0-2.3, indicating a product that was relatively free from protein and other contamination. The concentration of DNA extracted from each slough sample was estimated to be approximately 50-75 ng/µl, giving a total DNA yield of approximately 10-15 µg of DNA per cm^2 of slough.

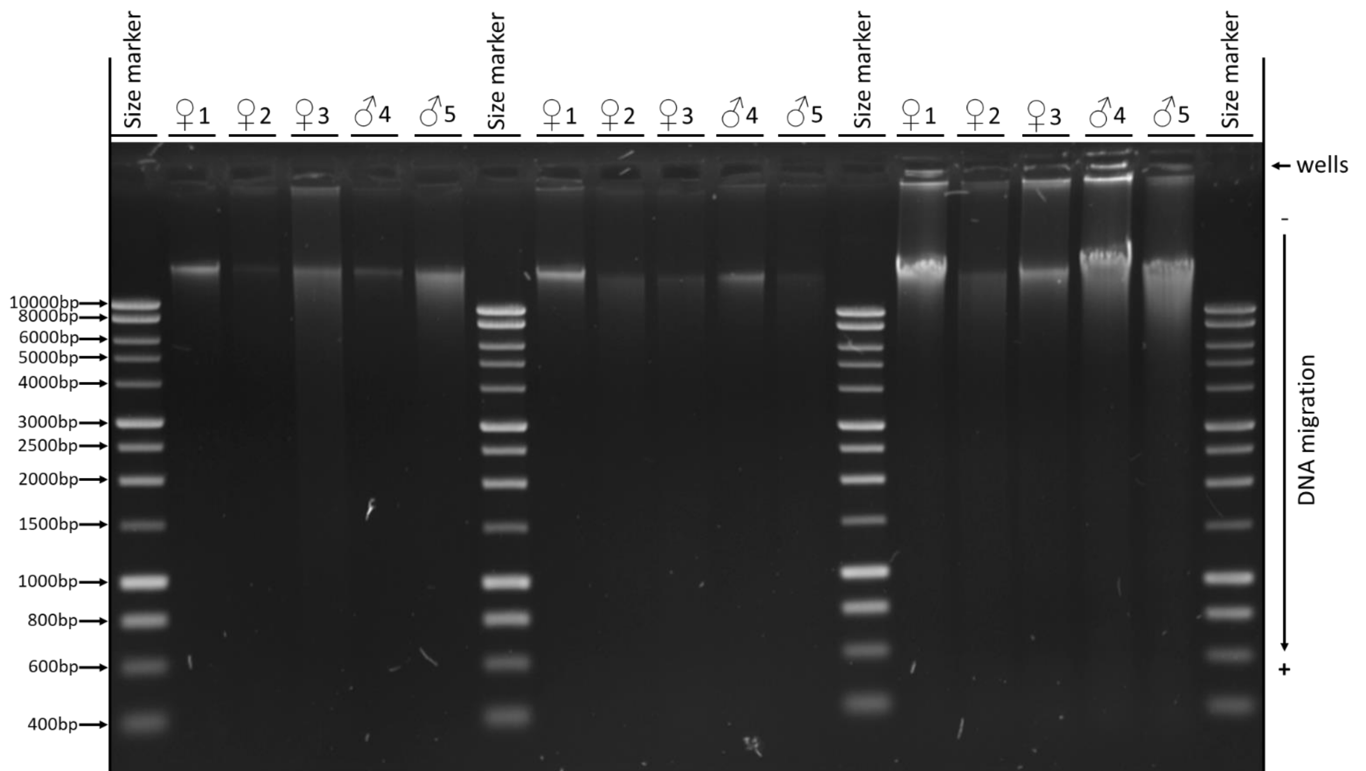

Figure 2-2 Snake slough DNA extraction

Figure 2-2 shows three groups of DNA extraction from three different regions of the snake sloughs were electrophoresed on an agarose gel. Lanes 2 to 6 (left to right); DNA extraction using the small scales of the head 1cm^2 pieces. Lanes 8 to 12; using large scales from the mid-section 1 cm^2 pieces. Lanes 14 to 18 large scales from the mid-section 2.5 cm^2 pieces. Five snake sloughs were used in the extraction, three females and two males.

RAPD Analysis

RAPD analysis was performed on DNA extracted from 7 snakes, 3 males and 4 females, with a negative control included, giving a total of 8 PCR reactions (Figure 2-3). RAPD primers from kits A, B, C, D and E were investigated, along with 3 others primers, OPD-10, OPAB-09 and OPAB-18, which have been able to amplify sex specific fragments in other species (Bello and Sanchez 1999; Leon Huynen, Craig D. Millar, and David M. Lambert 2002; Griffiths et al. 1998; Griffiths and Tiwari 1993). This gave a total of 102 RAPD primers, which were tested in the initial study (Table 2-1). Each set of RAPD fragment profiles was visualised using agarose gel electrophoresis.

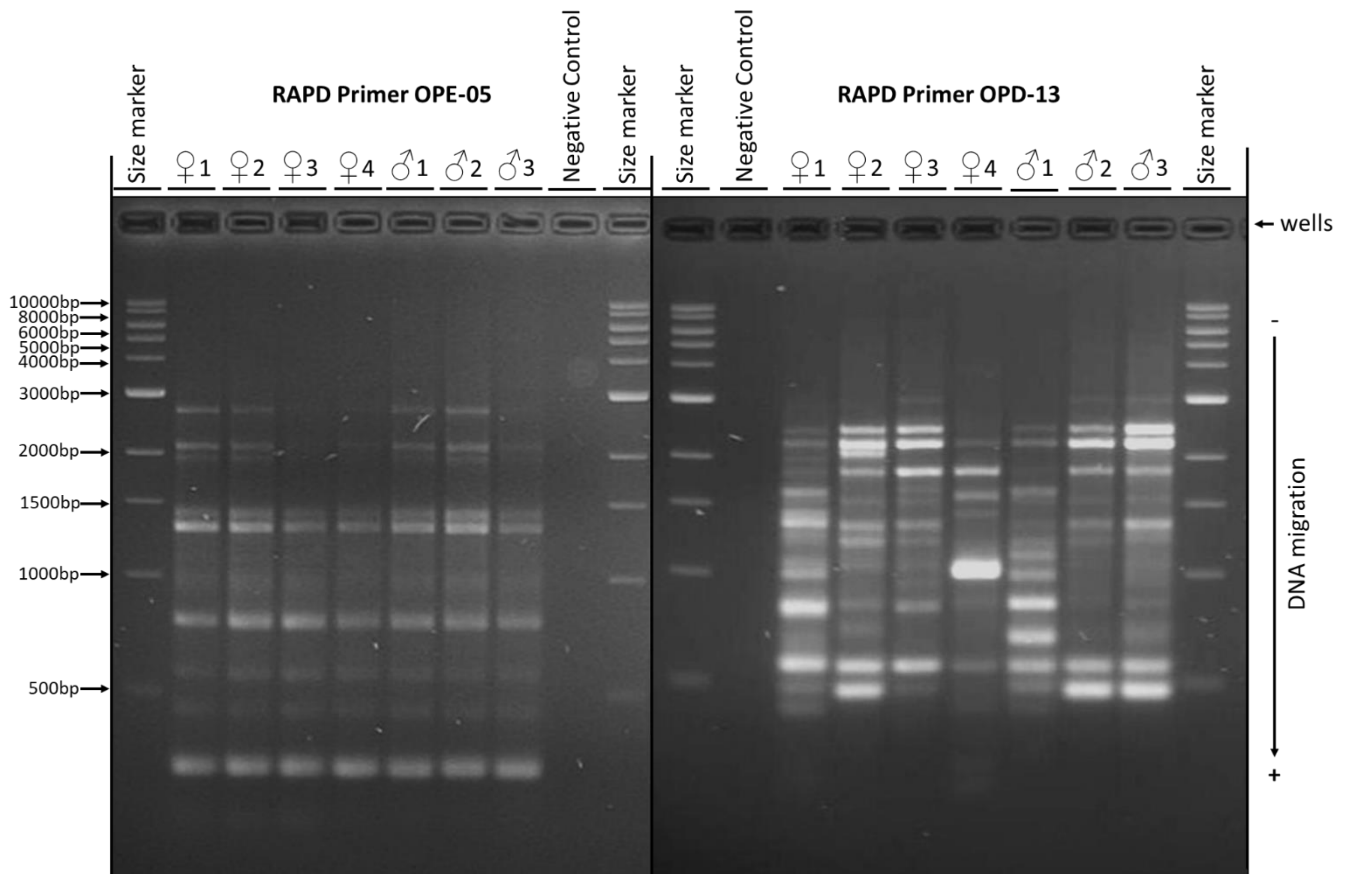

Figure 2-3 RAPD fragment profile of OPE-05 and OPD-13

Figure 2-3 show a comparison of the RAPD fragment profile produced for two RAPD primers. OPE-05, which shows almost no variation between the fragment profiles of individuals tested. OPD-13, which shows great variation in the fragment profile produced between individuals tested. The DNA from the same three males and four females was used in each of these RAPD profiles as labelled, allowing direct comparison between the RAPD primers.

The 102 RAPD primers produced 881 individual fragments, with the results summarised in Table 2-2. Of these, 221 (approximately 25%) appeared to be polymorphic. The average number of fragments observed per RAPD primer was 9. The amplified fragments were observed to range in size from 300bp to 2500bp, with an average observed size range of between 1,000bp and 1,500bp. However, smaller, and larger DNA fragments did occur infrequently, the smallest observed being approximately 100bp and the largest approximately 5000bp in length.

Table 2-2 RAPD fragments identified

RAPD primer sets	Total Number of Bands	Number of Polymorphic Bands	Average Number of Bands	Average Number of Polymorphic Bands
OPA	202	71	10	4
OPB	171	42	9	2
OPC	143	41	7	2
OPD	157	41	8	2
OPE	188	23	9	1
Total (102 RAPDs)	881	221	9	2

Python reticulatus has a genome which is estimated to be approximately 1.44Gb (Castoe, de Koning, et al. 2011; Gregory et al. 2007). *Python reticulatus* is the most closely related snake species to *M. spilota* that has had its genome size investigated (Castoe, de Koning, et al. 2011; Gregory et al. 2007). With the RAPD fragments having an average size range of between 1,000bp and 1,500bp and over 881 bands identified, it can be estimated that the RAPD PCR sets used amplified between 881,000bp and 1,321,500bp of the *M. spilota* genome. Using the genome size of *Python reticulatus* as a reference, it was determined that between 0.06% and 0.09% of the *M. spilota* genome was analysed.

For all the 102 RAPD primers tested, only one DNA fragment was identified as being potentially sex-linked. This fragment was within fragment profiles generated using the RAPD primer OPA-17 and was approximately 440bp in length. The fragment is identified with an asterisk in Figure 2-4. This fragment was observed in all PCR amplifications of female DNA but was absent from amplification profiles of all male DNA tested (Figure 2-4). This amplification product was therefore subject to further investigation.

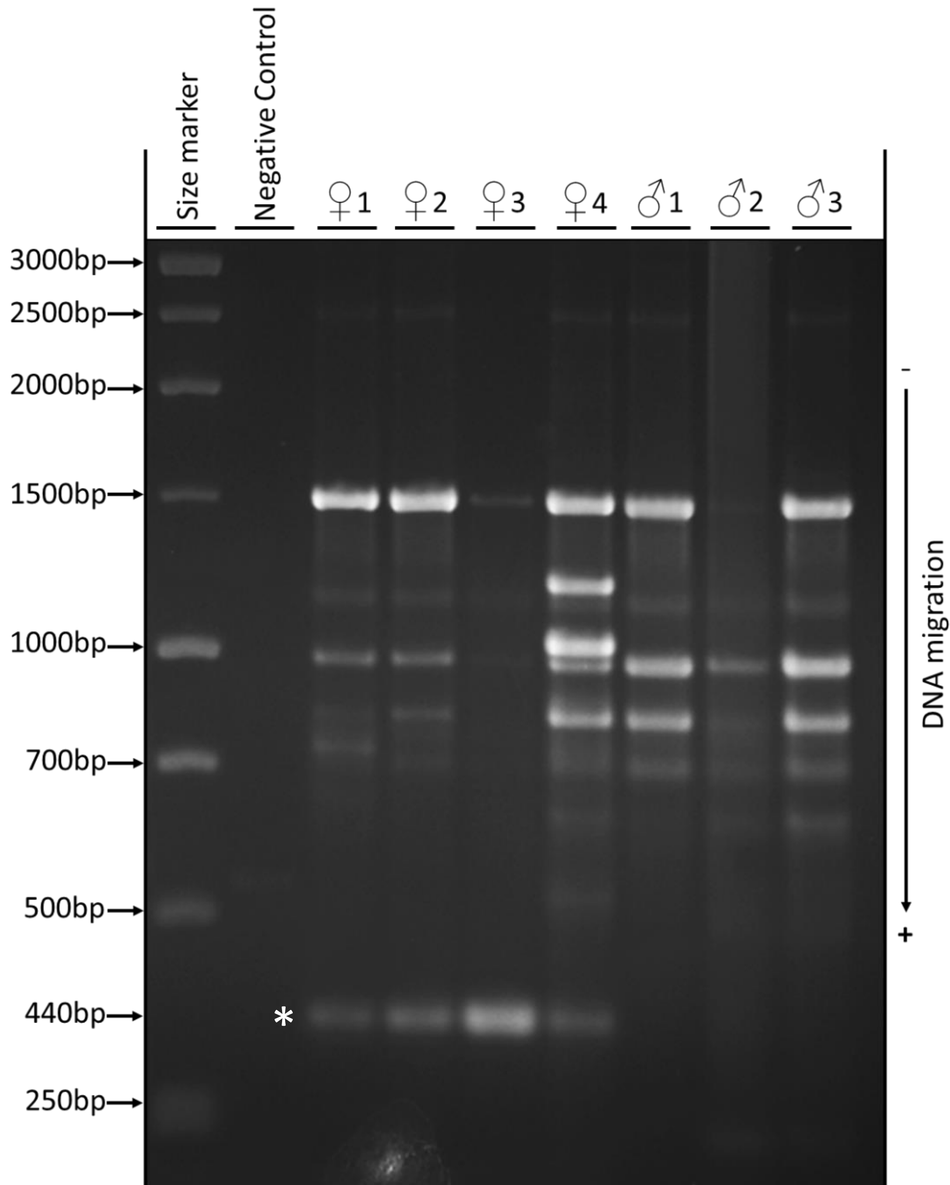

Figure 2-4 RAPD fragment profile OPA-17

*Figure 2-4 shows that RAPD fragment profile produced when amplifying genomic DNA using the RAPD primer OPA-17 from 7 individuals; 4 females and 3 males. In this initial experiment a 440bp DNA fragment was identified that is present in all 4 female individuals amplified (identified by the *) but absent from all 3 males amplified.*

For each of the 4 females investigated, the 440bp female-associated DNA fragment was excised from the agarose gel, column purified and sequenced by Macrogen, Korea. However, this sequence did not produce readable sequencing results. Therefore, the purified DNA fragment from female 4 was ligated into the cloning vector pCR™4-TOPO™ using the TOPO™ TA Cloning™ and used to transform TOP10 *E. coli*. The DNA from 10 plasmid clones were purified and sequenced. The sequencing results

produced from the individual clones were inconsistent. Some clones did not produce sequencing outputs that were readable, while others were only partially readable. Readable forward and reverse sequencing results for four clones were produced. Two of the four clones produced the exact same readable DNA sequence, as shown in Figure 2-5.

Analysis of the inserted DNA sequence showed that it had a predicted length of 440bp and contained three separate microsatellite DNA regions; a TG dinucleotide microsatellite containing 20 repeats, a G mononucleotide microsatellite containing 11 repeats and a CTGT tetranucleotide microsatellite containing 7 repeats. The two other clones with readable DNA sequence, differed in number of microsatellites repeats they contained. Sequence analysis identified two complimentary DNA regions 22bp in length at the beginning and end of the DNA sequence. However, no DNA sequence matching the RAPD primer OPA-17 used to amplify this genomic DNA fragment was observed, without using mismatched DNA nucleotides.

Using the mega-BLAST algorithm analysis no similarity sequences were identified in GenBank (NCBI Resource Coordinators 2017). However, using the BLASTn algorithm identified a variety of different DNA sequence from different vertebrate species. Most sequences showed sequence similarity to the three microsatellite repeats within the 440bp DNA fragment. Interestingly, this analysis also showed three matching sequences for the short complimentary DNA region at the beginning and the end of the DNA fragment. Sequence similarity with a *Verasper moseri* (Barfin flounder) microsatellite marker (AB434984) was identified. The sequence had 100% sequence identity for the first 24bp of the DNA fragment and 96% (27/28) for the last 28bp. As well as showing 41/49 (84%) sequence identity for *Moniliophthora roreri* (fungus) microsatellite DNA, clone Mr31 (FN252888), for the first 49bp of the DNA sequence. Further BLAST analysis of just the 22bp complimentary DNA sequence showed 100% sequence identity to *Oropsylla hirsuta* (Flea; FJ233086), *Podarcis erhardii* (Erhard's wall lizard; AY924400) and again *V. moseri* (AB434984). Analysis using RepeatMasker identified the simple repeats within the DNA fragment, but no other repetitive elements were identified.

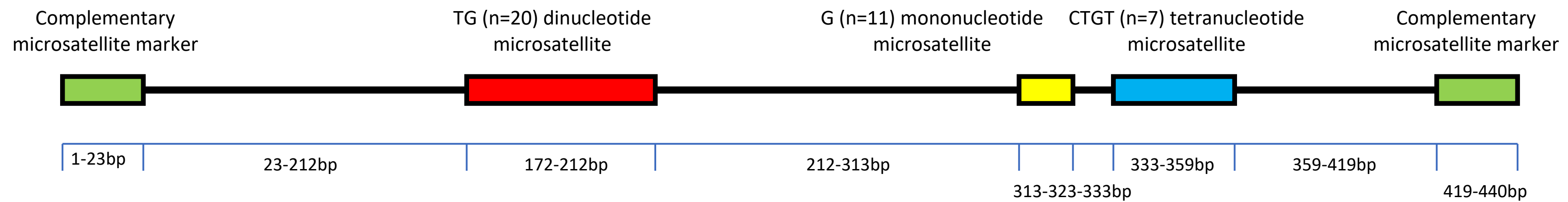

Figure 2-5 Diagrammatic and sequence representation DNA fragment produced by OPA-17

Figure 2-5 shows a both a diagrammatic and a DNA sequence representation of the potentially sex-linked 440bp DNA fragment identified from the RAPD fragment profiles produced by the primer OPA-17. Important features include the three polymorphic microsatellite regions and the complimentary 22bp regions at the beginning and end of the DNA sequences, as identified in the figure.

Due to the repetitive nature of the DNA sequence identified, no suitable specific PCR primers could be identified for the DNA region. Instead the RAPD primer OPA-17, which originally produced the female-associated DNA fragment, was used to PCR amplify genomic DNA templates from all 17 known sex *M. spilota* individuals. The initial PCR resulted in fragment profiles showing differences in overall fragment intensity between some individuals (female 3 and male 2), as shown in Figure 2-4. For this reason, the RAPD PCR was performed again, with fresh DNA extractions from these individuals and DNA extractions from 10 individuals that had not been previously analysed. The amount of template DNA in the PCR was increased from 30 ng to 60 ng to better define the 440bp fragment intensity differences. Using this modified PCR protocol, the 440bp DNA fragment was observed in all male and female samples (Figure 2-6). However, there appeared to be two distinct DNA intensities for this fragment when observing the RAPD fragment profiles produced. The previously used female samples; females 1, 2, 3 and 4 all showed a DNA intensity approximately double that of the previously tested male samples: males 1, 2 and 3. While the 10 new individual samples did not follow the same sex-linked pattern. Female 5 was observed to have a DNA fragment intensity of approximately half of other female samples, while females 6 and 7 were comparable with other female samples. Males 7, 8 and 9 showed results that were indicative of previously tested male samples, with DNA fragment intensity approximately half of females. While males 4, 5, 6 and 10 had a DNA fragment intensity like the previously tested female samples. The results showed that 6 out of 7 females were observed to have a DNA fragment intensity that was greater, whilst 5 out of 10 males were observed to have DNA fragment intensity approximately half that of most female samples.

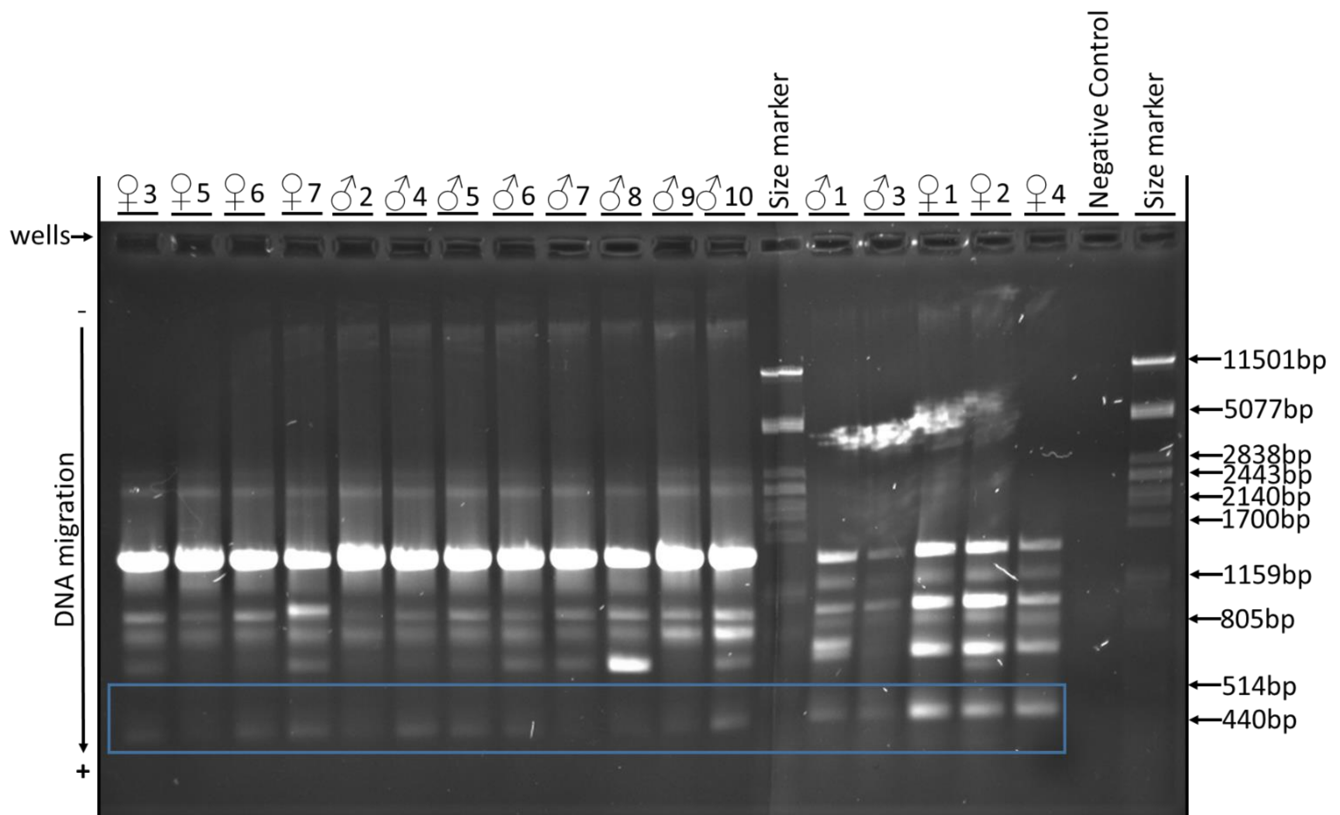

Figure 2-6 RAPD OPA-17, 17 M. spilota templates of previously determined sex

Figure 2-6 shows the agarose electrophoresis gel results produced from the secondary RAPD PCR using the primer OPA-17. The blue box highlights the potentially sex-linked 440bp fragment identified. Ten male and seven female DNA samples were used in the experiment as labelled.

Discussion

The use of snake slough as a source of DNA was very successful throughout this experiment. With this method of DNA extraction, we were able to isolate approximately 10-15 μg of DNA, free from protein and other organic contamination. Furthermore, it had a high average molecular weight and appeared to be stable during freeze/thawing. Provided the DNA extraction was performed on a piece of snake slough no larger than 1cm^2 and the DNA sample produced was free of protein contamination. The DNA extraction produced consistent and reproducible results during RAPD analysis, indicating its suitability for this type of experiments. In addition, only a very small amount of starting material was required for the extraction, which allowed multiple DNA extraction to be performed from a single sampled slough. This method of DNA extraction is advantageous over other commonly used methods because it is inexpensive, relatively quick, simple to perform and does not require the use of hazardous chemicals such as organic solvents (Fetzner 1999; Bricker, J. et al. 1996; Clark, A 1998; Eguchi and Eguchi 2000).

Alternative sampling techniques, many of which are invasive, such as scale clipping or tail tissue sampling leaves snakes permanently disfigured. Consequently, reptile owners, such as reptile parks and zoos, are usually unwilling to allow this type of tissue sampling to be performed on their animals. Other non-disfiguring sampling techniques, such blood sampling and mouth swabs, require specialised expertise such as a veterinarian to perform sampling and are still invasive to the animal. This method of sampling the snake's DNA by using their discarded slough is non-invasive and does not require veterinarian expertise. Because sloughing is a natural process for snakes, sample collection does not affect the animal's health or welfare and in the case of sampling venomous snakes, is not dangerous to the person performing the sampling. Furthermore, we found that this method was favoured by the snake keeping community and allowed them to have direct involvement in the research, by collecting, drying, labelling, and storing the sloughs. This engaged their interest, which was positive for our relationship and ultimately led to acquiring more sloughs for DNA sampling and further research collaboration.

This DNA extraction method can also be used in the field when trying to collect DNA samples of rare, elusive or endangered snake species (Maxine P. Piggott and Andrea C. Taylor, n.d.; Jose L. Horreo, Maria L. Pelaez, and Patrick S. Fitze 2015). Previous research has shown that DNA can be obtained from snake sloughs exposed to the elements (Clark, A 1998). It was observed that the DNA within the snake slough was very stable over time, supporting this finding. Provided the slough was dried, free from faecal matter and stored correctly, away from insects, we managed to extract usable DNA from sloughs that were over 3 years old. The DNA extracted was reduced in quality and appeared to be increasingly fragmented with age. However, after 3 years, the DNA extracted was still usable for PCR applications, such as RAPD analysis. If the sample DNA needs to be stored longer than a period of three years, we suggest that enough DNA is extracted from the sample when the snake slough is initially recovered, and the DNA stored pure, as DNA can be considered more stable.

Random Amplified Polymorphic DNA Analysis

Bulked Segregation Analysis (BSA) is a technique that is commonly coupled with RAPD analysis, used to identify sex-linked markers (Michelmore, Paran, and Kesseli 1991). The BSA technique involves bulking the DNA of males together and females together, creating two sex specific DNA pools (Michelmore, Paran, and Kesseli 1991). This allows for the investigating of observable differences between the RAPD fragment profiles produced from a male and female population, rather than investigating differences between multiple individual males and females separately, increases the

overall efficiency of analysis (Michelmore, Paran, and Kesseli 1991; Hormaza, Dollo, and Polito 1994; Bello and Sanchez 1999). It is theorised that the pooling of the two population will result in the collection of all possible individual polymorphisms in the two populations and the only consistent difference between the individuals that make up each population will be the sex-linked locus. The BSA technique has been successful in identifying sex-linked markers in *Actinidia chinensis* (Kiwi fruit), *Pistacia vera* (Pistachio), *Humulus lupulus* (Hop) and the *Struthio camelus* (Ostrich) (Bello and Sanchez 1999; Gill, Harvey, and Fraser 1998; Hormaza, Dollo, and Polito 1994; Polley, Ganal, and Seigner 1997).

Initially, the coupling of the BSA technique was assessed in the RAPD analysis, by investigating the suitability of the BSA technique. The RAPD fragment profile produced from male and female pools were compared and in the various individuals used to constitute the pools. It was observed that some RAPD fragments, were present in less than 50% of the individuals making up that pool, were not identifiable within the RAPD fragment profile produced using the pooled populations. Similarly, RAPD fragments of a lower than average intensity, which were found in more than 50% of individuals from the pooled population, were sometimes completely absent from the DNA fragment profile produced from the pooled population. This finding indicated that the BSA technique in some cases, resulted in RAPD fragments that are present in multiple individuals, being absent from the RAPD fragment profile produced when using a pooled population and therefore were not identified. As the sex determination system of snakes is still undefined and the BSA technique has the potential to over-simplify the RAPD results, the potential to miss a sex-linked marker was high. Therefore, a decision was made to not incorporate the BSA technique into the RAPD experiments.

In the present experiments, 102 individual RAPD primers were used to produce RAPD fragment profiles from the DNA of 7 individual snakes. The decision was made to investigate 100 RAPD primers, a number that fitted well within previously published research, investigating sex-linked DNA markers using RAPDs, as represented in Table 2-3. Two additional RAPD primers were included from previously published research bringing the final number to 102 RAPD primers investigated.

Table 2-3 RAPD primers investigated in published research

Species Tested	Number of RAPDs Investigated	Outcome	Reference
***Piper longum* (Long Pepper)**	40	Successful	(Banerjee, Manoj, and Das 1999)
***Salix viminalis* (Basket Willow)**	88	Successful	(Semerikov et al. 2003)
***Tetraodon nigroviridis* (Green Spotted Pufferfish)**	2300	Unsuccessful	(Y. Li et al. 2002)
***Gracilaria gracilis* (Red Seaweed)**	69	Successful	(Martinez E. A. et al. 2003)
***Pistacia vera* (Pistachio)**	700	Successful	(Hormaza, Dollo, and Polito 1994)
***Simmondsia chinensis* (Jojoba)**	60	Successful	(Agrawal et al. 2007)
***Carica papaya* (Papaya)**	25	Successful	(Urasaki et al. 2002)
***Huso huso* (Beluga)**	310	Unsuccessful	(Keyvanshokooh, Pourkazemi, and Kalbassi 2007)
***Struthio camelus* (Ostrich)**	200	Successful	(Bello and Sanchez 1999)
***Oncorhynchus mykiss* (Rainbow Trout)**	900	Successful	(Iturra et al. 1997)

It was observed that different sets of RAPD primers produced differing variability between individuals within the test group. It was observed that some primers produced no discernible difference in the fragment profiles between all 7 individuals tested, however other primers produced fragment profiles that appeared highly variable between individuals (Figure 2-3). In some instances, the fragment profiles produced by a single RAPD primer were variable enough to differentiate between all 7-individual snakes within the test group and could be considered almost hypervariable. These RAPD primers could possibly be applied in future studies to determine systematic relationships (Handrys, Balick, and Schierwater 1992). For example, RAPD primers that produce constant results between individuals could possibly be used to identify snakes of a genus, species or sub-species level (Handrys, Balick, and Schierwater 1992; Bardakci and Skibinski 1994). Specifically, for this snake species, RAPD primers could be applied to identify sub-species of *M. spilota*, of which there are 6, which are presently known to interbreed both in the wild and in captivity (Handrys, Balick, and Schierwater 1992; Bardakci and Skibinski 1994; Hoser 1999). They could also be used to infer relatedness and identify unique populations of sub-species and isolated island populations, known to have unique physical features (Handrys, Balick, and Schierwater 1992; D Pearson, Shine, and How 2002; David Pearson, Shine, and Williams 2002). Similarly, RAPD primers that produced highly variable results could be used to identify individuals within a given population, similar to a genetic fingerprint (Handrys, Balick, and Schierwater 1992). More specifically, highly variable RAPD primers could identify individuals within a population that have produced offspring

asexually through parthenogenesis (Handrys, Balick, and Schierwater 1992). Parthenogenesis has been observed to occur in snakes held within captivity but has never been confirmed in the wild (Groot, Bruins, and Breeuwer 2003). However, it is important to note that all experiments used to determine relatedness are only valid between taxa and individuals that have had their RAPD fragment profiles determined.

Each RAPD profile generated was examined in detail to identify any DNA fragment that was observed only in one sex. Indeed, if the current theory of snake sex determination is correct, then we would expect to find a RAPD product that is present in all females but absent in all males. This would suggest that the product band is most probably from a unique portion of the proposed W chromosome and could possibly be linked to a sex determining locus in snakes (Matsubara et al. 2006). However, because distinguishable sex chromosomes have never been observed in our model snake species *M. spilota* and no sex determination system has been shown to be conserved amongst all snake species, the opposite could also be possible, in that a RAPD product may be specific only in males and absent in females.

RAPD Primer OPA-17

One RAPD primer, OPA-17, was identified as producing what appeared to be a female-associated RAPD fragment of 440bp in length. This DNA fragment was present in the 4 females but absent in the 3 males used for the initial RAPD investigation (Figure 2-4). Further investigation using OPA-17 was conducted on all 17 individuals of known sex, using altered RAPD PCR conditions with double the concentration of template DNA (Figure 2-6). The secondary investigation showed the DNA fragment could be amplified from the genomic DNA irrespective of the sex. This suggests that the DNA region is present in all individuals tested. However, the DNA fragment was observed to be present in varying DNA intensities using agarose gel electrophoresis. Six of the 7 females tested produced a DNA fragment of moderate intensity, while only 5 out of 10 males produced a moderate intensity DNA band. These results indicate that the DNA fragment produced by OPA-17 is unlikely to be sex specific but has the potential to be female-associated.

Agarose gel extraction, cloning and sequencing of the RAPD fragment from one female individual, reveal the DNA sequence to be 440bp in length. Sequencing revealed the region contained three distinct microsatellites. DNA sequencing of multiple clones produced from the same female fragment revealed at least three different copies of the DNA region, which presented with a different number of microsatellite repeats, but did not differ in 22bp complimentary DNA sequence.

This result is evidence that the 440bp DNA region is present within the genome of female *M. spilota* in at least 2 copies, but most likely more.

The 440bp sequence identified also contained 22bp complimentary DNA sequences. These elements span 22 nucleotides at the 5-prime and 3-prime end of the DNA sequence and were 100% complementary to one-another, which is a similar feature observed in retrotransposon elements (Finnegan 2012). Nucleotide BLAST search analysis of this short sequence identified similar elements found to be associated with polymorphic microsatellite regions in a *Oropsylla hirsuta* (a species of flea), *Verasper moseri* (Barfin flounder) and the more closely related *Podarcis erhardii* (Erhard's wall lizard) (Jones, Wallace, and Bitten 2009; Poulakakis et al. 2005).

Using the DNA sequence data obtained an attempt was made to designed PCR primers specific for the DNA region being investigated. However due to the repetitive nature of the OPA-17 RAPD fragment, including the presents of three microsatellite and a potential transposon element at each end of the sequence, PCR primers that would amplify the DNA region with confidence were unable to be designed. It was decided that the only PCR primer that would confidently amplifying the correct DNA region being investigated was the original OPA-17 RAPD primer.

The absence of the RAPD fragment in original male samples and the potentially female associated RAPD fragment intensity identified in the secondary experiment, could be explained by three possible theories. It is possible that the results produced by the RAPD investigations are artifactual and caused by slight differences in template DNA concentrations used in each RAPD PCR reaction. Therefore, it may be coincidental that the initial RAPD investigation resulted in the DNA fragment being observed in the 4 female test samples and absent from the 3 male samples. The results produced when doubling the template DNA concentration in our secondary RAPD investigation tend to support this theory, because with an increase in starting material the DNA fragment was identified in all individuals tested despite their sex. This theory is further supported in the literature, with examples of low intensity RAPD fragments being influenced by slight differences in the concentration of template DNA (Handrys, Balick, and Schierwater 1992).

An alternative theory is that there are multiple copies of the 440bp DNA fragment in the *M. spilota* genome. This theory is supported by the results presented, which show that the DNA region is amplified in all individuals tested but has an inconsistent DNA fragment intensity between different individuals. Although the RAPD PCR results show that the DNA fragment is located within all males

and females tested, these results also show three different polymorphic examples of the region from a single female individual tested.

The DNA region, in most cases, was amplified at a stronger intensity in more females than males and therefore it is logical to suggest that one or more of the copies of this region may be female-linked. In the literature, there are some examples of RAPD fragments, most commonly observed in plant species, which are not sex specific, but are strongly associated with one sex (Mandolino et al. 1999; Sakamoto et al. 1995; Kafkas et al. 2015). Most examples of these RAPD fragments remain unstudied, because they cannot reliably discriminate between the sexes of individuals. One example of a male-associated RAPD fragment, with retrotransposon features identified within the fragment DNA sequence, was discovered in *Cannabis sativa L.* (Mandolino et al. 1999). Southern blotting revealed the DNA region to be present in both males and females, but RAPD analysis showed the amplification of the RAPD fragment to be repeatedly male-associated in most plant cultivars (Mandolino et al. 1999). The authors suggest small changes to the RAPD priming sites such as a single nucleotide polymorphism (SNP), could account for the observed results in the RAPD fragment profile, even though the DNA region is present in both sexes (Mandolino et al. 1999).

It is a general feature of the sex defining chromosomes, Y or W, to accumulate repetitive elements, the most common being retrotransposon related elements (Bachtrog et al. 2008). The 440bp DNA region may be an example of such a retrotransposon related element that has been appropriated by the non-recombining region of the sex defining chromosome. Examples of repetitive elements invading and proliferating on the W chromosomes of snakes with cytologically differentiated sex chromosomes, include; Bkm, DMRT1-related, CTNNB1-related, 18S-related, and chicken W chromosome similar sequences (O'Meally et al. 2010). The 22bp complimentary microsatellite marker at either end of the DNA fragment is a similar feature observed in retrotransposon elements (Finnegan 2012). In retrotransposons these DNA sequences function as priming sites for reverse transcription and are a direct result of the process of these elements inserting into the genome (Finnegan 2012; Adelson, Raison, and Edgar 2009; Xiong and Eickbush 1990). The hypothesis is that the 440bp DNA region identified may behave in a similar way, copying itself throughout the genome and one or more of these copies has become associated with the female sex.

The final theory is that *M. spilota* employs a male heterogametic sex chromosomes system, and the 440bp DNA region is localized to the X chromosomes. The results support this hypothesis with the DNA fragment being readily identified in female snakes initially. A more intensive investigation

showed the DNA region was present in all individuals tested, but appears to be PCR amplified to a generally higher intensity in females (XX), which may contain two copies, when compared to males (XY), which may only contain one. This theory is further supported in the literature, with research using RAD-seq showing a XX/XY GSD system likely exists in two snake species *Boa imperator* and *Python bivittatus*, belonging to the snake families Boidae and Pythonidae and shown to have cytologically indistinguishable sexes, like *M. spilota* (Gamble et al. 2017).

Conclusion

The RAPD technique relies upon genetic differences between individuals presenting a unique phenotypic trait. In this experiment, we hypothesised that there are genetic differences in RAPD profiles between male and female *M. spilota*. It is logical to assume that due to the random nature of the RAPD technique, the greater the genetic difference between the two sexes, the greater the chance of identifying a sex-linked RAPD marker. Using RAPD analysis to identify sex specific DNA markers has proved to be unsuccessful in four different fish species; Sturgeon, Atlantic Salmon, Beluga and Pufferfish (McGowan and Davidson 1998; Y. Li et al. 2002; Sven Wuertz et al. 2006; Keyvanshokooh, Pourkazemi, and Kalbassi 2007; Olmo and Signorino 2010). All four fish species used in these examples and our model species *M. spilota* lack cytologically distinguishable sex chromosomes (McGowan and Davidson 1998; Y. Li et al. 2002; Sven Wuertz et al. 2006; Keyvanshokooh, Pourkazemi, and Kalbassi 2007; Olmo and Signorino 2010). In this work we were unable to identify any sex specific RAPD markers. The lack of identifiable sex specific RAPD markers in this experiment, along with supporting cytological and genetic evidence, suggests that the genetic difference between snake species with cytologically indistinguishable sex chromosomes, such as *M. spilota* is very small or possibly even absent (McGowan and Davidson 1998; Y. Li et al. 2002; Sven Wuertz et al. 2006; Keyvanshokooh, Pourkazemi, and Kalbassi 2007; Olmo and Signorino 2010; Vicoso et al. 2013).

The 440bp DNA fragment produced from the RAPD primer OPA-17 has the potential to be female-linked and should be investigated in further detail. If the DNA region does exist as multiple copies in the genome of *M. spilota*, as predicted, identifying a female-associated allele using PCR and sequencing techniques would be very difficult. To investigate the sex specific nature of the DNA region, florescent *in situ* hybridisation could be used to map and compare the presence in both male and female karyotypes. This would allow the region’s chromosomal location and potentially its genomic copy number to be better understood. This might give greater insight to the fragment’s sex-association. To determine the DNA regions exact genomic location and potential function or

association with sex determination, the DNA surrounding the DNA region would have to be investigated, which may involve the use of a large insert genomic library.

Chapter 3 - Representational Difference Analysis

Introduction

Representational Difference Analysis (RDA) is a variation of the molecular genetics technique, subtractive hybridisation, first published in 1984 by Lamar and Palmer (Lamar and Palmer 1984). The technique combines subtractive hybridisation and PCR to generate a process of PCR-mediated kinetic enrichment that is 1000-fold more specific than subtractive hybridisation (Lamar and Palmer 1984; Lisitsyn, Lisitsyn, and Wigler 1993). The technique has been shown to be capable of identifying genetic differences between two complex genomes and has been used to identify viruses, genetic mutations and sex specific differences (Lisitsyn, Lisitsyn, and Wigler 1993). In this chapter, a Representational Difference Analysis method was applied to identify DNA sequences specific to one sex in the snake species *Morelia spilota*.

Representational Difference Analysis (RDA), like other subtractive hybridisation techniques, first requires two genetically different DNA populations (Lisitsyn, Lisitsyn, and Wigler 1993). The “tester”; a representation of a genome that contains the unique DNA sequences of interest and the “driver”; representing a genome that is closely related to the tester but is distinguishable to the tester by the lack of unique DNA sequences that are being sought. In the experimental work described within this chapter, the aim was to identify sex specific differences in the snake species *M. spilota*. Although it has been hypothesised that all snakes share a common female heterochromatic sex chromosome system, no sex chromosomes have previously been identified in our model snake species (Matsubara et al. 2006; Olmo and Signorino 2010). Therefore, the RDA experiment was performed twice. The first using the female genome as the tester and the male genome as the driver, targeting DNA sequences unique to females and the second using the reverse orientation to target DNA sequences unique to males.

It has been previously shown that individual polymorphisms can be a common issue when trying to determine differences between two closely related genomes using a subtractive hybridisation based technique such as RDA (Michelmore, Paran, and Kesseli 1991; Hormaza, Dollo, and Polito 1994). This is due to individual polymorphisms also being enriched which can then out-compete the DNA trying to be targeted for enrichment (Michelmore, Paran, and Kesseli 1991; Hormaza, Dollo, and Polito 1994). To prevent this problem in these current experiments, a bulked segregate approach, in which a pool of 8 individuals is used to create both tester and driver DNA populations. The tester and driver genome representations are created from genomic DNA isolated from 8 females and 8 males

of known sex. This reduces the possibility of individual polymorphisms also being enriched during the RDA experiment (Michelmore, Paran, and Kesseli 1991; Hormaza, Dollo, and Polito 1994; Donnison et al. 1996).

RDA Protocol Preparation

Representations of male and female genome pools were prepared using a modified method based upon two publications; Yuan Chang, 2002 and Lisitsyn N, Lisitsyn N and Wigler M, 1993 (Lisitsyn, Lisitsyn, and Wigler 1993; Yuan 2002). Both driver and tester DNA populations were digested with the same restriction endonuclease (RE). A pair of oligonucleotide adaptors, complementary to the single stranded DNA overhang resulting from the RE digestion, were ligated to both driver and tester digested DNA and amplified through PCR (Table 3-1). The adaptors of the driver amplicons were removed by digestion with the same RE used initially, and the DNA purified, ready for hybridisation. The adaptors of the tester amplicon were then removed, the tester amplicon was size selected (DNA fragments between 100bp and 2000bp), and the amplicon was ligated to a new set of oligonucleotide adaptors, complimentary to the initial RE site but differing in DNA sequence compared to the first set of adaptors utilised. The adaptor set was only covalently bound at the 5' end to the tester amplicon.

An important note is that the tester population has been PCR amplified and subject to size selection, therefore it no longer represents the entire genome of the animal sampled, but instead is a simplified smaller representation of the genome under investigation. The tester amplicon has a considerably reduced complexity compared with the original genome and therefore has the potential to be lacking the unique DNA sequences being targeted. For this reason, three different restriction endonucleases, all with different recognition sequences; Bgl II (5' A^GATCT 3'), Bam HI (5' G^GATCC 3') and Hind III (5' A^AGCTT 3'), were used in separate experiments. This decreases the chances of missing any unique DNA sequences in the tester amplicon.

RDA Hybridisation

The prepared tester and driver amplicons were combined, to a ratio of 1 of tester to 10 of driver, for the first round of hybridisation (Figure 3-1). The combined amplicon reaction was heat denatured at 98°C for 30 minutes, before the temperature was slowly reduced from 98°C to 72°C over a period of 30 minutes. This allowed complementary DNA strands to re-associate but not necessarily with their original complimentary DNA strand. The result produces three different types of double stranded DNA. Excess driver DNA will hybridise back with its complimentary driver DNA. Simultaneously, DNA

sequences common between the tester and driver amplicons will hybridise together and DNA sequences found only within the tester amplicon will rehybridise. Some single stranded DNA from both driver and tester amplicons, as well as the disassociated adaptor not covalently bound to the tester amplicon will also remain. By adding an excess of the driver amplicon, theoretically the complementary sequences between the driver and tester will become saturated, leaving only DNA fragments unique to the tester to be available for PCR enrichment.

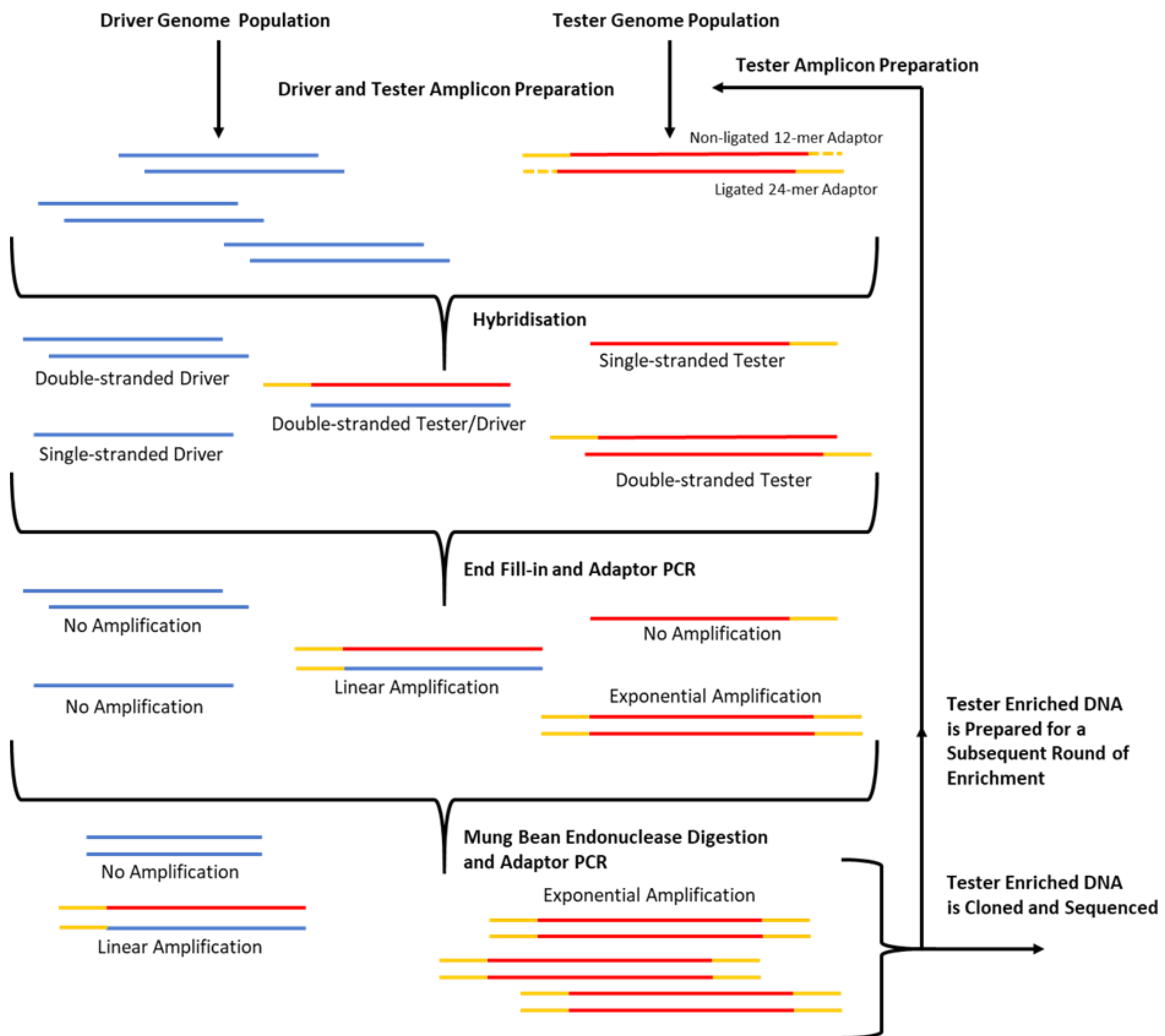

Figure 3-1 A simplified representation of the RDA enrichment process

Figure 3-1 is a simplified representation of the RDA protocol, highlighting what is theories to occur to each of the DNA molecules created during the hybridisation, PCR enrichment and mung-bean digestion phases. Showing the resulting DNA population after RDA enrichment and how it can be either recycled through the hybridisation process again or clones and sequenced for identification and subsequent investigation.

In the next step of the reaction, a 24-mer adaptor corresponding to the adaptor ligated to the tester amplicon was added to the hybridisation reaction together with Taq DNA polymerase. The Taq DNA polymerase extends any double stranded DNA with incomplete 3' ends. A PCR reaction using the same tester 24-mer adaptor as a primer is performed. This PCR reaction is followed by digestion with mung bean endonuclease, removing single stranded DNA, leaving only blunt ended double stranded DNA. A second PCR is then performed using the tester 24-mer adaptor as the primer. This results in the exponential amplification of the double stranded DNA comprising the two complimentary strands of tester amplicon only (theoretically unique to the tester). Double stranded DNA formed by complimentary sequences of the driver and tester will also be amplified in the PCR reactions, but because only one strand of the DNA molecule contains the priming sequence, it is amplified linearly. Driver DNA containing no adaptor will not be amplified.

Following each round of hybridisation and PCR, the tester enriched DNA can either be cloned and the sequences identified using DNA sequencing or subjected to a further round of hybridisation and PCR amplification enrichment, further enriching the unique tester DNA sequences. After each round of hybridisation, the adaptors on the enriched tester DNA sequences are removed and new adaptors of different sequence ligated. This ensures that the selective subtractive kinetic enrichment of the unique tester DNA occurred from fragments containing the newly ligated adaptors and not from any un-cleaved adaptors carried over from the previous rounds of hybridisation.

DNA sequence information pertaining to *M. spilota* or snakes in general, is limited and therefore molecular genetics techniques that require prior DNA sequence knowledge are impractical for this type of investigation. The RDA technique does not require any previous DNA sequence knowledge for either the driver or tester DNA populations. This experimental chapter describes the use of this technique to investigate the presence of sex-linked DNA regions in *M. spilota*.

Materials and Methods

Blood sampling was conducted on the 19th of January 2011 following approval by Curtin University's Animal Ethics Committee; approved project number AEC_2011_25, approved on the 16th of January 2011. Dr Karen Payne conducted all blood sampling. All animals were closely monitored during and after sampling procedures, to identify any infections or other complications. No animals showed any adverse effects during or after sampling.

The animals sampled were kindly made available to us by Mieke and Glen Gaikhorst of Armadale Reptile & Wildlife Centre, 304-308 South Western Highway, Wungong WA, 6112. We endeavoured to take blood samples from 24 snakes. Blood samples were obtained from 23 *M. spilota imbricata* individuals; 15-male and 8-female. Approximately 0.5 ml of blood was taken from the ventral coccygeal vein using a 22-gauge needle and a 1 ml syringe. The blood was transferred to a 1.8 ml CryoTube and approximately 120 µl of 10 mg/ml EDTA solution was added (to a final concentration 1.5 - 2.0 mg/ml EDTA) to prevent coagulation and improve DNA stability during storage. Samples were stored on ice and then at 4°C and -20°C until required.

The reptile park staff identified the sex of each snake using observations of mating and egg laying activity. In addition, the sex of each animal was confirmed at the time of blood sampling, through probing the snake's cloaca to determine the presence or absence of a hemipenis (Dellinger and Hegel 1990). Sex determination by probing, was performed by veterinarian Dr Karen Payne with assistance by Kristy Gaikhorst.

Genomic DNA Preparation

DNA was extracted from 8 female blood samples (F1 to F8) and 8 male samples (M1 to M4, M6 to M9) using a phenol/chloroform extraction method. From each blood sample 5 µl of whole blood was pipetted into a 1.5 ml microfuge tube containing 741 µl of Lysis buffer and 3.75 µl of 20 mg/ml proteinase K (New England Biolabs) making a final volume of 750 µl. Samples were very gently mixed by slow inversion of the tubes. The tubes were incubated at 50°C overnight with gentle inversion approximately three times during the digestion process.

When the digestion was visibly complete (no visible red clumps), the microfuge tubes were removed from the heating block and allowed to cool to room temperature. One volume (750 µl) of Tris buffered Phenol (Ajax) was added to each tube. The microfuge tubes were gently inverted for 5 minutes and centrifuged at 3000 g at 4°C for 30 minutes. The aqueous layer was removed and transferred into a fresh microfuge tube. The phenol extraction was repeated once more. The samples were then extracted with one volume of Chloroform (Ajax), three times, using the same process. The aqueous layer from the final extraction was transferred to a fresh 1.5 ml microfuge tube. The DNA was precipitated by the addition of 750 µl ice cold Isopropanol (Ajax) and the tubes mixed by inversion and left at -20°C overnight. The precipitated DNA was pelleted by centrifugation at 3000 g for 30 minutes at 4°C. The isopropanol was carefully decanted, and each microfuge tube was place upside down on a clean laboratory bench, with the lid open, to allow the DNA pellet to

slowly air dry. The DNA pellet was then resuspended in 500 µl hpH_2O. The quality and quantity of the extracted DNA was evaluated using both agarose gel electrophoresis, 0.7% (w/v) agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer and spectrophotometrically using a NanoDrop spectrophotometer.

Two DNA pools were created: representing male and female snakes. The two pools contained 0.625 µg of DNA extracted from 8 females or 8 males respectively. Each pool contained a total of 5 µg of genomic DNA in a final volume of 240 µl with hpH_2O. Triplicates of each DNA pool were made. Two RDA experiments were performed, investigating the unique DNA sequences in both male and female genomes. The first experiment used female DNA to create the tester population and male DNA to create the driver population. The second RDA experiment was performed in the reverse, using male DNA to create the tester population and female DNA to create the driver population.

The RDA procedure was repeated three times using the different Restriction Endonucleases (RE; Bgl II, Bam HI and Hind III) and appropriate adaptors, as represented in Table 3-1, to create three different representations of the genome. The REs (Bgl II, Bam HI and Hind III), were used to digest the tester and driver pool respectively, using the following reaction conditions; 1 µg of genomic DNA, 20 U of RE (Promega) and 2 µl 10X reaction buffer (Promega), in a total volume of 20 µl. Each reaction was incubated at 37°C for 2 hours. The digestion was evaluated using 1% (w/v) agarose gel electrophoresis in TAE buffer followed by staining with 0.5 µg/ml ethidium bromide in TAE, to visually ensure that the genomic DNA was completely digested. Once digestion had been completed, the reaction was stopped by the addition of 20 µl of chloroform.

One volume of chloroform was then added to each 1.5 ml microfuge tube. The microfuge tubes were gently inverted and then centrifuged at 3000 g for 30 minutes. The top aqueous layer was removed and transferred into a fresh 1.5 ml microfuge tube. A small proportion of the top aqueous layer was left to ensure that no interphase or bottom layer was transferred. Sodium acetate (0.1 volumes, 3M) was added to each sample for a final concentration of 0.3 M sodium acetate and the tubes were mixed gently. The tubes were then centrifuged at 3000 g at 4°C for 30 minutes and the supernatant removed from the precipitant by transferring to a fresh 1.5 ml microfuge tube.

The DNA was precipitated by adding 2.5 volumes of ice-cold ethanol. The tubes were mixed gently and placed on ice for 30 minutes. The microfuge tubes were centrifuged at 3000 g at 4°C for 30 minutes. The supernatant was carefully removed with a 1 ml pipette, making sure not to disturb the

DNA pellet. Thirty µl of room temperature 70% ethanol was added to each tube to wash the DNA pellet. The tubes were again centrifuged at 3000 g at 4°C for 30 minutes, the supernatant removed and discarded and the DNA dried, by inverting the tubes and leaving them open upside-down on the bench for approximately 15 minutes. Each DNA pellet was re-suspended in 15 µl of sterile hpH_2O.

Tester and Driver Amplicon Preparation

Oligonucleotide adaptors used in this experiment, were described in the RDA method of Yuan Chang, 2002 (Yuan 2002). The primers were purchased from GeneWorks and purified using high performance liquid chromatography (HPLC; Table 3-1). Each primer was resuspended in sterile hpH_2O to a final concentration of 100 µM. The R-adaptor sets were used for the construction of representative genomes for both tester and driver. The J-adaptor and N-adaptor sets were used in the RDA hybridisation and enrichment processes, alternating between odd and even rounds of hybridisation.

Table 3-1 RDA adaptors

Adaptor Type	Primer Name	Sequence (5' to 3')
Representation		
24-mers	RBgl-24	AGCACTCTCCAGCCTCTCACCGCA
	RBam-24	AGCACTCTCCAGCCTCTCACCGAG
	RHind-24	AGCACTCTCCAGCCTCTCACCGCA
12-mers	RBgl-12	GATCTGCGGTGA
	RBam-12	GATCCTCGGTGA
	RHind-12	AGCTTGCGGTGA
Odd cycle		
24-mers	JBgl-24	ACCGACGTCGACTATCCATGAACA
	JBam-24	ACCGACGTCGACTATCCATGAACG
	JHind-24	ACCGACGTCGACTATCCATGAACA
12-mers	JBgl-12	GATCTGTTCATG
	JBam-12	GATCCGTTCATG
	JHind-12	AGCTTGTTCATG
Even cycle		
24-mers	NBgl-24	AGGCAACTGTGCTATCCGAGGGAA
	NBam-24	AGGCAACTGTGCTATCCGAGGGAG
	NHind-24	AGGCAGCTGTGGTATCGAGGGAGA
12-mers	NBgl-12	GATCTTCCCTCG
	NBam-12	GATCCTCCCTCG
	NHind-12	AGCTTCTCCCTC

Underlined base pairs indicate restriction endonuclease cleaving sites

The R-adaptor pairs were ligated with the DNA pools, in a 200 µl PCR tube using the following conditions; 500 ng of digested DNA, 2 µl of 10X reaction buffer (Promega), 1 µl of each of the 12-mer and 24-mer adaptors was made up to a final volume of 19.5 µl. The tube containing the ligation reaction was heated to 55°C for 10 minutes before being slowly reduced to 16°C at a constant rate, over a period of 1 hour. Once the reaction had reached 16°C, 0.5 µl of T4 DNA Ligase (Promega) was added to the ligation reaction and the reaction left overnight at 16°C.

The tester and driver ligation reactions were amplified by PCR. The PCR contained 10 µl of 5X MyTaq Buffer (Bioline), 0.5 µl of the 100 µM stock of the corresponding 24-mer adaptor used in the ligation (final concentration 1 µM), 5 µl (25ng) of linker ligated DNA, 0.5 µl of MyTaq DNA polymerase (5 Units/µl Bioline, non-hotstart), made to a final volume of 50 µl. The PCR reactions conditions were; 3 min at 72°C, 20 cycles of 95°C for 1 min and 72°C for 3 min and a final extension at 72°C for 10 min. Four replicates of the driver PCR reaction were made to generate enough driver amplicon to perform all hybridisations required in the RDA protocol. After the PCR was complete, the reactions were evaluated using agarose gel electrophoresis, 0.7% (w/v) agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer.

The tester PCR amplification was then digested with the appropriate RE to remove the R-adaptors. This reaction was performed under the following conditions; 40 µl of the amplified tester PCR product was added to a 1.5 ml microfuge tube along with 5 µl of 10X RE buffer (Promega), 30 U of the appropriate RE made to a final volume of 50 µl. The reaction was incubated at 37°C for 2 hours.

Half of the volume of tester amplicon digestion was then size separated by electrophoresis on a 1.5% (w/v) agarose gel in TAE buffer, stained with 0.5 µg/ml ethidium bromide in TAE buffer and compared to a size standard DNA ladder (Axygen 100bp). The tester amplicon was size selected, between 100bp and 2000bp in length, by excising from the agarose gel using a sterile scalpel. The gel slice was placed into a 1.5 ml microfuge tube and the DNA extracted using a QIAquick Gel Extraction Kit (Qiagen), according to the manufacturer's instructions, the product eluted with sterile water and reconstituted to a concentration of approximately 50 ng/µl.

The size selected tester amplicon was then ligated to a new J-set of adaptors. This was performed in a 200 µl microfuge tube at a final volume of 50 µl, containing 5 µl of 10X ligation reaction buffer (Promega), 1.0 µl of each appropriate 12-mer and 24-mer J-adaptor (100 µM). The tube was placed into a thermal cycler, heated to 55°C for 10 minutes and the temperature reduced to 16°C at a

constant rate, over a period of 1 hour. Once the reaction had reached 16°C, 0.5 µl of T4 DNA Ligase (Promega) was added to the ligation reaction mixture and incubated at 16°C overnight. Upon completion the T4 DNA ligase was heat inactivated at 65°C for 20 minutes. The ligation reaction was purified using a QIAquick PCR Purification Kit (Qiagen), according to the manufacturer's instructions and the product reconstituted at a concentration of approximately 10 ng/µl.

The four driver amplicon PCR replicates were combined and purified using a QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions. The final product was resuspended in hpH_2O at a concentration of 50 ng/µl.

Hybridisation Round One

Tester amplicon, approximately 50 ng, was added to a 200 µl microfuge tube together with 500 ng of prepared driver amplicon, 25 µl of 5X MyTaq Buffer (Bioline), and hpH_2O added to a final reaction volume of 49 µl. The tube was placed into a thermal cycler, heated to 98°C for 30 minutes, and the temperature reduced from 98°C to 72°C over a period of 30 minutes. After the tube had reached 72°C, 0.5 µl of MyTaq DNA polymerase (5 Units/µl Bioline, non-hotstart) and 0.5 µl of the 100 µM of the corresponding 24-mer J-set adaptor was added to make a final reaction volume of 50 µl. The hybridisation PCR reaction was then cycled through a PCR profile of; 72°C for 3 min, 10 cycles of 95°C for 30 s and 72°C for 2 min, followed by a final extension step at 72°C for 10 min.

All single stranded DNA in the PCR reaction was removed by digestion with mung bean nuclease (MBN), using the following conditions; 34 µl of the PCR product from the hybridisation/PCR, 2 µl (20 U) MBN (Promega), 4 µl 10X nuclease buffer (Promega), in a total volume of 40 µl. The reaction was incubated at 37°C for 30 minutes. Once completed, the reaction was diluted 1 in 5 with 50 mM Tris.HCl (pH 8.9) to a final volume of 200 µl and heat denatured for 5 minutes at 95°C.

Two identical PCR reactions were then prepared from the digestion to increase the quantity of targeted DNA. This was achieved using the following conditions; 40 µl of MBN treated DNA, 80 µl of 5X MyTaq Buffer (Bioline), 4 µl at 100 µM of the corresponding 24-mer J adapter and 273 µl of hpH_2O to make a final volume of 397 µl. The PCR reaction was heated to 72°C for 5 min and 3 µl (15 U) of Taq DNA polymerase (Bioline, non-hotstart) added. The reaction was incubated at 72°C for a further 5 min before amplified using the following parameters; 20 cycles of 95°C for 1 min and 72°C for 3 min. Final extension at 72°C for 10 minutes. The final PCR product was combined and purified using a QIAquick PCR Purification Kit (Qiagen), according to the manufacturer's instructions and

eluted using 50 µl of hpH_2O. The eluted product was diluted to a concentration of approximately 0.1 µg/µl. These PCR products were evaluated by agarose gel electrophoresis, 1.0% (w/v) agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer and compared to the previous tester preparation to identify changes in the enrichment.

Hybridisation Round Two to Six

For each of the subsequent rounds of hybridisation the protocol remained relatively consistent, with just three features differentiating between each round of hybridisation:

The type of adaptor ligated to the tester amplicon. Odd hybridisation cycles; 1, 3 and 5 had the J-adaptor ligated and even hybridisation cycles; 2, 4 and 6 the N-adaptor was ligated, as described in Table 3-1.

The amount of prepared tester amplicon added to the hybridisation reaction was reduced 10-fold with every subsequent round of hybridisation. However, the amount of prepared driver remained constant. This altered the overall driver to tester ratio in each hybridisation reaction, as described in Table 3-2. In the later hybridisation reactions, this required the prepared tester amplicon to be diluted to achieve the appropriate tester concentration. The number of cycles used for the PCR conducted after the hybridisation reaction and the secondary PCR conducted after the mung bean nuclease digestion was increased for subsequent hybridisation rounds according to Table 3-2.

Table 3-2 Reaction conditions for hybridisation rounds

Hybridisation	Driver DNA (ng)	Tester DNA (ng)	Driver/Tester Ratio	Primary PCR	Secondary PCR
Round 1	500	50	10:1	10 cycles	20 cycles
Round 2	500	5	100:1	10 cycles	20 cycles
Round 3	500	0.5	1,000:1	10 cycles	30 cycles
Round 4	500	0.05	10,000:1	20 cycles	30 cycles
Round 5	500	0.005	100,000:1	30 cycles	30 cycles
Round 6	500	0.0005	1,000,000:1	30 cycles	30 cycles

RDA Enriched DNA Cloning

The sixth round of hybridisation was conducted according to the same protocol as all other hybridisations. The final PCR products were purified using a QIAquick PCR Purification Kit (Qiagen), following the manufacturer's instructions, eluted using hpH_2O and diluted to a concentration of approximately 0.1 µg/µl. After visual confirmation by agarose gel electrophoresis, 1.0% (w/v)

agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer, the purified PCR products were cloned into pGEM T Easy.

The PCR product generated were ligated into Promega's pGEM-T Easy Vector. This ligation reaction was performed according to the following conditions; 25 ng of purified DNA from hybridisation round six was combined with 50 ng of vector DNA, in a 0.7 ml microfuge tube, 1 µl of T4 DNA Ligase (Promega), 5 µl of 2X rapid ligation buffer added and made up to a final volume of 10 µl. The reaction was prepared on ice and incubated at room temperature for 1 hour.

The ligated vector-insert DNA was transformed into *E. coli* JM109. Five-microliters of the ligation reaction was aliquoted into a 1.5 ml microfuge tube and left on ice. The tube containing the JM109 *E. coli* cells was thawed on ice for approximately 5 minutes before 50 µl of the cells were transferred to the ligation aliquot tube. The treated cells were then heat shocked in a water bath at 42°C for 30 seconds before being immediately returned to the ice for 2 minutes. Pre-warmed SOC medium, 900 µl (37°C) was added to the transformation reaction. The transformed *E. coli* cells in SOC were incubated at 37°C for 1 hour with shaking. Then 100 µl of the transformed cells were spread, using a sterile glass spreader, onto each of ten LBA plates containing Amp, X-Gal and IPTG. The plates were incubated inverted at 37°C overnight.

Pure white colonies (no blue or blue centred colonies) were selected from the LBA plates. Half of each colony was picked and re-streaked onto a fresh LBA plate containing Amp, X-Gal and IPTG. These plates were incubated at 37°C overnight and stored as a retention plate at 4°C. The other half of the colony was tested using PCR to determine the size of the DNA insert. A fresh 200 µl pipette tips was used to pick each of the remaining colonies and each colony was mixed into 500 µl of hpH_2O by pipetting up and down. A PCR was conducted on each colony using the following amplification conditions; 1 µl of 10X PCR Buffer containing dNTPs and $MgCl_2$ (Roche), 0.1 µl (0.5 Units) of Taq DNA polymerase (Roche), 0.1 µl of pUC/M13 forward (5'-GTTTTCCCAGTCACGAC-3') and pUC/M13 reverse (5'-CGCCAGGGTTTTCCCAGTCACGAC-3') primers (final concentration 1 µM) and 1.0 µl of the colony suspension as the template DNA. The reaction was made to a final volume of 10 µl with hpH_2O. The PCR reactions were placed into a thermal cycler and cycled as follows; 95°C for 10 min, 30 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 2 min, final extension at 72 °C for 10 min. The PCR products were visualised on a 1% (w/v) agarose gel electrophoresis in TAE and stained with 0.5 µg/ml ethidium bromide in TAE.

Clones that contained the same DNA insert could be identified according to their matching PCR insert size. Three clones of each DNA insert size were subsequently sequenced. The retention plates created earlier were used to sample the identified clones and samples picked from each clone were incubated in 5 ml of LB broth containing 100 µg/ml ampicillin, within a 15 ml tube. The inoculants were incubated at 37°C for 8 hours with continual shaking. The plasmids contained within each *E. coli* culture were extracted using a Qiagen Plasmid Mini Kit according to the manufacturer's instructions.

A PCR was performed to confirm the presence of a DNA insert before the plasmids were sequenced. The PCR was conducted under the following conditions; 1 µl of 10X PCR Buffer containing dNTP's and $MgCl_2$ (Roche), 0.1 µl (0.5 Units) of Taq DNA polymerase (Roche), 0.1 µl of pUC/M13 forward and pUC/M13 reverse primers (final concentration 1 µM) and 0.5 µl plasmid template DNA (approximately 30 ng), made to a final reaction volume of 10 µl with hpH_2O. The PCR reactions were cycled using the following conditions; 95°C for 10 min, 30 cycles at 95°C for 30 s, 55°C for 30 s, 72°C for 2 min, final extension at 72 °C for 10 min. The PCR products were visualised using by 1% (w/v) agarose gel electrophoresis in TAE buffer, stained with 0.5 µg/ml ethidium bromide in TAE buffer.

Plasmid extractions that produced a positive PCR result were DNA sequenced by Macrogen Korea; 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511, Rep. of Korea. Sequencing used the following primers; pUC/M13 forward (5'-GTTTTCCCAGTCACGAC-3') and pUC/M13 reverse (5'-CGCCAGGGTTTTCCCAGTCACGAC-3'). Geneious (Biomatters Limited) software was used to visualise and interrogate the DNA sequences. The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST), RepeatMasker, Genscan and Mfold Web Server were all used to investigate features within each DNA sequence.

Genomic DNA Investigation

Using the DNA sequencing results, oligonucleotide PCR primers were designed specifically to amplify the DNA regions that were identified. Primers were designed using the National Center for Biotechnology Information online program Primer-BLAST; https://www.ncbi.nlm.nih.gov/tools/primer-Blast/. Each primer pair was chosen to reflect the greatest coverage of the specific DNA target sequence (Table 3-3).

Table 3-3 Male and female RDA insert primers

Female Oligo Name	Oligo Sequence 5' to 3'	Length (bp)	GC %	Tm (°C)
RDA_BglII_1F	TCCACACTTACAAGCCCTGG	20	55.0	52.5
RDA_BglII_1R	GGCTTGCAAAGCACCAAGAC	20	55.0	54.8
RDA_BglII_2F	CGCCCACAGCCATAAAAACC	20	55.0	56.7
RDA_BglII_2R	GAGGATTTGTCGTTCCCACG	20	55.0	54.2
RDA_BglII_3F	GTTTCTGCCTGCTGAAAGGA	20	50.0	52.0
RDA_BglII_3R	TCCTTACCCTTCCTGGGTGAT	21	52.4	53.9
RDA_BamHI_1F	AAACGGGGGTGCGCTAATAA	20	50.0	56.0
RDA_BamHI_1R	TTCCTTGCCCAGACGAAGAC	20	55.0	54.1
RDA_HindIII_1F	GGGGCAGTCTCACTCCAAAA	20	55.0	53.9
RDA_HindIII_1R	AAAGGTGTCGCCATGTCAGT	20	50.0	51.8
Hind12F-F	GGTTCAGCCGTAAACAAGGT	20	50.0	50.9
Hind12F-R	GCTGTGGTATCGAGGGAGAAG	21	57.1	52.9
RDA_HindIII_2F	AGCTTCGTTCCCCAACAGAG	20	55.0	53.3
RDA_HindIII_2R	TGGGCATGTTGCTCTGATGT	20	50.0	53.6

Male Oligo Name	Oligo Sequence 5' to 3'	Length (bp)	GC %	Tm (°C)
BglM1-F	ACCTTGCTGCTTGTCATCCT	20	50.0	51.2
BglM1-R	TCCGCTTTGCTGTTCTTCTCT	21	47.6	53.4
BglM2-F	GGCTACCAGAAGAGGGGGATA	21	57.1	53.8
BglM2-R	TGAGAGGATGGGAGAAAGCG	20	55.0	54.2
BglM3-F	TCCCTGCTCACTTCCACTTTG	21	52.4	54.2
BglM3-R	CCAAGCACAGGGTTAAAAGGC	21	52.4	55.3
BglM4-F	TCTTCATCGCTCTTCCACCAA	21	47.6	54.7
BglM4-R	CACAGTCAGCCACAAGCATC	20	55.0	51.2
BglM5-F	ATATATGGTGGCACACCCTGC	21	52.4	53.7
BglM5-R	CCTCCAGCAATCATCACCCA	20	55.0	55.5
BamM1-F	AGCCAGCCCTCTCTCTCTTTA	21	52.4	51.8
BamM1-R	AAGCTGATACACCCCCGAAC	20	55.0	53.0
BamM2-F	TTCCTGACCCACCAGAGAAAG	21	52.4	53.2
BamM2-R	CTCAAGCTTGTGTTTGGCTGC	21	52.4	55.0
BamM3-F	ATGCCAGCTGACAAAGGCAA	20	50.0	55.6
BamM3-R	ATGTGGCGGGGTCTAGTAGT	20	55.0	50.2
BamM4-F	CTGGTGTGGCTGAGTTAGCA	20	55.0	51.1
BamM4-R	CACTAGGTGCATGAGCCCAA	20	55.0	53.5
HindM1-F	GAACAAGCTTCGCAATGGGC	20	55.0	57.0
HindM1-R	TGTTCGCAGTCACCGTCTAC	20	55.0	50.9
HindM2-F	GTGCAGGAGTGATGTCCGAT	20	55.0	52.0
HindM2-R	AAGCTTTTTGCAAGGTTGCCT	21	42.9	54.9
HindM3-F	GCCTGAGCAAGCCTGGTAAA	20	55.0	54.5
HindM3-R	GCAGCATCTCCCTCGTTTCT	20	55.0	53.4
HindM4-F	AGCTACATCCCATTTGGGCG	20	55.0	56.3
HindM4-R	CCAGAAAAGGAAGCCCCGTT	20	55.0	56.9
HindM5-F	ACCTTGACTCCACATGGACAG	21	52.4	51.2
HindM5-R	GTTCGCCACCACCTCTTTCA	20	55.0	54.9

The PCR primers designed to target the RDA enriched DNA clones, were optimised using temperature gradient PCR, with annealing temperatures ranging from approximately 5°C above and 5°C below the calculated primer Tm values (Table 3-3). This PCR optimisation process was repeated using DNA from one male and one female snake. This was to ensure that no sex-associated PCR products were missed in the process of PCR optimisation. Seven PCR reaction were performed for each DNA template, evenly spread across the 10°C differential in annealing temperature. The negative control (template DNA absent) was included at the lowest annealing temperature. Each PCR was performed in a total volume of 10 µl containing 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10× PCR buffer, 0.1 µM of each primer and 30 ng of template genomic DNA. The thermal profile was 95°C for 10 min, followed by 45 cycles of: 95°C for 30 s, annealing temperature gradient for 30 s, 72°C for 90 s, final extension at 72°C for 10 min. The PCR products were examined using 1% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer. Each PCR primer set was optimised using both male and female DNA templates, side by side. Using the same PCR reagents at the same volume and concentration, the PCR annealing temperatures, the number of PCR cycles and cycle times were modified to produce a single PCR product for each of the primer sets (Table 3-4).

Table 3-4 Male and female RDA insert optimised PCR conditions

Female PCR Primer Pairs	Tm (°C)	Denaturation (s)	Annealing (s)	Extension (s)	PCR Cycles
RDA_BglII_1F & RDA_BglII_1R	51	30	30	60	45
RDA_BglII_2F & RDA_BglII_2R	51	30	30	60	45
RDA_BglII_3F & RDA_BglII_3R	51	30	30	60	45
RDA_BamHI_1F & RDA_BamHI_1R	51	30	30	60	45
RDA_HindIII_1F & RDA_HindIII_1R	No PCR product could be produced				
Hind12F-F & Hind12F-R	51	30	30	60	35
RDA_HindIII_2F & RDA_HindIII_2R	51	30	30	60	50

Male PCR Primer Pairs	Tm (°C)	Denaturation (s)	Annealing (s)	Extension (s)	PCR Cycles
BglM1-F & BglM1-R	56	30	60	90	50
BglM2-F & BglM2-R	56	30	60	90	50
BglM3-F & BglM3-R	56	30	60	90	50
BglM4-F & BglM4-R	52	30	60	90	50
BglM5-F & BglM5-R	70	30	60	90	45
BamM1-F & BamM1-R	54	30	60	90	50
BamM2-F & BamM2-R	56	30	60	90	50
BamM3-F & BamM3-R	56	30	60	90	50
BamM4-F & BamM4-R	50	30	60	90	50
HindM1-F & HindM1-R	60	30	60	90	45
HindM2-F & HindM2-R	60	30	60	90	45
HindM3-F & HindM3-R	60	30	60	90	45
HindM4-F & HindM4-R	60	30	60	90	45
HindM5-F & HindM5-R	60	30	60	90	45

DNA sequencing was performed on the optimised PCR reactions, using template DNA from three male and three females. Five identical PCR reactions were combined, visualised using agarose gel electrophoresis and sequenced (Macrogen Korea). Geneious (Biomatters Limited) was used to analyse the DNA sequences. RepeatMasker, nBLAST, Genscan and Mfold Web Server were all used to identify features within each DNA sequence. Similarity between DNA sequences was determined using (Burmese python) unplaced genomic scaffold database made available for download from NCBI Python_molurus_bivittatus-5.0.2 reference Annotation Release 101, published in 2013 (Castoe et al. 2013).

Results

Initially the protocol used for our RDA experiment was an unaltered version of Yuan Chang RDA protocol, published in Current Protocols in Molecular Biology, 2002 (Yuan 2002). The study initially targeted female-specific DNA only, using all three restriction endonucleases detailed in the protocol;

Bgl II, Bam HI and Hind III. Five rounds of hybridisation were performed, without any modifications to the method. The results produced from using this protocol appeared satisfactory leading up and including the third round of hybridisation. However, after the third round of amplification, it was difficult to clearly visualise the enriched DNA fragments within the agarose gel (Figure 3-2).

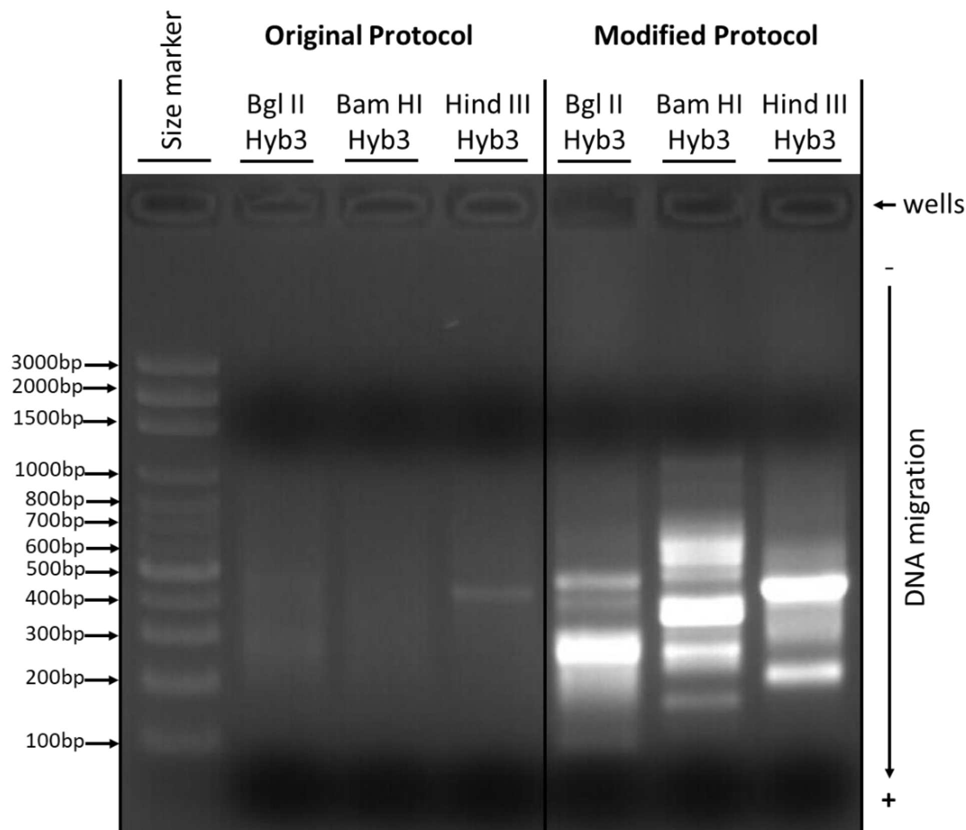

Figure 3-2 Comparison of RDA protocol after hybridisation round 3

Figure 3-2 shows a comparison agarose electrophoresis gel of the original and modified RDA protocols after hybridisation round 3. Each protocol is represented by the Bgl II, Bam HI and Hind III RDA hybridisations as labelled. It is important to note that both protocols exhibit a similar banding pattern produced for each RE used. With the modified protocol presenting a greater DNA band intensity than the original.

Results from the original RDA hybridisation protocol, after rounds one, two and three and the corresponding PCR amplifications, showed that the concentration and complexity of the enriched DNA populations appeared to be reduced. After hybridisation round one, the representational DNA population appeared to be a smear ranging in size from 100bp to 2000bp, and with subsequent hybridisation rounds, DNA bands were observable within this smear. However, after the fourth round of hybridisation, the overall DNA concentration had fallen so low that no DNA, smear, or

enriched DNA bands, could be visualised within the agarose gel. Finally, after the fifth and sixth round of hybridisation, there was no observable increase in the DNA concentration between the post-hybridisation sample and post-PCR sample.

It was suspected that the PCR may have been less efficient than was suggested in the published protocol. To address this issue, the number of PCR cycles after the RDA hybridisation of round 3 and the subsequent mung bean digestion were increased. Ten cycles to 20 cycles post hybridisation and from 20 cycles to 30 cycles post mung bean digestion were evaluated. However, this adjustment to the protocol produced a result that is difficult to explain. The change in protocol produced an increase in overall DNA concentration, however the size of the smear shifted considerably from a size range of 100bp to 2000bp, to a range of 800bp to approximately 10kb (Figure 3-3). Similar results were also observed in subsequent rounds of RDA hybridisation. This led to a complete re-evaluation of the original published RDA method.

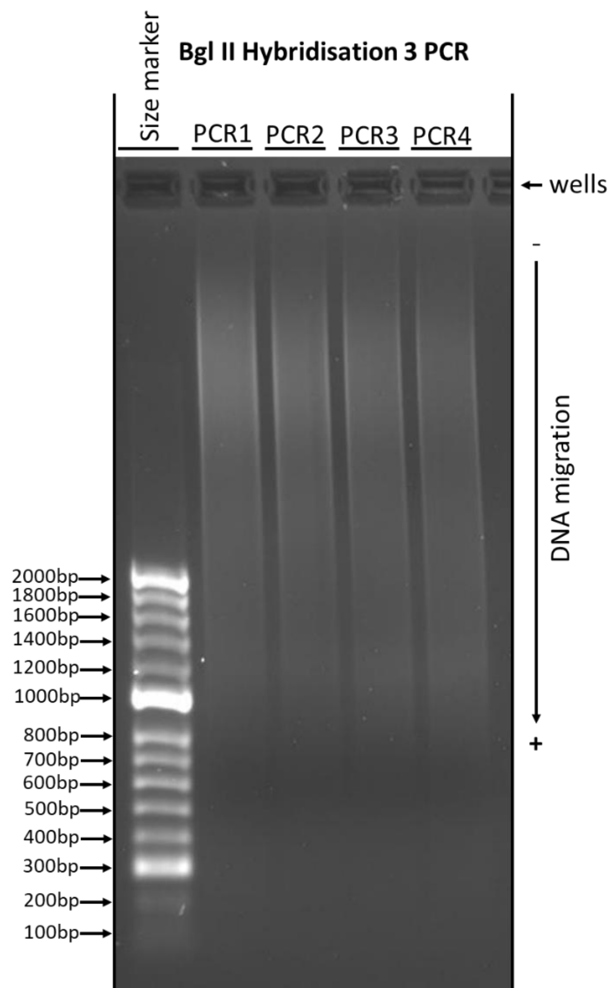

Figure 3-3 RDA hybridisation round 3 with increased PCR cycles

Figure 3-3 shows the result produced after increasing the PCR cycles post RDA hybridisation round 3 using Yuan Chang's original RDA protocol. The agarose electrophoresis gel shows an increase in overall DNA concentration, but also a significant increase in the size range of the smear. Indicating unwanted DNA interactions within the PCR.

Multiple modifications to Yuan Chang's original RDA protocol were performed, through a process of trial and evaluation. With unmodified and modified protocols running side by side for comparison and the results visually compared using agarose gel electrophoresis. Modifications tested and included in the final protocol were;

1. Reducing the amount of DNA being used in the experiment.
2. Reducing the ratio of tester to driver DNA in the hybridisation reactions.
3. Combining the hybridisation and subsequent PCR reaction into a single reaction.
4. Increasing the number of PCR cycles performed after the hybridisation reaction and after the mung bean nuclease digestion reaction.
5. Increasing the number of rounds of hybridisation performed.

These modifications to the originally described RDA protocol resulted in a consistently enriched tester-specific DNA fragment population. Figure 3-2 shows that the modifications increased the specificity of the DNA being enriched, which appeared to increase after each round of hybridisation. The outcomes of the modified RDA protocol were observed after each round of hybridisation using agarose gel electrophoresis as shown in Figure 3-4.

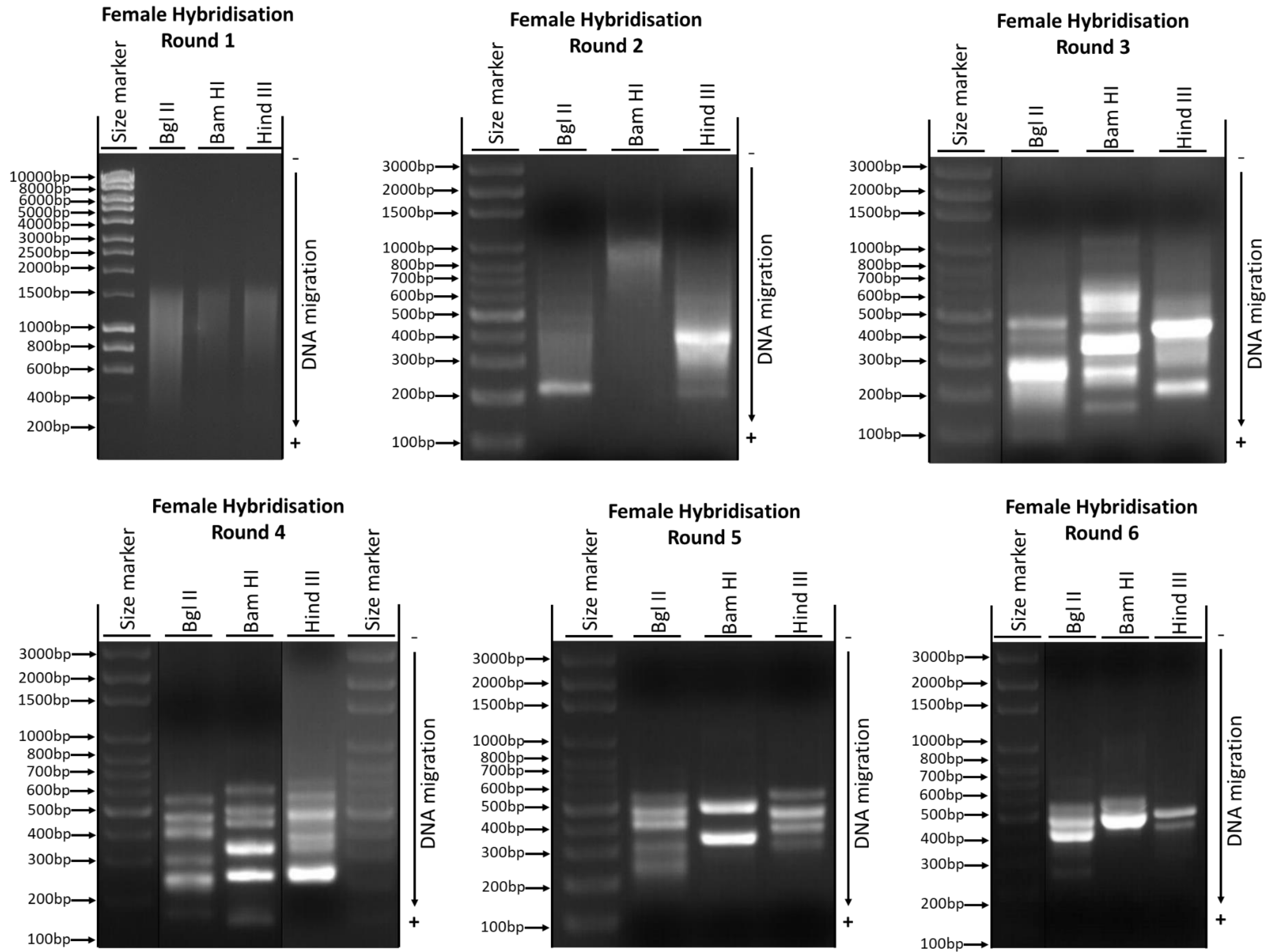

Figure 3-4 Modified RDA protocol hybridisation rounds 1 to 6

Figure 3-4 is a collection of 6 different agarose electrophoresis gels, representing the observed DNA products present after each round of RDA hybridisation and post-PCR amplification, targeting female-specific DNA fragments. Each agarose gel contains a DNA marker; hybridisation round one contains a 1kb DNA marker (Axygen) and hybridisation round 2 to 6 contain a 100bp DNA marker (Axygen). The samples in each agarose gel are arranged in the same order, with the DNA digested with restriction endonucleases Bgl II, Bam HI and Hind III, left to right. Agarose gels representing hybridisation round 3, 4 and 6 have been altered by deleting images of some unrelated DNA samples between the hybridisation samples and the DNA marker, to avoid possible confusion.

The modified RDA protocol revealed that subsequent rounds of hybridisation resulted in the overall complexity of the DNA population appearing to be reduced. Each genomic DNA gel profile started with a smear of multiple fragments ranging in size from 100bp to 2000bp. After each round of hybridisation, the complexity of enriched DNA appears to decrease, however, the overall DNA concentration remained relatively constant. Individual DNA bands progressively became visible after hybridisation round two, though they were not always completely distinguishable from a background DNA smear produced by the wider DNA population until post hybridisation round 5. After 6 rounds of hybridisation, the diverse DNA population had been reduced to several distinct and quite intense DNA fragments. These remaining DNA bands were considered as tester-specific enriched DNA. All genomic representations produced a different pattern or subset of enriched DNA fragments. Once we were confident in the results produced from our modified RDA protocol, a repeat experiment enriching male-specific DNA was performed. Similar results were observed when testing for male-specific DNA fragments, but a different banding pattern was observed for each RE experiment.

The RDA enriched DNA fragments from each of the three genomic representation analyses, for both male and female, were cloned into the vector pGEM and *E. coli* transformed. The size of the insert in twenty-two clones obtained from the different genomic representations was determined using M13-primer PCR and the fragments visualised by agarose gel electrophoresis. This allowed the diversity of captured DNA inserts from each representation to be determined. At least one sample from each different DNA insert size was then sent for direct plasmid DNA sequencing (Macrogen, Korea). The sequencing results were analysed using Geneious. All sequences were also analysed for repetitive DNA using the online

program RepeatMasker. The origin of each DNA sequence was investigated using the NCBI's BLAST and the top result for each sequence reported. The results produced from each of these DNA fragments are shown in Table 3-5.

Table 3-5 RDA enriched DNA regions RepeatMasker and BLAST results

Female Enriched DNA	Length (bp)	RepeatMasker Results	BLAST Results	BLAST - Sequence Identity
Bgl II Insert 1	318		PREDICTED: *Chrysemys picta DPP9* mRNA	28/30 (93%)
Bgl II Insert 2	459	DNA/PIF - Harbinger (139-172) LINE/CRI (226-281)	PREDICTED: *Lonchura striata TGFBR3* mRNA	38/41 (93%)
Bgl II Insert 3	481	LINE/L2 (80-224)	PREDICTED: *Chenopodium quinoa ITPR2* mRNA	34/40 (85%)
Bam HI Insert 1	425		PREDICTED: *P. bivittatus FSTL1* mRNA	185/198 (93%) 63/65 (97%)
Hind III Insert 1	380	LTR/Gypsy (39-371)	*Echis coloratus* Ty3/gypsy retrotransposon partial sequence	184/258 (71% (w/v))
Hind III Insert 2	349		*Syphacia muris* genome assembly Scaffold 0000252	25/25 (100%)

Male Enriched DNA	Length (bp)	RepeatMasker Results	BLAST Results	BLAST - Sequence Identity
Bgl II Insert 1	337	LINE/RTE - BovB (125-203) Microsatellite (GAAAA)6 (267-298)	PREDICTED: *P. bivittatus ANKS6* mRNA	128/165 (78%) 113/166 (68%)
Bgl II Insert 2	664	LTR/ERV1 (2-298)	PREDICTED: *Fulmarus glacialis* LOC104079068 mRNA	50/62 (81% (w/v))
Bgl II Insert 3	358		PREDICTED: *P. bivittatus OSBPL6* mRNA	140/141 (99%)
Bgl II Insert 4	304		*Psammodromus hispanicus* locus 17 genomic sequence	69/92 (75%)
Bgl II Insert 5	335		*Trimeresurus gramineus gTgTBP* gene complete cds	137/176 (78%)
Bam HI Insert 1	430		*Ovis canadensis* chromosome 12 sequence	43/51 (84%)
Bam HI Insert 2	508		*Tetrahymena thermophila CLASP* mRNA	30/33 (91% (w/v))
Bam HI Insert 3	180	LINE/Penelope (1-157)	PREDICTED: *P. bivittatus RABGAP1L* mRNA	100/107 (93%)
Bam HI Insert 4	261		*Bacillus pumilus* strain SH-B9 complete genome	33/38 (87%)
Hind III Insert 1	312		PREDICTED: *P. bivittatus DNAH2* mRNA	155/167 (93%)
Hind III Insert 2	326		*Apteryx australis* genome assembly scaffold 87	57/71 (80%)
Hind III Insert 3	294		*Hydrogenophaga sp.* RAC07 complete genome	31/33 (94%)
Hind III Insert 4	429		PREDICTED: *P. bivittatus BCL6B* mRNA	167/177 (94%)
Hind III Insert 5	358	LINE/L2 (73-356)	PREDICTED: *P. bivittatus MTG2* mRNA	61/79 (77%)

Inserts highlighted in blue are comprised partially or entirely of repetitive elements.

All sequence lengths exclude RDA oligonucleotide adaptors.

Sequences for each DNA insert, in both forward and reverse direction, were determined. Except for the DNA from Female Hind III - Insert 1, The sequence obtained from Female Hind III - Insert 1 is discussed further in the following paragraphs. The 5' and 3' RDA adaptors were identified in each DNA insert sequence. All had the correct adaptor DNA sequence with no mismatches. The DNA sequence length corresponded to the observed length seen on the agarose gel electrophoresis image produced from the final hybridisation PCR, round 6 (Table 3-5). The length of the DNA fragments produced ranged from 180bp to 664bp, with the average fragment length being 375bp.

Using RepeatMasker, DNA inserts that contained repetitive DNA elements were identified and these RDA fragments are highlighted (in blue) in Table 3-5. It was observed that 7 of the 20 DNA fragments contained repetitive DNA elements. These repetitive DNA elements comprised between 15% and 87% of the total length of the obtained DNA sequence. None of the repetitive elements identified were repeated in different DNA inserts.

NCBI BLAST was used to determine the identity of the sequenced DNA fragments (Table 3-5). Every DNA fragment returned a BLAST match. Generally, DNA fragments that returned a BLAST result using the mega BLAST algorithm were matched to DNA sequences from other snake or reptile species, with an identity percentage of between 70% to 90% across approximately 100bp to 200bp. If the DNA insert was not matched using the mega BLAST algorithm, the BLASTn algorithm was used to determine somewhat similar DNA sequence matches. DNA matches from these searches were generally above 80% identity, however these DNA sequence matches were considerably shorter, comprising only 25bp to 50bp in length, which are most likely to be random in nature. These shorter sequence matches returned using the BLASTn algorithm, identified much more distantly related organisms and in some cases species of bacteria. No BLAST result inferred any relatedness between any DNA fragments identified. The BLAST results indicated that some of the DNA sequences had associations with sex determination, sexual differentiation, sexual development, or fertility.

The DNA sequences were further analysed using the genome of the Python_molurus_bivittatus-5.0.2 reference Annotation Release 101 (Table 3-6). All DNA sequences showed a strong mega BLAST match. These BLAST matches had high percentage sequence identity with the lowest of 88% and the highest showing 97% identity across the entire query sequence length. Three of the 20 DNA sequences returned BLAST matches to

multiple NCBI *P. bivittatus* Annotation Release 101 scaffolds. These three sequences are composed of repetitive elements and therefore this result was not surprising. Four aligned to exonic regions, eight aligned to intronic regions, three aligned to non-coding regions near or between genes and five aligned to unidentified non-coding genome regions. One of the sequences; Male Bgl II - Insert 1, showed high sequence similarity to the ends of two different scaffolds and appeared to bridge the gap between the missing DNA sequence from the NCBI *P. bivittatus* Annotation Release 101.

Table 3-6 RDA enriched DNA regions NCBI P. bivittatus annotation release 101 BLAST results

Female Enriched DNA	**Length (bp)**	***P. bivittatus* Unplaced Genomic Scaffold**	**BLAST – Sequence Identity**	***P. bivittatus* Scaffold Results**
Bgl II Insert 1	318	Scaffold 489	299/318 (94%)	Noncoding region between *ADAMTS17* and *CERS3* genes
Bgl II Insert 2	459	Scaffold 35	429/460 (93%)	Intronic region of *RALGDS* gene
Bgl II Insert 3	481	Scaffold 659	438/465 (94%)	Intronic region of *SORBS1* gene
Bam HI Insert 1	425	Scaffold 380	404/425 (95%)	Exon 1 and 2 of *FSTL1* gene
Hind III Insert 1	380	Scaffold 12399, 9816, 12447, 4330, 6280, 6353, 11869, 8882, 24689, 6987, 12568 and 29316	284/320 (89%)	Noncoding region found in 12 different scaffolds
Hind III Insert 2	349	Scaffold 823	323/350 (92%)	Noncoding region near *AKAP6* gene
Male Enriched DNA	**Length (bp)**	***P. bivittatus* Unplaced Genomic Scaffolds**	**BLAST – Sequence Identity**	***P. bivittatus* Scaffold Results**
Bgl II Insert 1	337	Scaffold 5223 Scaffold 568	182/191 (95%) 147/154 (95%)	Noncoding region found in 2 different scaffolds
Bgl II Insert 2	664	Scaffold 11227, 8744, 23692, 11305, 9694, 10863 and 12505	343/373 (92%) 1-373bp 235/261(90%) 404-664bp	Noncoding region found in 7 different scaffolds
Bgl II Insert 3	358	Scaffold 2169	347/358 (97%)	Exon 17 of *OSBPL6* gene
Bgl II Insert 4	304	Scaffold 211	290/302 (96%)	Intronic region of *CCDC141* gene
Bgl II Insert 5	335	Scaffold 3997	256/284 (90%)	Unidentified noncoding region
Bam HI Insert 1	430	Scaffold 5959	392/444 (88%)	Intronic region of *LRCH4* gene
Bam HI Insert 2	508	Scaffold 431	485/504 (96%)	Intronic region of *MBTD1* gene
Bam HI Insert 3	180	Scaffold 7749, 2463, 1598, 5270, 8021, 3917, 8621, 498, 5609 and 2597	148/157 (94%)	Noncoding region found in 9 different scaffolds
Bam HI Insert 4	261	Scaffold 50	247/259 (95%)	Noncoding region adjacent *JPH2* gene
Hind III Insert 1	312	Scaffold 749	291/312 (93%)	Exon 22 and 23 of *DNAH2* gene
Hind III Insert 2	326	Scaffold 1046	317/327 (97%)	Intronic region of *SLAIN1* gene
Hind III Insert 3	294	Scaffold 3405	275/291 (95%)	Intronic region of *MED12L* gene
Hind III Insert 4	429	Scaffold 2188	405/429 (94%)	Exon 3 of *BCL6B* gene
Hind III Insert 5	358	Scaffold 227	325/358 (91% (w/v))	Intronic region of *LOC103063639* unidentified gene

Male and Female Enriched DNA Fragments

To investigate the sex specific nature of the DNA fragments identified from the RDA experiments, sets of PCR primers were designed for all 20 DNA regions. These primer pairs were optimised, using both male and female DNA templates. This was done to ensure any sex bias produced within the PCR optimisation process was not unintentionally disregarding.

The final optimised PCR conditions are represented in Table 3-4. At the beginning of the optimisation process, the PCR primers for regions that were made up partially or entirely of repetitive elements, produced a variety of different sized PCR products. These DNA fragments were more difficult to optimise than their non-repetitive counterparts, although all were eventually optimised to produce only one DNA fragment. Except for PCR primer pair RDA_HindIII_1F & RDA_HindIII_1R. There was no observable difference in size, the presents or absents of DNA bands or a difference in DNA band intensity, which might indicate a sex-linked DNA marker, from the male and female DNA amplicons using agarose gel electrophoresis.

The potential genetic differences between male and female derived PCR products were analysed through amplifying and DNA sequencing. PCRs targeting the enriched DNA regions from Female Bgl II Insert 1, Female Bgl II Insert 2, Female Bgl II Insert 3, Female Bam HI Insert 1, and Female Hind III Insert 1 were DNA sequenced. For comparison, this was repeated using genomic DNA from 3 males and 3 females. Each PCR reaction was reproduced 5 times, combined, column purified, checked by agarose gel electrophoresis, and then submitted for sequencing (Macrogen Korea). A summary of the results obtained is shown in Table 3-7. Analysis of the sequences between individuals revealed nucleotide differences between the genomic derived PCR products and the original RDA derived DNA insert sequences.

Table 3-7 Genomic PCR products sequenced

Female Enriched DNA	Length (bp)	PCR Primer Pairs	PCR Product Length (bp)	Individual Polymorphisms	Sequencing Differences Between Cloned and PCR DNA
Bgl II Insert 1	318	RDA_BglII_1F & RDA_BglII_1R	177	SNP C/T position 121 SNP A/T position 133 Indel A/T position 94 Sex-linked SNP C or T-Males, A-Females	Substitution C/T position 165
Bgl II Insert 2	459	RDA_BglII_2F & RDA_BglII_2R	294	None	Substitution A/G position 205 Substitution G/A position 234 Deletion T position 309
Bgl II Insert 3	481	RDA_BglII_3F & RDA_BglII_3R	348	None	Substitution C/T position 336 Substitution C/T position 401
Bam HI Insert 1	423	RDA_BamHI_1F & RDA_BamHI_1R	217	None	Deletion T position 143 Substitution G/A position 159 Substitution T/C position 235 Substitution T/C position 268 Substitution G/A position 283 Substitution T/C position 293 Substitution CTT/TCC position 302-304 Substitution TATA/CGTC position 311-314
Hind III Insert 1	380	Hind12F-F & Hind12F-R	228	G/A position 114 G/A position 129 C/T position 153 T/C position 196 T/A position 254 G/A position 289 A/G position 293 T/C position 319	None

Female Hind III - Insert 1

The DNA insert Female Hind III - Insert 1, could only be DNA sequenced in one direction. All attempts to DNA sequence the insert within the pGEM vector using M13 primers failed in the forward direction but gave good sequence when the reverse primer was used. Furthermore, direct DNA sequencing of the PCR products was only successful in the reverse direction. It was observed that DNA sequencing reaction performed using either the forward M13 primers or the forward PCR primer (RDA_BglII_1F), gave no clean sequencing signal characterised by a low-quality score of <3%.

The sequencing results obtained in the reverse direction showed that after approximately 280bp of good sequence, the intensity of the sequencing signal declined to approximately one quarter of its original strength. From this point, the sequence signal continued to decrease and after a further 100bp the sequencing software was unable to differentiate individual nucleotides from background. Therefore, only 380bp of the DNA fragment was

sequenced in both the forward and reverse orientation, which is considerably smaller than the estimated size of the cloned fragment. However, the longest sequencing result obtained in the reverse orientation was 901bp in length, before the sequencing reaction reached its maximum length limit. After the removal of the RDA adaptor and the pGEM vector sequence the sequence insert totalled was 776bp. The RDA adaptor in the reverse orientation was easily identified, however after 776bp no forward RDA adaptor or pGEM vector could be clearly identified in this sequence. This indicates that the length of this DNA insert is considerably larger than 776bp, even though after several sequencing attempts this result could not be improved.

DNA sequencing comparison using PCR products produced from male and female samples identified 8 single nucleotide polymorphisms (SNPs; Table 3-7). From these sequencing results it was determined that none of the 8 SNPs identified appeared to be sex-linked. Repeat masking results for the reliable 380bp sequence and the larger 776bp sequence showed that 87% and 66% of the two DNA sequences respectively, aligned with retroviral elements belonging to the family LTR/Gypsy. BLAST analysis supported this result with two BLAST results showing similarities with DNA sequence having an identity score of 70% from 475-767bp and 75% from 329-427bp, against the DNA sequence Ty3/gypsy retrotransposon reverse transcriptase and RNase H pseudogene belonging to *Vipera palaestinae* (Palestine viper). While the remaining 320bp showed BLAST matches with 16S ribosomal RNA sequences from a variety of different types of bacterium commonly found in soil, the highest being 75% across 187bp belonging to *Curtobacterium flaccumfaciens*.

Further investigation of the snake sequences coding Ty3/gypsy retrotransposons revealed examples of similar sequences belonging to 4 more snakes species; *Echis coloratus* (Saw-scaled viper), *Natrix tessellata* (Dice snake), *Boa constrictor* (Red-tailed boa) and *Python molurus* (Indian python) (Gorinsek, Gubensek, and Kordis 2004). These related DNA sequences were aligned, and the identity scores are shown in Table 3-8. It was observed that all five DNA sequences aligned to the DNA insert between 329bp and 776bp. The Ty3/gypsy retrotransposons showing greatest similarity belonged to the *B. constrictor* with an identity score of 58.8%. However, examples of Ty3/gypsy retrotransposons identified in other species also showed high identity scores. The most closely related snake species in this group, *P. molurus,* had the lowest alignment identity of 54.6%.

Table 3-8 Sequence similarity between Female Hind III - Insert 1 and TY3/Gypsy sequences

Species	Accession Number	Length (bp)	Identity (%)
N. tessellata	AY158713	907	56.6
E. coloratus	AY158729	936	57.3
V. palaestinae	AY158744	1044	57.0
B. constrictor	AY158731	924	57.9
B. constrictor	AY158732	925	55.2
B. constrictor	AY158733	943	58.8
P. molurus	AY158744	896	54.6

Female Hind III - Insert 1 was analysed against NCBI *P. bivittatus* Annotation Release 101. Fourteen BLAST hits across 12 different genomic scaffolds were observed, all showing a strong sequence identity of between 82% to 89% with the DNA insert sequences. The BLASTn algorithm, used to identify similar DNA sequences, produced 216 BLAST hits across 100 different genomic scaffolds from the NCBI *P. bivittatus* Annotation Release 101, all with identity matches greater than 73%. However, all the BLAST hits identified aligned with the DNA insert across two thirds of its sequence (329bp through to 776bp), with no examples of DNA matches to the prior third of the sequence being observed.

Male Bgl II - Insert 2

The DNA insert Male Bgl II - Insert 2 was observed to have a DNA sequence of 644bp in length, and exhibited strong sequence similarity with the repetitive region ERV1-10_AMi-I. This repetitive element belongs to the repeat family LTR/ERV1 and was identified at the beginning of the sequence, ending 298bp downstream. This DNA repeat is a type of endogenous retroviral element, like Ty3/gypsy and also usually has long terminal repeats flanking the beginning and end of the retroviral element (Belshaw et al. 2004). This DNA sequence showed strong sequence similarity to 7 separate DNA scaffolds from NCBI *P. bivittatus* Annotation Release 101, but all 7 scaffolds lacked annotation data and therefore it could not be determined what these DNA regions represent.

Analysis of the DNA insert Male Bgl II - Insert 2 using the mega BLAST algorithm identified no similar DNA sequences. Using BLASTn, similar DNA sequences previously associated with sex determination were identified. The BLAST results, presented in Table 3-9, shows sequence similarity to the human gene Protein Kinase, Y-Linked, Pseudogene (PRKY) located on the human Y chromosome and the *Oryctolagus cuniculus* (European Rabbit) *SRY* gene, as well as regions from both chicken chromosome Z and W. A pairwise alignment of the Male Bgl II –

Insert 2 and the rabbit complete protein coding sequence representing the *SRY* gene showed that the two sequences had a sequence identity of 52.2%, with the much shorter region identified in the BLAST alignment having an identity of 82.2% over 45 nucleotides (12). Further investigation showed that the region did not align with any functional part of the *SRY* gene in *O. cuniculus*, but it was aligning to a retroviral element present near the *SRY* gene (12).

Table 3-9 BLAST results Male Bgl II - Insert 2

BLAST Sequences Description	Score (bits)	Identities	Gaps	Matching DNA region
PREDICTED: *Fulmarus glacialis* uncharacterized LOC104079068, partial mRNA	59.0 (64)	50/62(81%)	0/62(0%)	1-62
PREDICTED: *Tauraco erythrolophus* endogenous retrovirus group K member 25 Pol protein-like (LOC104381738), partial mRNA	50.0 (54)	42/52(81%)	0/52(0%)	1-52
***Pan troglodytes* BAC clone CH251-31E16 from Y, complete sequence**	48.2 (52)	40/49(82%)	0/49(0%)	101-149
***Homo sapiens PRKY* gene, partial sequence, in clone Y813, complete sequence**	48.2 (52)	40/49(82%)	0/49(0%)	101-149
***Oryctolagus cuniculus* clone BAC 828D7 *SRY* (*SRY*) genes, complete cds**	46.4 (50)	37/45(82%)	0/45(0%)	181-225
***Gallus gallus* BAC clone CH261-140N15 from chromosome w, complete sequence**	44.6 (48)	36/44(82%)	0/44(0%)	2-45
***Gallus gallus* BAC clone CH261-64K10 from chromosome z, complete sequence**	44.6 (48)	34/39(87%)	1/39(2%)	514-552

RDA Fragments DNA Secondary Structures

The structure and function of some of the identified retrotransposons revealed that these repetitive elements possibly regulate their own replication using Primer Binding Site (PBS) and Polypurine Tract (PPT), shown to bind to themselves creating double stranded DNA secondary structures from a single strand of DNA (Finnegan 2012). An investigation of these features in the DNA inserts was performed using the Mfold Wed Server (1995-2017) hosted by The RNA Institute, College of Arts and Science, State University of New York at Albany, USA. Created by Michael Zuker & Nick Markham, © Rensselaer Polytechnic Institute and supported by the SUNY Albany Research IT Group (http://unafold.rna.albany.edu/?q=mfold/dna-folding-form) (Zuker 2003; Santa 1998; Peyret 2000). In this analysis the parameters of our RDA hybridization reaction were replicated within the server's system; Na^{+} = 50 mM, Mg^{+2} = 1.5 mM at 72°C and the DNA molecule was represented as a polymer in the system; Table 3-10.

Table 3-10 DNA secondary structure results

Female Enriched DNA	Delta G Value (kcal/mol at 72°C)	Predicted Melting Temperature (°C)
Bgl II Insert 1	-0.30	74.8
Bgl II Insert 2	0.10	70.8
Bgl II Insert 3	-0.40	77.4
Bam HI Insert 1	-0.76	75.4
Hind III Insert 1	-6.47	75.8
Hind III Insert 2	0.30	67.6

Male Enriched DNA	Delta G Value (kcal/mol at 72°C)	Predicted Melting Temperature (°C)
Bgl II Insert 1	-0.07	72.7
Bgl II Insert 2	-0.53	77.0
Bgl II Insert 3	0.95	66.0
Bgl II Insert 4	-0.44	74.7
Bgl II Insert 5	0.06	71.3
Bam HI Insert 1	-0.42	75.1
Bam HI Insert 2	0.96	62.8
Bam HI Insert 3	1.14	59.4
Bam HI Insert 4	0.29	65.6
Hind III Insert 1	-0.02	72.1
Hind III Insert 2	0.83	64.5
Hind III Insert 3	-0.56	75.1
Hind III Insert 4	-0.51	74.1
Hind III Insert 5	0.73	64.5

Delta G represents the minimum free energy required for folding of the modelled DNA secondary structure to occur (Zuker 2003; Santa 1998; Peyret 2000). If the delta G value is greater than zero it is very unlikely that any secondary DNA structure will be formed. While, the more negative the delta G value, the more likely the modelled DNA secondary structure will occur (Zuker 2003; Santa 1998; Peyret 2000). The Tm value is the predicted melting temperature of the modelled DNA secondary structure (Zuker 2003; Santa 1998; Peyret 2000).

Mfold analysis indicated that the DNA region, Female Hind III – Insert 1, had the lowest delta G value of any other DNA insert of -6.47 kcal/mol. In comparison, all other DNA inserts had a delta G value of between -0.76 and 1.14. The predicted DNA secondary structure of the DNA region Female Hind III – Insert 1 is shown in Figure 3-5. It shows that the first 321bp of the DNA region has minimal secondary structure. However, the second half of the DNA region from 321bp to 671bp, most likely has a strongly bound DNA secondary structure. Analysis of

this region determined that this DNA secondary structure would have a melting temperature of 75.8°C, within our modelled RDA hybridisation protocol.

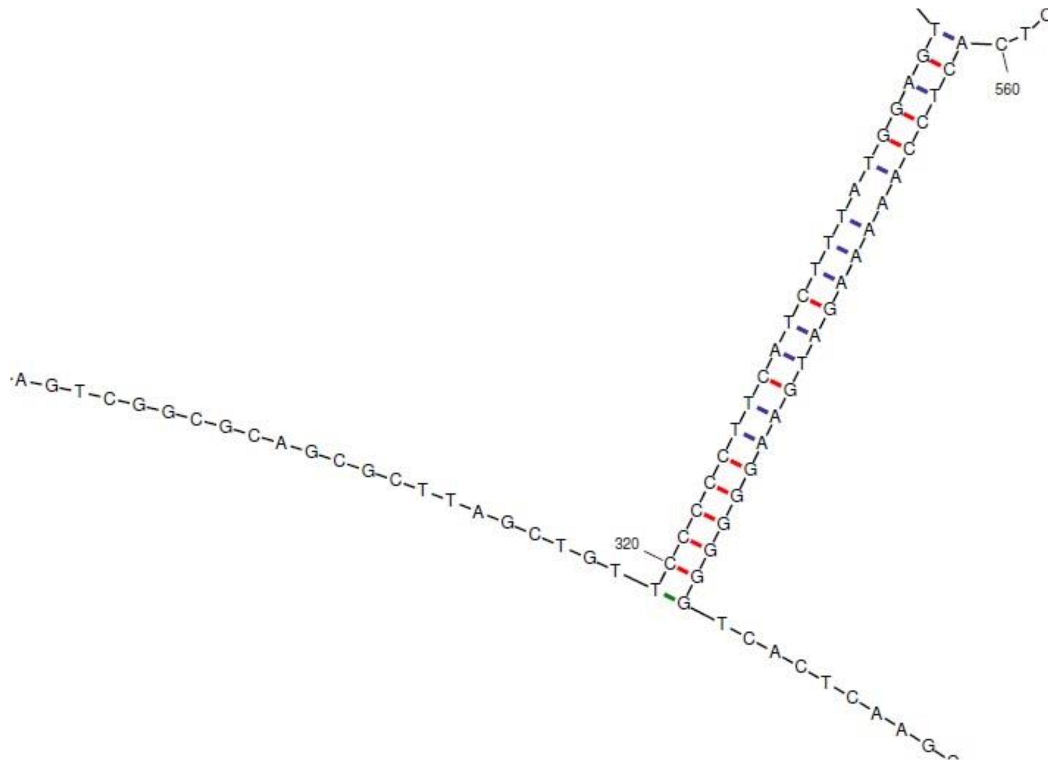

Figure 3-5 DNA secondary structure Female Hind III - Insert 1

Figure 3-5 is a representation of the predicted DNA secondary structure formed by the DNA region Female Hind III – Insert 1, produced by Mfold Wed Server (1995-2017) hosted by The RNA Institute (Zuker 2003; Santa 1998; Peyret 2000)*. The delta G value and melting temperature of the linear DNA fold model was determined with a temperature parameter of 72°C and an ionic condition of Na^+ = 50 mM, Mg^{+2} = 1.5 mM. Replicating the experimental conditions of the modified RDA hybridization reaction as closely as possible. Standard errors are roughly ±5%, ±10%, ±11% (w/v) and 2-4 °C for free energy, enthalpy, entropy, and Tm, respectively.*

Discussion

The Representation Difference Analysis (RDA) protocol initially used in this experiment was based upon the publication; Representational Difference Analysis, by Yuan Chang, 2002, which was in turn based upon the original RDA protocol publication; Cloning the differences

between two complex genomes, by Lisitsyn N, Lisitsyn N and Wigler M, 1993 (Lisitsyn, Lisitsyn, and Wigler 1993; Yuan 2002). Our early experiments used Yuan Chang's RDA protocol but failed to produce satisfactory results. This resulted in the development of a new RDA protocol, which is simpler, took less time to perform and used considerably fewer reagents.

The most substantial alteration made to the RDA protocol was combining the hybridisation reaction and then subsequent PCR reaction into a single reaction. The original protocol specified hybridising the tester and driver amplicons over a period of 24 hours, with multiple temperature reduction and incubation steps, down to a final temperature of 4°C. However, the subsequent PCR reaction, using an aliquot of the hybridisation mixture, calls for the sample to be re-heated back to 72°C, as quickly as possible and then continues onto subsequent rounds of PCR amplification. It was considered unnecessary to slowly reduce the temperature of the hybridisation reaction to 4°C over a period of 24 hours, only to subsequently increase the reaction temperature back to 72°C, as quickly as possible, in the next step of the protocol. It was observed that by combining the two reactions together the reaction time was reduced by 23 hours for each round of hybridisation and resulted in a PCR product that was more concentrated than the original protocol.

Combining the hybridisation and PCR reaction meant that the hybridisation reaction no longer occurred in a solution of 1M NaCl, but instead took place in a reaction made up of primarily 1X MyTaq Buffer, which contains 50 µM dNTPs, 1.5 mM $MgCl_2$, 50 mM KCl, PCR stabilizers and PCR enhancers. This solution has an overall reduction in the concentration of positively charged ions in the hybridisation reaction. Theoretically, this reduction of positively charged ions would have resulted in the overall negative charge from the DNA strands being shielded less and therefore, double stranded DNA would have had a higher melting temperature in our hybridisation protocol, compared to the original protocol, possibly resulting in a more specific hybridisation result (Tsuruoka and Karube 2003; Fuchs et al. 2010). This reduction in the positively charged ion concentration was a necessity for the hybridisation and the PCR reactions to be combined.

The total DNA concentration (both tester and driver DNA) was reduced in the hybridisation reaction and the subsequent PCR, used in this RDA protocol. The original protocol used 0.5 µg of tester DNA amplicon and 40 µg of driver DNA amplicon in a hybridisation reaction

volume of 5 µl (8.1 µg/µl). A reduction to 0.05 µg of tester DNA amplicon and 0.5 µg of driver DNA amplicon in a total volume of 49 µl (0.011 µg/µl) was produced in this modified protocol. The results obtained showed that the DNA concentration used in the original protocol was too high and it was observed to be negatively impacting the subsequent PCR reaction. The reduction in the overall DNA concentration was a found to be required when combining both hybridisation and PCR reactions, however it also ensured that the subsequent PCR reaction was not inhibited by high template DNA concentrations. Furthermore, this reduction in overall DNA concentration made our RDA protocol more efficient than the original, requiring less starting material and less reagents for an improved result.

The first round of hybridisation in the original protocol used a ratio of 1 to 80 (tester to driver) and this ratio was reduced to 1 to 10 in the modified protocol. This ratio was reduced to increase the amount of DNA that was being selectively enriched during the hybridisation reaction. It was observed when performing the original protocol that using a ratio of 1 to 80 resulted in no DNA being recovered after the fourth or fifth round of hybridisation. After six rounds of hybridisation the ratio of tester to driver DNA being used in the hybridisation reaction was similar between both protocols.

The impact of combining the hybridisation and PCR reactions in our protocol, compared to the original RDA protocol is difficult to theoretically determine. However, the protocol changes that were made in this study were performed sequentially and the results of these changes at each step were critically evaluated. The changes to the protocol found to either produce no difference in the results or appeared to improve the results were subsequently incorporated, as presented in Figure 3-2, which shows a comparison of the two protocols. All protocol changes that resulted in an undesirable effect were discontinued in the final protocol.

The phenol-chloroform DNA purification and ethanol precipitation described in the original protocol was replaced with the use of a QIAquick PCR Purification Kit (Qiagen). For all examples using this substitution, the goal was to purify the PCR fragments from PCR enzymes and primers. While the phenol-chloroform extraction and ethanol precipitation would have completed this task, the method uses harmful chemicals, is labour intensive and results in reduction in quantity of the final product. The QIAquick PCR Purification Kit does

not require the use of harmful chemicals, can be completed in approximated 30 minutes, and has a theoretical 95% recovery rate. This increased the amount of DNA being retained from one step to another in the protocol, allowing further reductions in the amount of DNA being manipulated from the previous protocol step. Furthermore, the QIAquick PCR Purification Kit has been designed specifically to purify PCR products, removing enzymatic protein, salt, dNTPs and primers, while also protecting overhanging single-stranded DNA ends.

Representational Difference Analysis

The results presented, have shown that the modified RDA protocol enriched DNA fragments that were present in the tester populations. Each round of hybridisation showed a sequential reduction in the complexity of the tester DNA population, which was initially represented as a smeared DNA population and was observed visually being reduced in size and concentration after each round of hybridisation. The enrichment of specific DNA fragments is evident, represented as a successive increase in the DNA concentration of specific DNA fragment, which persisted through multiple rounds of the hybridisation enrichment.

PCR reaction protocols were developed targeting each of the 20 DNA regions identified from the two RDA experiments. However, no DNA regions identified were proven to be sex specific as was the original aim of this experiment. With hindsight, it would have been more prudent to use an animal system with a chromosomally defined sex specific difference as a control in our modified RDA protocol. For example, *Notechis scutatus* (Tiger snake), which has been shown to have cytologically distinguishable sex chromosomes and has previously described sex specific DNA regions. The addition of this snake species in the experiment would have provided validation of the protocol and would have provided a direct comparison with the results produced from *M. spilota*. Despite this, the RDA experiments, yielded 20 genomic DNA regions from *M. spilota*, which still have the potential to be sex-linked. An analysis of these DNA regions was performed which categorised them into two groups. The first category is DNA regions that are comprised partially or entirely of repetitive elements and the second is DNA regions that are comprised of either coding and/or noncoding DNA regions of the genome.

Group 1 – Repetitive Elements

Seven of the 20 DNA regions identified in our female (three) and male (four) RDA experiments were shown to contain repetitive elements. These repetitive elements comprised between 19% and 87% of the DNA sequences and were identified using the online database RepeatMasker. Long interspersed nuclear elements (LINE) belonging to the families CRI, L2, RTE (BovB) and Penelope were identified, as well as an autonomous transposon element from the family PIF/Harbinger, a retroelements with long terminal repeats from the Ty3/Gypsy family as well as an endogenous retroviral sequence 1 with long terminal repeats from the ERV1 family (Gorinsek, Gubensek, and Kordis 2004; Finnegan 2012; St. John and Quinn 2008; Lovsin, Gubensek, and Kordi 2001; Adelson, Raison, and Edgar 2009; Evgen'ev et al. 1997; Grzebelus et al. 2007; Llorens et al. 2011; Knisbacher and Levanon 2016). One simple repetitive element in the form of a penta-nucleotide microsatellite sequences was also found in the DNA fragment Male Bgl II – Insert 1 and was associated with an RTE (BovB) LINE. These seven DNA regions making up Group 1, all contain different examples of retrotransposable elements.

Retrotransposons are DNA elements that are thought to be able to amplify themselves within the genome using a 'copy and paste' mechanism (Castoe, Hall, et al. 2011; Finnegan 2012). They perform this process by first transcribing themselves into a complimentary RNA molecule and then using reverse transcription to produce a complementary DNA molecule, which can then integrate itself back into the genome at a different, but target specific site (Castoe, Hall, et al. 2011; Finnegan 2012). This replication process can induce mutations by inserting new copies of the retrotransposon near or within genes within the genome (Castoe, Hall, et al. 2011; Finnegan 2012). Retrotransposons are found ubiquitously in many eukaryotic genomes including snakes, which have been observed to have a surprising amount of variation between species (Castoe, Hall, et al. 2011; Finnegan 2012).

The genome of *Agkistrodon contortrix* (Copperhead) is comprised of approximately 45% transposon elements, whilst the genome of *P. bivittatus* contains approximately 21% of these elements (Castoe, Hall, et al. 2011). Transposon elements in snakes have been showed to be extremely active, with even examples showing horizontal genetic transfer of these elements occurring between species of vertebrates (Castoe, Hall, et al. 2011; Adelson, Raison, and Edgar 2009). Studies on five snake species belonging to the family Viperidae have shown that the retrotransposon elements exhibit a dynamic, lineage-specific,

expansion and that some transposons have brain-specific expression, along with closely associated genes (Yin et al. 2016). It is hypothesised that this rapid expansion of transposons in snakes, was not only associated, but may also have facilitated the adaptive evolution of olfactory, venom and thermal sensing genes and also functional degeneration of genes associated with vision and hearing, as the ancestors of snakes made their way from a terrestrial habitat to a subterranean one (Yin et al. 2016).

In general, PCR primers designed to target the DNA regions from Group 1 were more difficult to optimize than DNA regions from Group 2, producing multiple DNA bands during the optimization process. This result is expected when trying to amplify a retrotransposon element using PCR, as it most likely has many related copies dispersed throughout the genome. However, all PCR reactions were optimised to produce a single DNA band.

Female Hind III – Insert 1

Female Hind III – Insert 1 was one of the most interesting DNA regions identified through the RDA experiments. This DNA region shows sequence similarity and therefore is assumed to belong to the Ty3/Gypsy family of retrotransposon (Gorinsek, Gubensek, and Kordis 2004). This is the first time this type of retrotransposon has been documented in snake species *M. spilota*.

The Ty3/Gypsy are a family of retrotransposons that are comprised of retroviruses bearing long terminal repeats and are widely distributed in plants, fungi and some groups of vertebrates (Gorinsek, Gubensek, and Kordis 2004). These retrotransposons, which are closely related to endogenous retroviruses usually have a genome size of between 4-15kb, which are flanked by long tandem repeats on both sides (Llorens et al. 2011).

The structure of the Ty3/Gypsy begins with a 5’ long terminal repeat approximately 100-2000bp, followed by a non-coding region about 75-250bp, which corresponds to the first portion of the retrotranscribed genome (Llorens et al. 2011). A Primer Binding Site (PBS) of 18bp follows, which is complementary to a specific region of a transcribed RNA molecule normally provided by the host cell to start the retro-transcription (Llorens et al. 2011). Following is an open reading frame, which codes for a group-specific antigen (gag) and a polymerase (pol) gene (Llorens et al. 2011). An env gene, which encodes a protein forming a viral envelope and other accessory genes are sometimes seen as well (Llorens et al. 2011). A

region of approximately 10bp of A or G nucleotides called the Polypurine Tract (PPT) is where the synthesis of the proviral DNA strands begins (Llorens et al. 2011). And finally, the 3' long terminal repeat, which is again 100-2000bp in length (Llorens et al. 2011).

It was not possible to determine the entire length of the Ty3/Gypsy retrotransposons isolated in this study, the largest DNA sequence obtained being 776 base pairs and containing no obvious beginning. The BLAST results showed that Female Hind III – Insert 1 has sequence similarity to Ty3/Gypsy retrotransposons found in other closely related and unrelated snake species belonging to the snake families Pythonidae, Boidae, Colubridae and Viperidae. While the DNA sequence isolated in this study showed clear signs of divergence from these other snake retrotransposons. Investigation into this retrotransposon in the NCBI *P. bivittatus* Annotation Release 101 showed a strong correlation to 12 examples of related DNA sequences, as well as over 200 more distantly related DNA sequences, which may contain related or divergent retrotransposons.

Investigation of potential DNA secondary structures produced from the DNA region Female Hind III – Insert 1 showed that the DNA region likely forms a DNA secondary structure that starts at the exact point our DNA sequencing efforts failed. Our modelling shows that this DNA secondary structure weakens once you pass 671bp. However, it is likely that this structure continues past this point. The Mfold modelling did not determine this feature as the remaining DNA sequence data after 776bp is missing. This secondary structure possibly explains why this fragment was obstinate to DNA sequencing in the reverse direction (past 320bp), but not in the forward direction. Examples are published in the literature revealing how this type of DNA secondary structure can cause Taq DNA polymerase to stall during DNA replication, rendering DNA sequencing past this point problematic (Voineagu et al. 2008). This occurs because the DNA secondary structure inhibits the enzyme's ability to unzip the double stranded DNA molecule and continue synthesising DNA (Voineagu et al. 2008).

We hypothesise that single stranded tester amplicons that can self-anneal to produce DNA secondary structures that are stable above 72°C, will have an increased probability of being enriched during the RDA protocol (Figure 3-6). During the hybridisation process as the temperature of the system is reduced, it is assumed that two disassociated complimentary DNA strand realign and form double stranded DNA. However, if a single stranded DNA

molecule can also bind to itself and thereby cause a reduction in enthalpy of the system, close to the change in enthalpy caused by it binding to its complimentary DNA strand, an equilibrium will be formed with both products being present in the system. Furthermore, the RDA hybridisation reaction also requires adequate proximity of two complimentary DNA strands, before they will become reassociated. A single stranded DNA molecule is always near itself and therefore has a greater probability of forming an intramolecular structure compared to an intermolecular structure. Our evidence indicates that this process probably occurred at least once in the RDA experiments, thereby resulting in the enrichment of Female Hind III – Insert 1.

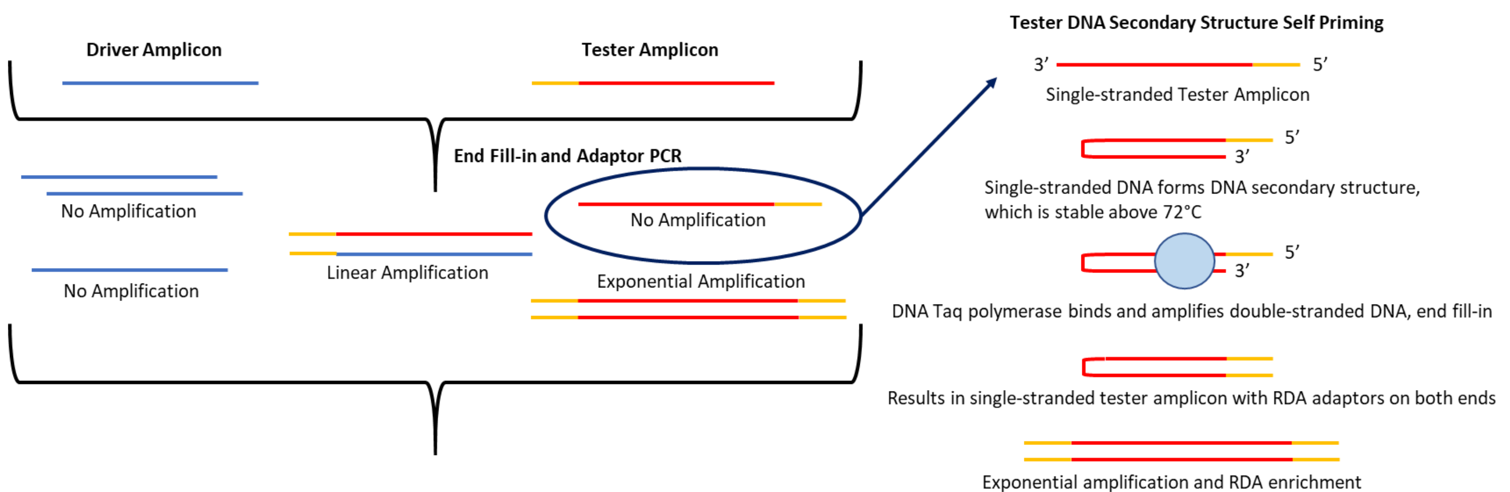

Figure 3-6 Tester DNA secondary structure self-priming and RDA enrichment

Figure 3-6 is a simplified representation of the RDA protocol during hybridisation. Showing the theorised process and resulting exponential amplification of single stranded DNA from the tester population forming a DNA secondary structure.

Male Bgl II – Insert 2

Investigation into the DNA region Male Bgl II – Insert 2 revealed that this DNA region has sequence similarity to multiple regions associated with both sex chromosomes and sex determination in mammals and birds; *Homo sapiens* BAC clone from chromosome Y (AC010104), *G. gallus* BAC clone from chromosome W (AC239887) and *G. gallus* BAC clone from chromosome Z (AC186840) (NCBI Resource Coordinators 2017). This also includes the human gene Protein Kinase, Y-Linked, Pseudogene (*PRKY*) located on the human Y chromosome. This gene is the non-functional version of the X chromosome gene *PRKX* (Schiebel et al. 1997). It is classified in humans as a transcribed pseudogene, but it has lost a coding region which has resulted in nonsense-mediated decay and therefore is unlikely to be

translated into a protein product (Schiebel et al. 1997; NCBI Resource Coordinators 2017). This region no longer undergoes recombination with its X chromosome pair in humans, however, examples of atypical recombination events between these two genes on the X and Y chromosomes has produced examples of human XX males and XY females (Schiebel et al. 1997).

Male Bgl II – Insert 2 also has sequence similarity to a Bacterial Artificial Chromosomes (BAC) clone containing the Sex Determining Region Y (*SRY*) gene of *Oryctolagus cuniculus* (European Rabbit). The *SRY* gene is the primary sex determining gene in the rabbit and is located on the Y chromosome (Geraldes et al. 2010). A pairwise alignment of the BAC clone and Male Bgl II – Insert 2 showed a sequence identity of 52.2%. The smaller DNA region, 45bp in length, identified in the BLAST search, had a sequence identity of 82%. This DNA region represents a repetitive DNA sequences located within the non-coding DNA regions adjacent the *SRY* gene (Geraldes et al. 2010). Furthermore, Male Bgl II – Insert 2 also showed sequence similarity to two independent regions of the Chicken Z and W chromosome, but the function of these regions in the Chicken is unknown.

The 7 DNA regions identified from NCBI *P. bivittatus* Annotation Release 101, which had sequence similarity to Male Bgl II – Insert 2, provided limited information. All 7 DNA regions lacked sequencing data and were unannotated. However, identifying multiple scaffolds from the genome of *P. bivittatus* indicates the DNA region most likely represents a repetitive DNA element, most likely a retrotransposon, with multiple copies throughout the genome. The PCR amplification of a proportion of Male Bgl II – Insert 2 using genomic DNA from both male and female *M. spilota* resulted in no observable difference between the male and female PCR templates. However, no DNA sequence for these PCR products were generated and therefore small nucleotide differences could not be ruled out. A male-linked DNA region associated with a snake Y chromosome is supported by the most recent findings in the literature and considering the DNA region’s strong association with sex determining genes and particularly the Y chromosome of other vertebrate species, this DNA region should be investigated further (Gamble et al. 2017).

Group 2 - Coding and Noncoding Region of the Genome

Inserts containing coding and noncoding DNA regions were found to be the most numerous, totalling 13 out of the 20 DNA regions isolated. This DNA sequence result was proportionally more frequent in the male RDA experiment, accounting for 70% of the DNA regions in the male enrichment experiment, while only comprising about 50% DNA regions enriched in the female experiment. BLAST analysis of Group 2 DNA sequences showed that 4 of the 13 DNA regions had high sequence similarity with exonic regions of four genes. All four genes identified using BLAST are predicted mRNA sequences produced from NCBI *P. bivittatus* Annotation Release 101. These four genes were Follistatin-Like Protein 1 (*FSTL1*), Oxysterol Binding Protein Like 6 (*OSBPL6*), Dynein Axonemal Heavy Chain 2 (*DNAH2*) and B-Cell CLL/Lymphoma 6, Member B (Zinc Finger Protein; *BCL6B*).

Female Bam HI - Insert 1 showed sequences similarity to exon one, intron one, exon two, and the beginning of intron 2 of the *FSTL1* gene in *P. bivittatus*. The *FSTL1* protein's function is linked to tissue development in the lungs, ureter, central nervous system and the skeletal system, as well as heart regulation in humans (Geng et al. 2011, 1; Adams, Larman, and Oxburgh 2007). The regulation of the *FSTL1* gene is also associated with some diseases, with *FSTL1* transcription shown to be downregulation in some cancers and the gene product is thought to be an autoantigen and has been associated with rheumatoid arthritis in humans (Kudo-Saito et al. 2013, 1; Chaly et al. 2012). Studies on *FSTL1* gene in fish have shown that the gene is linked to sexual differentiation (Casas et al. 2016). The gene is expressed with a strong female bias in fish from the Amphiprioninae (clown fish) subfamily and *Oncorhynchus mykiss* (rainbow trout) (Casas et al. 2016). It is thought that the *FSTL1* protein may play a significant role in ovarian function by modifying steroidogenesis in follicles (Casas et al. 2016).

Analysis showed that the Male Bgl II – Insert 3 is representative of intron 16, exon 16 and intron 17 of the *OSBPL6* gene in *P. bivittatus*. The *OSBPL6* protein is a member of the oxysterol-binding protein family, and functions as an intracellular lipid receptor (Ouimet et al. 2016). The gene has very broad expression in humans, but its highest expression is in the brain and skin (Ouimet et al. 2016). To our knowledge no association between *OSBPL6* gene and sex has been shown.

Male Hind III – Insert 1 was found to be homologous with exon 22, intron 22 and exon 23 of the gene *DNAH2* in *P. bivittatus*. The *DNAH2* gene encodes for the dynein heavy chain, a protein component constituent of the ATP dependent motor, Dynein (Chapelin et al. 1997). These types of motors provide the driving force required for cellular movement within cilia and flagella (Chapelin et al. 1997). Low level expression of the *DNAH2* gene has be found in many different types of tissues in humans, but the highest expression levels of the gene are found in the testis and to a lesser extent ovaries, placenta and lungs (Maiti et al. 2000; Hu et al. 2017). The dynein heavy chain makes up a crucial component of sperm flagella, and is responsible for sperm motility (Maiti et al. 2000; Hu et al. 2017). While this gene has not been found to be associated with sexual definition or sex determination, the down regulation of *DNAH2* in the testis has a negative impact on sperm motility and therefore fertility in fish (Hu et al. 2017).

The DNA region Male Hind III – Insert 4 was found to contain a proportion of exon 3 and intron 3 of the *BCL6B* gene in *P. bivittatus*. As its name suggests, the gene encodes protein containing a zinc-finger motif (Fitzgibbon et al. 2000). This type of motif is commonly found in DNA binding proteins and are involved in regulation of DNA transcription (Fitzgibbon et al. 2000). Very little is known about the function of the *BCL6B* gene, although the gene has been linked with Lymphoma in humans (Fitzgibbon et al. 2000).

The remaining nine DNA regions from Group 2 had lower sequence similarity BLAST matches with a variety of different organisms, including some bacterial species, however over considerably shorter lengths. All DNA sequences returned BLAST matches with somewhat similar sequences using the BLASTn algorithm. BLAST results from NCBI *P. bivittatus* Annotation Release 101 showed that five of the eight DNA regions likely contained within intronic regions in five different genes; Coiled-Coil Domain Containing 141 (*CCDC141*), Leucine Rich Repeats and Calponin Homology Domain Containing 4 (*LRCH4*), Mbt Domain Containing 1 (*MBTD1*), SLAIN Motif Family Member 1 (*SLAIN1*) and Mediator Complex Subunit 12 Like (*MED12L*).

Male Bgl II - Insert 4 showed sequence identity to an intronic region of the Coiled-Coil Domain Containing 141 (*CCDC141*) gene in *P. bivittatus*. The protein produced from *CCDC141* has a coiled-coil structural domain (Hutchins et al. 2016). Although the exact function of the *CCDC141* gene is still unknown, proteins with similar coiled-coil domains are

known to be important in many biological functions, often regulating the expression of genes (J. Liu et al. 2006).

A mutation in the *CCDC141* gene has been shown to be associated with Kallmann Syndrome in both humans and mice (Hutchins et al. 2016). Kallmann Syndrome is a genetic disorder that prevents individuals from either starting or completed puberty and is usually combined with symptoms of partial or complete anosmia (Hutchins et al. 2016). The syndrome is caused by a deficiency of gonadotropin-releasing hormone (GnRH), which results in decreased production of testosterone in males and estrogen and progesterone in females (Hutchins et al. 2016). GnRH is expressed in both neurons and olfactory fibers (Hutchins et al. 2016). A mutation in the *CCDC141* gene has been shown to reduce GnRH neuronal migration (Hutchins et al. 2016). The study hypothesised that the *CCDC141* gene is important in embryonic migration of GnRH and mutations in the *CCDC141* gene may result in abnormal formation of hypothalamic neuronal network to initiate pulsatile GnRH secretion, ultimately leading to problems with reproductive function in both humans and mice (Hutchins et al. 2016).

Male Bam HI - Insert 1 displayed sequence homology to an intronic region of the Leucine Rich Repeats and Calponin Homology Domain Containing 4 (*LRCH4*) gene in *P. bivittatus*. This gene is ubiquitously expressed in humans and mice (Aloor et al. 2012). Structural predication of the protein produced from the *LRCH4* gene show that it contains a transmembrane domain and the protein is rich in leucine repeats, which suggests the protein may function as a cellular receptor (Aloor et al. 2012). Other research has found the *LRCH4* protein can be localised to the plasma membrane supporting this hypothesis (Aloor et al. 2012). Knock-down studies have shown that the *LRCH4* protein most likely makes up a sub-unit of a receptor complex and functions as a regulator of the innate immune response (Aloor et al. 2012).

BLAST analysis of Male Bam HI – Insert 2 showed it had sequence identity to the Mbt Domain Containing 1 (*MBTD1*) gene in *P. bivittatus*. The *MBTD1* gene is expressed ubiquitously in human tissue, but is expressed three times as high in the testis than any other tissue (Fagerberg et al. 2014). The *MBTD1* gene is a member of the Polycomb (PcG) proteins family and contains four MBT domains exhibiting an asymmetric rhomboid architecture (Eryilmaz et al. 2009, 1). The genes function is not well understood, but it is known that members of the Polycomb (PcG) proteins family function as subunits of a large protein complex, which is involved in establishing and maintaining of transcriptional

repression of developmental control genes (Eryilmaz et al. 2009, 1). MBT domains are known to bind to mono or dimethylated lysine histones and affect the transcription of genes accompanying the histone protein (Eryilmaz et al. 2009, 1; Luo et al. 2013, 1). In knockout studies, a lack of *MBTD1* has been shown to affect the transcription of genes associated with stabilisation of mouse oocyte meiotic maturation (Luo et al. 2013, 1). Specific mutations within the *MBTD1* gene have also been shown to cause endometrial stromal sarcoma and acute myeloid leukemia (Dewaele et al. 2014; de Rooij et al. 2016).

The Male Hind III – Insert 2 had sequence homology to the SLAIN Motif Family Member 1 (*SLAIN1*) from *P. bivittatus*. This gene is expressed ubiquitously in humans, but its expression is highly biases towards brain and testis tissue (Fagerberg et al. 2014). With expression of *SLAIN1* being approximately 25 times greater in the brain and 6 times greater in the testis than any other tissues (Fagerberg et al. 2014). Despite this association with brain and testicular tissue, *SLAIN1* has not yet been shown to be involved in sex determination or sexual differentiation. The function of this gene is not fully understood, but research suggests the gene, along with its homologue SLAIN2, is involved in differentiation of embryonic stem cells in mice, particularly during the early stages of embryonic development (Hirst et al. 2006).

Male Hind III – Insert 3 contained sequence homology with the Mediator Complex Subunit 12 Like (*MED12L*) gene in *P. bivittatus*. In humans, MED12L protein is a subunit of the macromolecular complex called Mediator (Daniels et al. 2013). This Mediator complex plays a role in the regulation of RNA Polymerase II activity (Daniels et al. 2013). The *MED12L* gene is a paralog of the gene *MED12* produced through a gene duplication even (Daniels et al. 2013). Whilst *MED12* is located on the X chromosome in humans, *MED12L* is located on chromosome 3 (Daniels et al. 2013; NCBI Resource Coordinators 2017). The *MED12L* gene as well as its paralog *MED12* have been shown to have a link with prostate cancer and particularly it's metastatic potential (Taylor et al. 2017; Barbieri et al. 2012). While mutations in the gene influencing *MED12L* regulation have also been shown to impact on the aggressiveness of the disease, as well as the aggressiveness of other diseases (Taylor et al. 2017; Barbieri et al. 2012).

Three of the remaining four DNA regions within Group 2 showed sequence similarity to non-coding DNA regions from NCBI *P. bivittatus* Annotation Release 101. Female Bgl II – Insert 1

showed sequence similarity to a non-coding DNA region between the two genes ADAM Metallopeptidase with Thrombospondin Type 1 Motif 17 (*ADAMTS17*) and Ceramide Synthase 3 (*CERS3*). Female Hind III – Insert 2 showed sequence similarity to a non-coding DNA region adjacent the A-Kinase Anchoring Protein 6 (*AKAP6*) gene. Male Bam HI - Insert 4 showed sequence similarity to a non-coding DNA region adjacent the Junctophilin 2 (*JPH2*) gene. The DNA region Male Bgl II – Insert 5 was unable to be identified in the genome of *P. bivittatus*. The only DNA sequences that Male Bgl II – Insert 5 had sequence similarity to, was the TATA-box-binding protein (*TBP*) from the snake species *T. gramineus*. The *TBP* gene is known to bind to specific DNA sequences called TATA boxes, which are usually located upstream of genes near the gene promotor and activate transcription of the downstream gene (Nakashima et al. 1995).

Twenty genes have now been mapped to the Z chromosome of the *Elaphe quadrivirgata* (Japanese four-striped rat snake) (Matsubara et al. 2012, 2006). These genes and their chromosomal location in humans are represented in Table 3-11
(NCBI Resource Coordinators 2017). The orthologue genes present on the q-arm of chromosome 17 in humans is the most well represented chromosomal location in the snake Z chromosome, with 8 orthologues located between human chromosome 17q11.2 and 17q23.3 (NCBI Resource Coordinators 2017). Orthologues present on the p-arm of chromosome 10, the p-arm of chromosome 3 and the p-arm and q-arm of chromosome 7 in humans are also represented on the snake Z chromosome (NCBI Resource Coordinators 2017).

Table 3-11 Human chromosome location of gene localised to snake Z and W chromosome

Gene abbreviations	Human Chromosome Location
GAD2	10p12.1
WAC	10p12.1
PIP4K2A	10p12.2
PTER	10p13
KLF6	10p15.2
MYO1D	17q11.2
ATP6V0A1	17q21.2
EIF1	17q21.2
TUBG1	17q21.2
RUNDC1	17q21.31
***KAT7* (aka *MYST2*)**	17q21.33
MRPL3	3q22.1
GH1	17q23.3
CTNNB1	3p22.1
RAB5A	3p24.3
CCDC127	5p15.33
AMPH	7p14.1
TAX1BP1	7p15.2
***CCDC126* (aka *LOC90693*)**	7p15.3
SRI	7q21.12

A comparison was performed with the human chromosomal locations of these 20 genes that mapped to the snake Z chromosome and the human chromosomal location of the genes identified through the BLAST analysis using our Group 2 DNA regions (Table 3-12). Two of the genes identified are located on chromosome 3 in humans, one located to chromosome 7, and another three are localized to chromosome 17. The most interesting result was produced for the gene *MBTD1*. An intronic region of the *MBTD1* gene was identified using our RDA protocol; DNA region Male Bam HI – Insert 2. This gene has been localized to the q-arm of chromosome 17 in humans, specifically at 17q21.33 (NCBI Resource Coordinators 2017). While one gene *KAT7* is located at this specific location, 17q21.33, in humans and it's orthologues have been localized to the snake Z chromosome (Matsubara et al. 2012, 2006). The *MBTD1* gene's location in humans is only approximately 1.4Mb from the *KAT7* gene (NCBI Resource Coordinators 2017). This suggests that the *MBTD1* gene may be located on the snake Z chromosome. Intron 13 of the *MBTD1* gene was also identified as being potentially sex-linked in our RDA experiments. While this information is promising, more research needs to be conducted to confirm this finding.

Table 3-12 Human chromosome location of RDA sequence gene homologues

Female Enriched DNA	*P. bivittatus* Scaffold Results	Human Chromosome Location
Bgl II Insert 1	Noncoding region between *ADAMTS17* and *CERS3* genes	15q26.3
Bam HI Insert 1	Exon 1 and 2 of *FSTL1* gene	3q13.33
Hind III Insert 2	Noncoding region near *AKAP6* gene	14q12

Male Enriched DNA	*P. bivittatus* Scaffold Results	Human Chromosome Location
Bgl II Insert 3	Exon 17 of *OSBPL6* gene	2q31.2
Bgl II Insert 4	Intronic region of *CCDC141* gene	2q31.2
Bam HI Insert 1	Intronic region of *LRCH4* gene	7q22.1
Bam HI Insert 2	Intronic region of *MBTD1* gene	17q21.33
Bam HI Insert 4	Noncoding region adjacent *JPH2* gene	20q13.12
Hind III Insert 1	Exon 22 and 23 of *DNAH2* gene	17p13.1
Hind III Insert 2	Intronic region of *SLAIN1* gene	13q22.3
Hind III Insert 3	Intronic region of *MED12L* gene	3q25.1
Hind III Insert 4	Exon 3 of *BCL6B* gene	17p13.1

Two of the thirteen DNA regions belonging to Group 2; Female Bgl II – Insert 1 and Female Bam HI – Insert 1 were PCR amplified, and DNA sequenced. DNA sequence was obtained for three males and three females for each DNA region. A PCR fragment approximately 177bp in length was produced when targeting the region Female Bgl II – Insert 1 and a PCR product approximately 217bp in length was observed when targeting the DNA region Female Bam HI – Insert 1. In both cases, the PCR sequencing results showed some sequencing discrepancies.

When targeting the region Female Bgl II - Insert 1, two single nucleotide polymorphisms (SNPs) and an indel were identified, which were present in all individuals. At position 165 a thymine nucleotide was present in all PCR sequences for both male and female and a cytosine nucleotide was present in all RDA cloned insert sequences. It is likely that this nucleotide substitution could be attributable to a Taq DNA polymerase incorporation error. However, at DNA sequence position 145, all male PCR sequences contained either a cytosine or a thymine residue, whilst all female PCR sequences in addition to the RDA cloned insert sequences contained an adenine nucleotide. This may be an example of a sex-linked SNP. To confirm if this DNA region is sex-linked or not, an investigation of a larger proportion of the DNA region using PCR combined with cloning and sequencing the PCR products to ensure a single DNA template is present within each sequencing reaction is required.

The sequencing results produced from targeting the region Female Bam HI - Insert 1 showed no individual polymorphisms. All sequences using the forward primer were good quality,

whilst all sequences that were obtained using the reverse primer were relatively low quality and could not be analysed further. The most interesting result was that the DNA sequences produced directly from the PCR products were substantially and consistently different from the sequencing results produced using the cloned RDA sequences. It was found that there was one deleted thymine nucleotide in the cloned RDA sequence as well as 5 individual single base pair substitutions and two regions constituting 3 and 4 base pairs respectively, having multiple and consecutive nucleotide substitutions. The RDA experiment generated sequences from three clones all representing this region. The sequencing results from these three clones revealed that all three clones contained a slightly different sequence, with all three differing by one or two nucleotide substitutions. While these three substitutions could be representative of Taq DNA polymerase incorporation mistakes, the PCR sequencing results show at least 12 consistently observed nucleotide substitutions, derived from both our male and female genomic PCR product sequences.

An estimation of the error frequency of the Taq DNA polymerase used in our RDA experiment was performed to determine if the nucleotide differences observed between the genomic PCR and the RDA fragment; Female Bam HI – Insert 1 could be artifactual. Comparing sequences from the two groups 12 nucleotide substitutions were identified over a region of 217bp, potentially one nucleotide error every 18bp. Using the following equation; error rate = 1/[(Taq error frequency/number of cycles) x 2] (assuming a constant Taq DNA polymerase error rate), the error rate that should be observe is 1 nucleotide mutation every 740bp amplified (Lundberg et al. 1991; Eckert and Kunkel 1991). Using the error frequency equation in reverse, calculations suggest that the Taq DNA polymerase error frequency would have to equal 1 nucleotide error in every 2,434bp amplified to produce the result observed, well outside the manufacturer’s described error frequency of 1bp for every 100,000bp amplified (Lundberg et al. 1991; Eckert and Kunkel 1991).

These calculations suggest that the differences observed between the genomic PCR and the RDA fragment; Female Bam HI – Insert 1 cannot been attributed to polymerase errors in the amplification. Therefore, the observed sequences differences are likely to represent two different copies of the same DNA region in the genome of *M. spilota*, with our RDA experiment identifying one copy and our PCR sequencing results possibly representing alternative copies found in males and females. This indicates that the Female Bam HI – Insert

1 DNA region may possibly be sex-linked and suggests that this region requires further investigation.

Conclusion

In conclusion, the modified RDA experiments enriched DNA regions that were present in both the male and female tester populations. The tested modifications to the original published protocol made this process cheaper and quicker. The DNA regions enriched and isolated in the RDA experiments can be classified into two broad groups. Those which contained proportions of retrotransposon elements; Group 1 and those that contained unrelated coding and non-coding DNA regions from the genome; Group 2. We suggest that at least one of the DNA regions belonging to Group 1, Female Hind III – Insert 1, was enriched during the RDA experiments because of a formation of DNA secondary structures produced by the single-stranded DNA molecule during the RDA hybridisation procedure.

We were unable to identify that the DNA regions enriched in our RDA experiments were sex-linked. Negative results produced from this type of molecular genetic experiment do not prove that the targeted DNA, in this case sex specific DNA, does not exist in the genome of *M. spilota*. However, it does provide support that both the male and female genomes of *M. spilota* are very similar and that any sex specific DNA regions that occur are very small, in comparison to previously identified vertebrate sex chromosomes.

Exonic and intronic DNA region representing proportions of seven different genes were identified using the RDA protocol. Investigation into the function of these genes has revealed four (*FSTL1*, *DNAH2*, *CCDC141*, *MBTD1* and *SLAIN1*) are functionality linked to sex reversal, fertility, reproduction functionality, oocyte meiotic maturation and embryonic stem cell differentiation (Casas et al. 2016; Maiti et al. 2000; Hu et al. 2017; Hutchins et al. 2016; Luo et al. 2013; Hirst et al. 2006). A DNA region with sequence homology to the *MBTD1* gene was identified, which is positioned at 17q21.33 in humans and near two genes previously localized to the snake Z chromosome having orthologues at the same location on chromosome 17 in humans. Furthermore, PCR sequencing results from two DNA regions indicate they may be nucleotide polymorphisms in these regions that are possibly sex-linked.

New research has shown that a male heterogametic sex chromosome system most likely exists in two snake species *Boa imperator* and *Python regius*, belonging to the snake families

Boidae and Pythonidae respectively (Gamble et al. 2017). If a related sex determination system also exists in *M. spilota* it gives one possible explanation for why no female-specific DNA fragments were identified in the initial experiments using the original less stringent research methodology (Lisitsyn, Lisitsyn, and Wigler 1993; Yuan 2002). If the male genome had been tested using the less stringent methodology with a positive result, this may have indicated a XX/XY sex chromosome system in M. spilota and the lack of enriched DNA when investigating the female genome was a true representative result. However, this experiment was not conducted.

In general, the DNA regions identified in this RDA experiment when targeting male-specific DNA regions were less likely to be made up of repetitive elements and more likely to have sequence homology to gene regions with sexual functions then DNA regions identified by targeting female-specific DNA. One of the DNA regions identified, Bam HI – Insert 2, which shows sequence homology to the gene *MBTD1*, is mostly likely localised to chromosome Z in snakes, but was identified by targeting male-specific DNA regions. This seemingly contradictory information makes the region the most interesting RDA fragment identified and should be investigated further.

Expanding the research from these observations would involve DNA sequencing of the PCR products produced from the male tester RDA experiment. Cloning the PCR products produced from both male and female RDA experiments and sequencing multiple clones to ensure sequencing quality and to confirm any potential sex-linked polymorphisms is required. Ultimately, use these PCR targets, to identify DNA regions in male and female cosmid libraries. This would allow sequencing of much larger sections of the DNA flanking our identified regions and is the first step required for an investigation using fluorescent *in situ* hybridisation of both male and female snake karyotypes, to determine the DNA region's chromosomal location.

Chapter 4 - Comparative Genomics

Introduction

The current genetics knowledge concerning the sex chromosomes of snakes is largely the result of Kazumi Matsubara's research (Matsubara et al. 2012, 2006). The research group has produced a high resolution genetic map of the snake species *Elaphe quadrivirgata* (Japanese four-striped rat snake), a species which has cytologically observable, female heterochromatic, sex chromosomes (Matsubara et al. 2006, 2012). The group achieved this genetic mapping through the creation of a cDNA library from *E. quadrivirgata* and mapping the corresponding gene locations using fluorescence *in situ* hybridization (Matsubara et al. 2012, 2006). To date, their research has identified the chromosomal location of 183 genes within the *E. quadrivirgata* genome (Matsubara et al. 2012).

In their 2006 study, 11 genes were mapped to the Z chromones of *E. quadrivirgata* and two other snake species; *Python molurus* (Burmese python) and *Trimeresurus flavoviridis* (Okinawa habu snake), represented in Figure 4-1 (Matsubara et al. 2006). No difference was seen in the gene location or gene order, between the proposed Z and W chromosome of *P. molurus*, which represents the relatively old snake lineage; Pythonidae (Matsubara et al. 2006). All 11 genes mapped to the Z chromosome of *E. quadrivirgata,* but only 3 could be mapped to its W chromosome, which belongs to the evolutionary newer snake lineage; Colubridae (Matsubara et al. 2006). While in *T. flavoviridis* all 11 genes were mapped to its Z chromosome, but not one could be identified on its W chromosome (Matsubara et al. 2006). This snake species belongs to the evolutionary youngest snake lineage of the three species investigated; Elapidae (Matsubara et al. 2006). This research showed that all three snake species had a well conserved Z chromosome, but their W chromosomes has differentiated to different degrees (Matsubara et al. 2006). This conclusion is further supported by the cytological evidence, which shows species belonging to the same three families, generally exhibit the same degree of Z and W chromosome differentiation (Matsubara et al. 2006; Olmo and Signorino 2010).

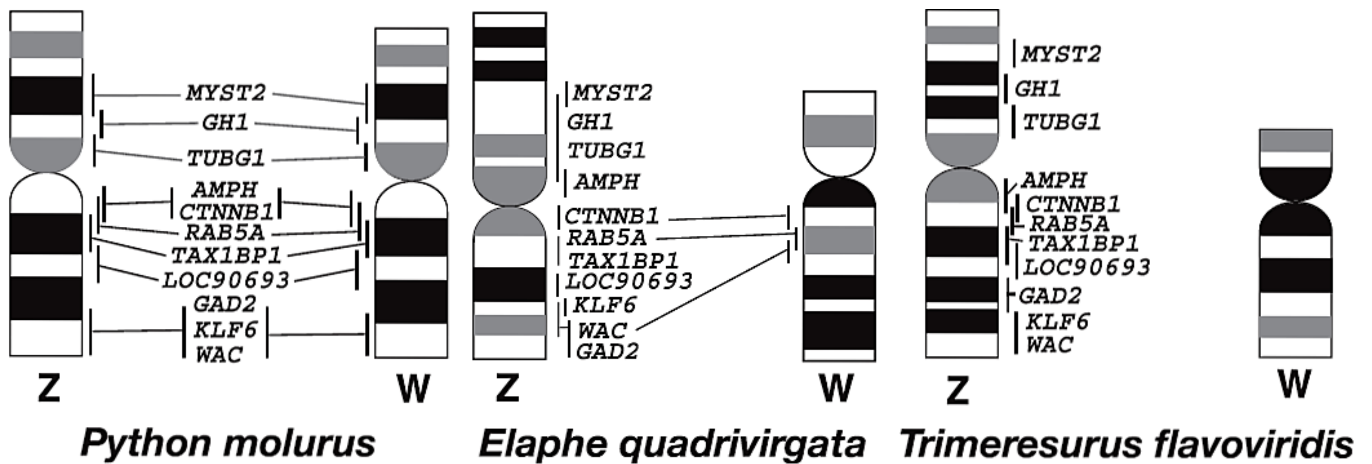

Figure 4-1 cDNA map of Z and W chromosomes of three snake species (Matsubara et al. 2006)

Figure 4-1 has been taken from the publication Matsubara K. et al. 2006. It shows a comparative cytogenic maps of 11 genes from the Z and W chromones of P. molurus, E. quadrivirgata, and T. flavoviridis, with G-banded patterns (Matsubara et al. 2006)*. The Z chromosome of E. quadrivirgata is depicted upside-down because it corresponds to the other two species* (Matsubara et al. 2006).

The research group's supplementary study in 2012 resulted in a total of 20 genes being mapped to the Z chromosome of *E. quadrivirgata* as represented in Figure 4-2 and Table 4-1 (Matsubara et al. 2012). This research revealed that the Z chromosome of *E. quadrivirgata* shows homology to autosomes 2 and 27 in chickens and the autosomes 3, 7, 10 and 17 in humans, providing supportive evidence that the sex chromosomes of snakes may have evolved independently from the sex chromosome system found in either birds or mammals (Matsubara et al. 2012, 2006).

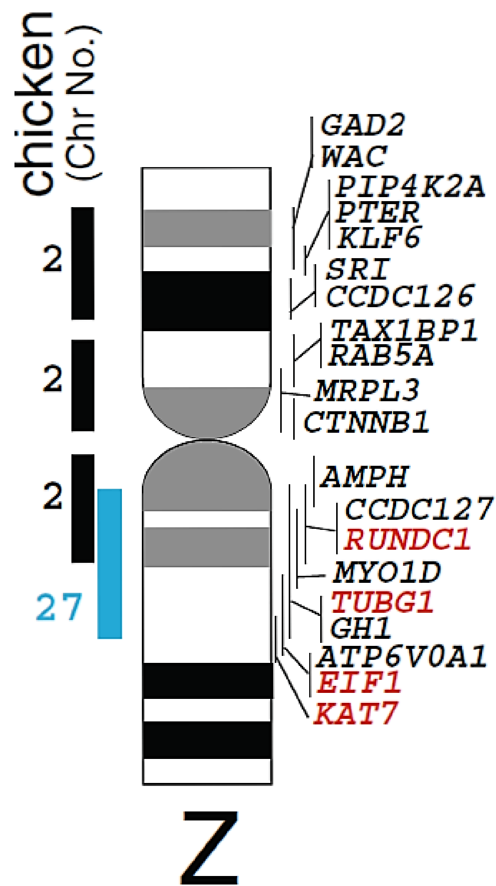

Figure 4-2 cDNA map of the Z chromosome of E. quadrivirgata (Matsubara et al. 2012)

Figure 4-2 has been taken from the publication Matsubara K. et al. 2012. The ideogram of the Z chromosome with G-banding patterns, from E. quadrivirgata (Matsubara et al. 2012)*. GC-rich (GC3 ≥ 50%) and GC-poor (GC3 < 50%) genes are shown in red and black, respectively* (Matsubara et al. 2012)*. Homologous chicken chromosomes and their chromosome numbers are indicated to the left of the snake chromosomes* (Matsubara et al. 2012)*.*

Table 4-1 Genes mapped to E. quadrivirgata Z chromosome

Gene (full name)	**Gene (abbreviation)**
Amphiphysin	*AMPH*
ATPase, H+ Transporting, Lysosomal (Vacuolar Proton Pump) Subunit 1	*ATP6V0A1*
Coiled-Coil Domain Containing 126	*CCDC126* (aka *LOC90693*)
Coiled-Coil Domain Containing 127	*CCDC127*
Catenin Beta 1	*CTNNB1*
Eukaryotic Translation Initiation Factor 1	*EIF1*
Glutamate Decarboxylase 2	*GAD2*
Growth Hormone 1	*GH1*
Lysine Acetyltransferase 7	*KAT7* (aka *MYST2*)
Kruppel-Like Factor 6	*KLF6*
Mitochondrial Ribosomal Protein L3	*MRPL3*
Myosin ID	*MYO1D*
Phosphatidylinositol-5-Phosphate 4-Kinase Type 2 Alpha	*PIP4K2A*
Phosphotriesterase-related	*PTER*
RAS-Associated Protein *RAB5A*, Member RAS Oncogene Family	*RAB5A*
RUN Domain Containing 1	*RUNDC1*
Sorcin	*SRI*
Tax1 Binding Protein	*TAX1BP1*
Tubulin Gamma 1	*TUBG1*
WW Domain Containing Adaptor with Coiled-Coil	*WAC*

In this study, 20 genes, previously mapped to the Z chromosome of *E. quadrivirgata*, will be analysed using our non-model snake species *Morelia spilota* (Matsubara et al. 2012). Using a comparative genomics approach, the cDNA sequences derived from the *E. quadrivirgata* study will be compared to mRNA sequences from other vertebrate species. This will allow conserved regions within the 20 genes of interest to be identified. Using the multiple sequence alignments of these conserved regions, PCR primers specific for each gene will be designed and PCR will be used to investigate the sex specific nature of these gene regions in male and female *M. spilota* DNA samples.

A comprehensive analysis of the gene Catenin Beta 1 (*CTNNB1*) will also be performed, because within the 20 genes being studied, the *CTNNB1* gene is the only gene that has been directly association with sexual differentiation (Chassot, Gillot, and Chaboissier 2014, 1; Chia-Feng et al. 2008). In addition, it is the only gene that has been previously shown to have a sex-linked nature in a snake species (O'Meally et al. 2010).

The *CTNNB1* gene is located on chromosome 3 in humans and produces β-catenin, which is known to comprise a portion of the protein complex that constitutes the adherens junctions in humans (Orsulic and Peifer 1996). While the exact function of the *CTNNB1* protein product remains elusive, DNA mutations and overexpression of the gene, have been associated with many different kinds of cancers in humans including ovarian cancer (Morin 1999; Irving et al. 2015; Na et al. 2017).

β-catenin is also known to play a key role in the transcriptional regulation of the Wnt signal transduction cascade, which has a crucial role in multiple biological functions, including embryonic development (Seung-Min et al. 2013; Brauner et al. 2010; von Gise et al. 2011). In mice, the *CTNNB1* gene is expressed in both male and female gonads (Chia-Feng et al. 2008). Studies in mice have shown that β-catenin is not essential for testis development, but necessary for ovarian development, with developing ovaries that lack β-catenin containing male-specific gonadal structures (Chia-Feng et al. 2008).

The *CTNNB1* gene has been localised to chromosome 2 in the chicken (Seung-Min et al. 2013). It has also been shown to be actively translated in the gonads of both male and female chicken embryos at different stages of embryonic development, post sexual differentiation (Seung-Min et al. 2013). It is hypothesised that β-catenin has an essential role in the differentiation of Sertoli cells during the formation of the seminiferous tubules during testicular development in male chickens (Seung-Min et al. 2013).

In snakes, the *CTNNB1* gene has been mapped to the Z chromosome of *P. molurus, T. flavoviridis* and *E. quadrivirgata*, while only mapped to the W chromosome of *P. molurus* and *E. quadrivirgata* (Matsubara et al. 2012). Investigation into the terminal intron of this gene in *Notechis scutatus* (Tiger snake), has revealed it likely contains a previously undescribed repetitive element (O'Meally et al. 2010). Repeats of this element are localised along the entire length of W chromosome of *N. scutatus*, but the repeat is absent in the Z chromosome (O'Meally et al. 2010). The repeat element was also localised to chromosome 6 in *N. scutatus* (O'Meally et al. 2010).

Materials and Methods

DNA was obtained from blood samples. For this project, high quality DNA of considerable fragment length was necessary for the genomic investigation. Alternative non-invasive DNA sampling methods in general, produce smaller DNA fragment lengths and were unsuitable for use in this project.

DNA Extraction

DNA extraction was performed on all 23 blood samples using the Qiagen DNeasy Blood & Tissue Kit, according to the manufacturer's instructions. Approximately 2 µl of whole blood was used in each DNA extraction, which is less than half the maximum volume recommended by the manufacturer when extracting DNA from blood containing snake nucleated erythrocytes. The quantity and quality of DNA extracted using the Qiagen kit was measured by agarose gel electrophoresis, 5µl of each DNA extraction was separated on a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer and spectral analysis using a NanoDrop spectrophotometer. The DNA extracts contained approximately 50-100 ng/µl of DNA. The maximum 200 µl of AE buffer was used to elute the DNA, giving approximately 10-20 µg of DNA per 2 µl of whole blood extraction. All DNA extracts had absorbance ratio 260 nm/280 nm of between 1.8 and 2.0, indicating that the DNA extractions were essentially free of protein contamination. The 260 nm/230 nm ratio results were disregarded because the AE elution buffer contains EDTA, which is known to affect this ratio.

Comparative Genomic Analysis

Oligonucleotide primers were designed to target the 20 genes of interest previously localised on the Z chromosome in *E. quadrivirgata* (Matsubara et al. 2012). This was achieved by using a comparative genomic approach. The original reference cDNA sequences used to create the *E. quadrivirgata* gene map were made available by the authors of the study and can be downloaded from the DNA Data Bank of Japan (Mashima et al. 2017). The National Center for Biotechnology Information was also used as a source of mRNA sequences, for each gene, from a variety of different vertebrates (NCBI Resource Coordinators 2017). The mRNA sequences from a minimum of 7 different species for each gene investigated were used. DNA sequences that had the status provisional, predicated, or inferred were not used. This is because these sequences may contain inaccurate sequence information as the database status describes. No mRNA sequences from any snake species, representing the 20 genes of interest were found,

despite an intensive search. Therefore, sequences from the nearest possible reptile species, for which data was available, was used in the comparative analysis. However, mRNA sequences from reptile species were extremely rare as well. Most mRNA sequences that were used came from model vertebrate species, such as mouse, human, rat, pig, cow, zebra fish, zebra finch and the clawed frog.

Each group of sequences, representing each of the 20 genes investigated, were aligned using the bioinformatics software program Geneious (Biomatters Ltd, Auckland, NZ). Areas of conservation within the sequences were identified that could be used to design PCR primers. For each gene the full annotated DNA sequence from *Gallus gallus* was obtained from the NCBI database, and our multiple sequence alignment was mapped to this reference. This allowed us to estimate the intron/exon configuration of the cDNA sequences and the potential PCR product size. The resulting multiple alignment and chicken reference map of the gene *EIF1* is represented in Figure 4-3 as a typical example of the analysis strategy. This alignment allowed DNA sequences for the forward and reverse primers to be identified, as shown in Figure 4-3.

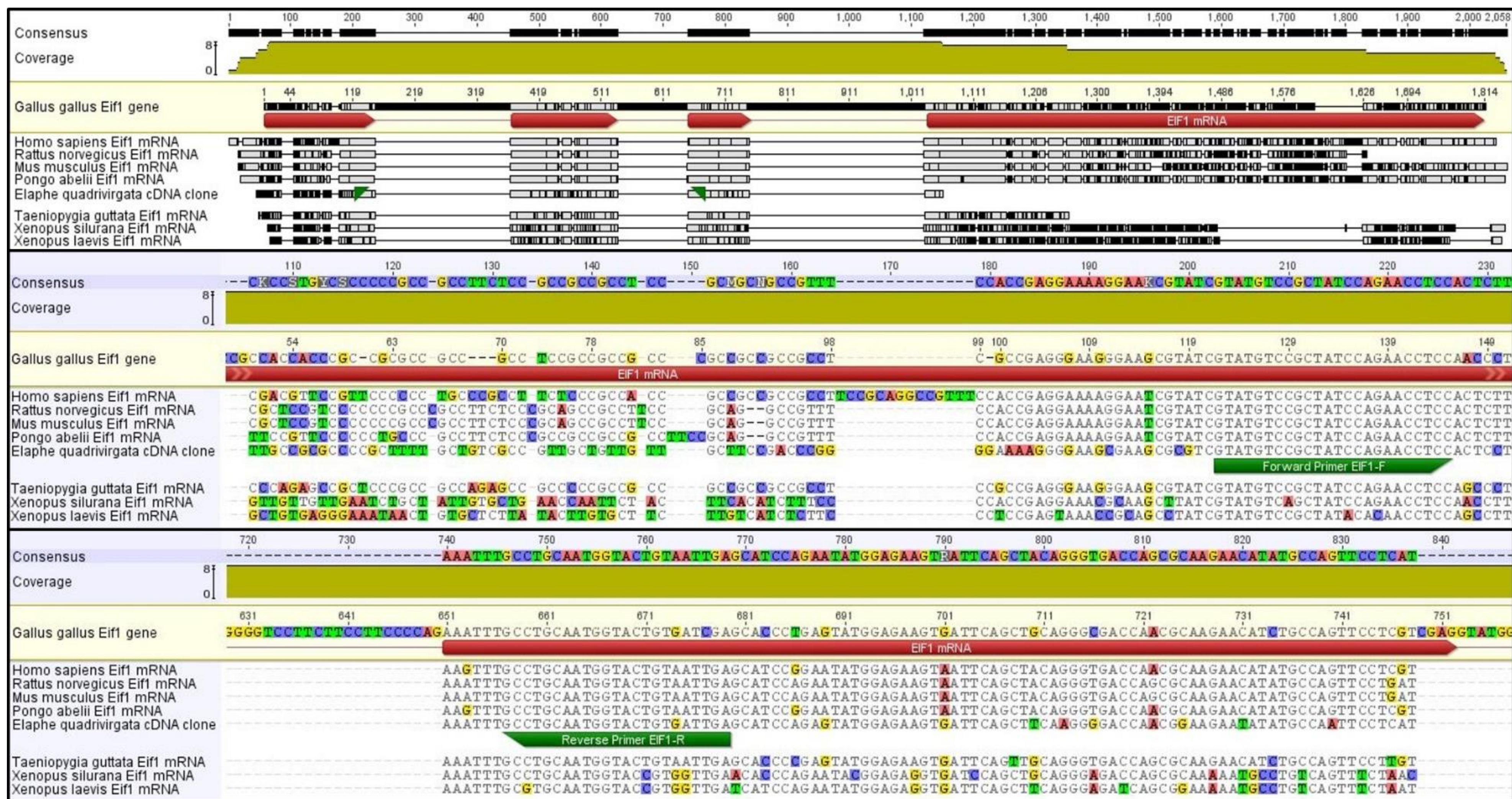

Figure 4-3 Multiple sequence alignment of reference EIF1 gene sequence

Figure 4-3 is the resulting multiple alignment of the EIF1 gene, including cDNA sequence from E. quadrivirgata, and mRNA sequence from Homo sapien, Rattus norvegicus, Mus musculus, Pongo abelii, Taeniopygia guttata, Xenopus silurana and Xenopus laevis using the Gallus gallus EIF1 gene sequence as a reference map. The figure is divided into three sections; first representing the whole EIF1 gene, second showing the forward primer selection and third showing the reverse primer selection.

Primer Design

A forward and reverse primer were designed for each of the 20 genes, from the multiple sequence alignments (Table 4-2). Due to the complicated nature of the 20 separate multiple sequence alignments, no single software program existed that could be used to pick out suitable PCR primer pairs. Therefore, each priming site was selected carefully, by visual analysis, to best fit the following criteria:

1. Areas of cDNA with the highest conservation between species were used in preference.
2. If nucleotides within the final primer sequence were less than 50% conserved across the different species investigated, redundant nucleotides were incorporated.
3. The target for primer length was between 19-25bp.
4. The target melting point of each primer was between 55-65°C.
5. The melting point of each primer was designed to not exceed 3°C of the melting point of its primer pair.
6. The targeted PCR product length was between 500-1500bp.
7. The target GC content was between 40-60%.
8. Primer to primer interactions were access using the oligo analyses feature of the software program Vector NTI, produce by Invitrogen. Primer dimers and hairpin loops were avoided were possible, but generally kept to a minimum.

Table 4-2 Primer pairs for gene targets

Gene Target	Oligo Name	Oligo Sequence 5' to 3'	Length (bp)	GC %	Calculated Tm (°C)
AMPH	AMPH-F	TGTGATGTRCTCTGGGAAGACTTCC	23	50.0	57.9
	AMPH-R	ACTGTCRTARTCCACNAGCTTCCTGC	22	51.9	60.1
ATP6V0A1	ATP6V0A1-F	TCATCCTGGAGTGTCCGGCC	20	65.0	59.3
	ATP6V0A1-R	ATCAATTCCAAAGGGATATGCTCC	24	47.7	56.4
CCDC126	CCDC126-F	CTGATCTGAAAAGAACAATTGCGG	24	41.7	56.8
	CCDC126-R	TATTTGTTGCTGAACCATTTACAACC	26	34.6	55.5
CCDC127	CCDC127-F	GAAGATCAGTGGCAAACAAAAGCC	24	45.8	58.0
	CCDC127-R	TCCCAATATCGAAGGTAGATCCACAT	26	42.3	57.9
CTNNB1	CTN1B-F	AGGAAGGCATGGAAGGTCTTCTGGG	25	56.0	63.4
	CTN1B-R	AGGCCAGTGGGATGGTGGATGC	22	63.6	63.7
EIF1	EIF1-F	GTATGTCCGCTATCCAGAACCTCC	24	54.2	56.9
	EIF1-R	TCAATCACAGTACCATTGCAGGC	23	47.8	56.4
GAD2	GAD2-F	GATTTGTTCCTTTTCTAGTAAGTGCC	26	38.5	53.4
	GAD-R	GACATAGTAGCCCACCGCCCC	21	66.7	59.2
GH1	GHI-F	CTGATTCTCATTCAGTCCTGGC	22	50.0	52.9
	GHI-R	CTGAAGCAGKAGAGCAGMCCG	21	61.9	56.6
KAT7	MYST2-F	GATGAGAGCATTGCCAAGGACATG	24	50.0	59.5
	MYST2-R	TGMTTATCHCGGCTCTGTGCTC	22	53.8	56.6
KLF6	KLF6-F	TGGGAAGGKTGTGAGTGGCG	20	62.5	59.2
	KLF6-R	CTCTTCATGTGMAGGGCCAGG	21	59.5	57.6
MRPL3	*MRPL3*-F	GGTGGGATGAGCATCTTTCAGAAG	24	50.0	57.9
	MRPL3-R	CCTGCTTAGTCCATAATGGCATCAT	25	44.0	57.4
MYO1D	MYO1D-F	GTAATATCAGGGGAAAGCGGAGC	23	52.2	57.1
	MYO1D-R	TCGCATTGCCAAAAGCTTCC	20	50.0	57.5
PIP4K2A	PIP4K2A-F	TGATGCCAGATGACTTCAAAGC	22	45.5	54.3
	PIP4K2A-R	TTGAATTTGAAATGGCTTGGC	21	38.1	54.3
PTER	PTER-F	GGAAATTGGCTGTTCTTGGCC	21	52.4	57.3
	PTER-R	ACCGTTTTTGAGATATCAGCCCC	23	47.8	57.3
RAB5A	RAB5AB-F	AGTCCTAACATTGTAATAGCTTTA	24	29.2	44.7
	RAB5AB-R	ATTGCCATAAATATTTCATTTAC	23	21.7	45.3
RUNDC1	RUNDC1-F	TGGTCATGGAGAGACAGCAGG	21	57.1	54.9
	RUNDC1-R	ACATTTCAAGATCCCGGATCTG	22	45.5	54.5
SRI	SRI-F	CTTTGGGCTATACTGAATGGCTGG	24	50.0	57.9
	SRI-R	GCCTGTGGACTCAATCTAAATCCC	24	50.0	56.9
TAX1BP1	TAXBP1-F	CAGGCTTTTGTTTTGATTCCA	21	38.1	52.1
	TAX1BP1-R	GGNGGGAACTGCTCGCTGC	19	71.1	59.8
TUBG1	TUBG1-F	GGTCAGYGGSCTMATGATGG	20	60.0	55.8
	TUBG1-R	GAGCAGTTCCGYAAGGAGGACA	22	56.8	57.6
WAC	WACB-F	ACATAATGACAGAGACTACAGACTGGCC	28	46.4	57.3
	WACB-R	GCTGATGATCTGACTGTAGAGAAAATGG	28	42.9	57.7

CTNNB1 Gene

Primer pairs were designed to target multiple regions within the gene *CTNNB1* (Table 4-3). Apart from intron number 1, which proved to be extremely variable between vertebrate species, primer pairs were designed targeting every intronic region within the *CTNNB1* gene. Each primer pair was design to amply intronic DNA by binding to the conserved region of DNA contained in each flanking exon, using the same comparative genomic method previously described. Once designed all primer pairs were purchased from the manufacturer GeneWorks.

Table 4-3 Primer pairs targeting the CTNNB1 gene

***CTNNB1* Targets**	**Oligo Name**	**Oligo Sequence 5' to 3'**	**Length (bp)**	**GC %**	**Tm (°C)**
Exon 2 to 3	CTN23F	CCAGCGTGRAMAATGGCTACSC	22	59.1	61.3
	CTN23R	GGACTGRGARAAGCCSTGSTCC	22	63.6	60.3
Exon 3 to 4	CTN34B-F	GTGGCAAGGGAAACCCAGAAGAGG	24	58.3	62.8
	CTN34B-R	CCGGGAACATGGCAGCACGC	20	70	65.1
	CTN34F	ACTGGCAGCARCAGTCTTACCTGG	24	56.3	59.5
	CTN34R	GCACGDGTGGCMAGTTCHGC	20	65.8	60.1
Exon 4 to 5	CTN45F	NGATGARGGCATGCAGATSCC	21	57.1	59
	CTN45R	ATGCCNCCAGAYTTAAAGATGGCC	24	50	60.7
Exon 5 to 6	CTN56F	GGCTGCDGTTATGGTNCASCAGC	23	60.1	61.6
	CTN56R	TGAAGSCAGTCTGTYGTSATDGCC	24	53.5	60.8
Exon 6 to 7	CTN67F	GAAGGAGCHAAAATGGCHGTNCG	23	52.9	60.5
	CTN67R	CTTCAGCACYCKSCTTGTGGTCC	23	60.9	61.5
Exon 7 to 8	CTN78F	ATCATACTGGCHAGTGGDGGACC	23	55.1	57.2
	CTN78R	GCATCTGANAGSTTYCTNAGAGTCC	25	50	55.9
Exon 8 to 9	CTN89F	CGTCTKGTTCARAASTGTCTTTGG	24	45.8	55.7
	CTN89R	AGAGGCCAGTGDGADGGHGG	20	65	57.2
Exon 9 to 10	CTN910B-F	GTGGAATTGAGGCTCTTGTGCGC	23	56.5	61.7
	CTN910B-R	GTCTTGGTATTGCACCTTGTTCACGC	26	50	61.6
	CTN910F	ATGTKGTVACBTGTGCHGCTGG	22	55.3	58.4
	CTN910R	AAAYTGCTGYTGBGTBCCACCC	22	56.1	60.8
	CTN1seq-F	TAAAATGATGGTGTGCCAAG	20	40	48.5
	CTN1seq-R	ACCATAATGAAGACGCACAG	20	45	47.5
Exon 10 to 11	CTN1011F	CDHTGCGNGARCAGGGTGC	19	66.7	58.6
	CTN1011R	AATYCGGTTGTGSACSTCHCG	21	56.3	59.1
Exon 11 to 12	CTN1112F	CCGCATGGARGARATHGTDGAAGG	24	52.8	61
	CTN1112R	GAGGAGCTGTDGCDCCYTCNGC	22	66.7	61.7
Exon 12 to 13	CTN1213F	AAASATCCAAAGAGTRGCHGCAGG	24	49.3	59
	CTN1213R	GBGGYTTGTCCTCAGACATDCGG	23	58.7	61.1
Exon 13 to 14	CTN1314F	TGCAGCTGCTGTTSTVTTCCG	21	55.6	57.8
	CTN1314R	GSGGTTCDCCYTGGGCACC	19	72.8	61.5
Exon 14 to 15	CTN1415F	GGTGCCCARGGHGAACCSC	19	72.8	61.5
	CTN1415R	CACAGCAGGTTACAAYAACTTTGGG	25	46	57.3
Exon 15	CTN15F	CCCARGASCTBATGGASGGG	20	65.8	58.4
	CTN15R	GTTCCCATAGGAAACTCAGCTTGG	24	50	57.1

PCR optimisation

Each primer pair went through PCR optimisation. This was to ensure that each primer pair produced a single DNA band during PCR amplification, to facilitate subsequent DNA sequencing. This process was performed using both male and female DNA templates, side by side. This is to ensure that no sex-linked DNA mutations that could potentially be identified

during the PCR optimisation were overlooked. This optimisation process was achieved by optimising the temperature and incubation times during denaturation, annealing and extension of the PCR reactions and the number of PCR cycles.

An initial temperature gradient PCR was performed at a series of different annealing temperatures, in a range approximately 5°C above and 5°C below the calculated primer pair average Tm. One negative control reaction (template DNA absent) was also included at the lowest annealing temperature in each gradient PCRs. Each PCR was performed in a total volume of 10 µl containing 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10X PCR buffer, 0.1 µM of each primer and 30 ng of template DNA. The thermal profile was 95°C for 10 min, followed by 35 cycles of: 95°C for 60s, annealing temperature gradient for 90s, 72°C for 120s, final extension for 10 min at 72°C. The PCR products were examined using agarose gel electrophoresis on a 1.5% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer.

DNA Sequencing

Once the PCR optimisation process resulted in the production of a single DNA band for a primer set; five identical 10 µl PCR reactions were performed, combined, visualised using agarose gel electrophoresis and sent for DNA sequencing. The PCR products were sequenced by Macrogen Korea, 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511, Rep. of Korea. PCR clean-up, sequencing reaction and DNA sequencing was performed on samples derived from at least 3 male and 3 female *M. spilota* for every primer pair optimised. Geneious (Biomatters Limited), was used to visualise and assess the resulting DNA sequences. RepeatMasker, NCBI's BLAST and Genscan were all used to investigate features within each DNA sequence produced (NCBI Resource Coordinators 2017; Smit, Hubley, and Green 2013; Burge and Karlin 1997).

Once sequence analysis was complete and a high-quality consensuses sequence derived, the consensus sequences were used as the reference in a sequence search using *Python bivittatus* (Burmese python) unplaced genomic scaffold databases 2011 and 2013, available for download from NCBI; Python_molurus_bivittatus-5.0.2 reference Annotation Release 101 (Castoe, de Koning, et al. 2011; Castoe et al. 2013). It is important to note that the 2013 version of *P. bivittatus* genome only became available after this research had been completed.

Primer Re-design

The primer sets which failed to produce a single PCR fragment during the PCR optimisation stage or failed to produce a DNA sequence of high enough quality during DNA sequencing, were re-designed (Table 4-4). Some DNA regions were also investigated in greater depth, by using the determined DNA sequences to produce new PCR primers. This was done to investigate possible sex-linked markers, or to extend the length or improve the quality of the DNA sequences previously generated.

Table 4-4 Primer pairs redesigned

Gene Target	Oligo Name	Oligo Sequence 5' to 3'	Length (bp)	GC %	Calculated Tm (°C)
AMPH	AMPH-F	ATTGCAAAAACTCCATCACCACC	23	43.5	56.7
	AMPH-R	GGAGGTCCTTTCCTCAGCTGTG	22	59.1	56.6
GAD2	GAD2-F	TGCAAAAAGTACAAGATCTGGATGCA	26	38.5	58.8
	GAD2-R	AGCAGCCCACCACCCC	16	75.0	53.5
GH1	GHI-F	CCTGCTCTGCCTGCCCTGG	19	73.7	60.3
	GHI-R	TTGTGCAGGTCTTTCTTGAAGC	22	45.5	53.8
KAT7	KAT7-F	AATGCCTCAGTACATGAGACAGGG	24	50.0	56.1
	KAT7-R	CCAGTACTTGAGCATCTGAAGGGC	24	54.2	54.0
KLF6	KLF6-F	ACCTGCGGAAAAAAGACCCC	20	55.0	56.7
	KLF6-R	AAGTCTGTTGCCAGTATTCCTCCAG	25	48.0	56.8
RAB5A	RAB5A micro F	AATGGCTGGGAAGCAACTGA	20	50.0	54.6
	RAB5A micro R	ACTCTTCAATTTCATCTGACAGCTC	25	40.0	52.8
TAX1BP1	TAX1BP1-F	CAGGCTTTTGCTTTGATTCCAGC	23	47.8	59.0
	TAXBP1-R	CTTGCTTTGATCATAGTTAGGAGG	24	14.7	51.3
TUBG1	TUBG1-F	TTCCCAAATCAACCAGCTGG	20	50.0	54.8
	TUBG1-R	GGTTGGATCCACTTCCCCC	19	63.2	55.1

To further analyse a microsatellite region observed within the gene *RAB5A*, a new primer set was designed (RAB5A micro F and RAB5A micro R) and real-time PCR, including melt-curve analysis performed on this DNA region. This experiment was performed using Applied Biosystems ViiA7 High Productivity Real Time PCR System and analysis achieved using their inbuilt software. The PCR reaction was performed using the following reagents; 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10X PCR buffer, 0.1 µM of each primer, 0.3 µl of 20X Eva Green qPCR dye and 30ng of template DNA. The thermal profile was 95°C for 10min, followed by 45 cycles of: 95°C for 15s, 58°C for 30s, 72°C for 30s, final extension at 72°C for 10min. All 23 *M. spilota* samples of known sex were investigated in this experiment using the above protocol.

In some cases, during PCR optimisation and visualisation by agarose gel electrophoresis two of more, strong distinct PCR products were observed. In these cases, agarose gel extraction was used to separate the different sized PCR products and the DNA from each band was extracted from the gel, so they could be DNA sequenced separately. A QIAquick Gel Extraction Kit, (Qiagen) was used according to the manufacturer's protocol.

In some cases, during DNA sequencing the results showed that certain PCR products produced a mixed signal, which rendered the results unusable. Further investigation determined that the cause of this problem was most likely a DNA sequencing reaction that contained more than one DNA template. In some situations, this problem was resolved by cloning the PCR products using the TOPO TA for sequencing cloning kit (Thermo Fisher Scientific), followed by sequencing several clones. This ensured a good DNA sequencing signal and sequencing result. The protocol was followed according to the manufacturer's protocols.

Results

Using a comparative genomic approach, PCR products from 17 of the 20 genes were obtained (Table 4-5). For several genes, the initial PCR optimisation process failed to produce a single DNA fragment. However, after re-designing the primer pairs, single PCR products were produced for several more gene targets. Some PCR reactions could not be optimised to a single PCR product even after two sets of primer pairs were tested. In these cases, the PCR products were cloned into bacterial plasmids or if the fragment size difference permitted, they were separated using agarose gel electrophoresis, excised and each fragment extracted using an agarose gel extraction kit, and sent for subsequent DNA sequencing. Despite multiple attempts using a variety of different techniques, no single PCR product for the three genes *CCDC126*, *CCDC127* and *MYO1D* were obtained. These three genes were also found to be the least well conserved among our model vertebrate mRNA and cDNA sequences and therefore the most bioinformatically challenging to design specific PCR primers to target.

Table 4-5 Primer pairs, optimised PCR conditions

PCR Primer Pairs	Tm (°C)	Denaturation (s)	Annealing (s)	Extension (s)	PCR Cycles
AMPH-F & AMPH-R	57	30	60	120	50
ATP6V0A1-F & ATP6V0A1-R	56	30	60	120	50
CCDC126-F & CCDC126-R		No Single PCR Product Produced			
CCDC127-F & CCDC127-R		No Single PCR Product Produced			
EIF1-F & EIF1-R	56	30	60	120	50
GAD2-F & GAD-R	55	30	60	120	35
GHI-F & GHI-R	51	30	60	120	40
MYST2-F & MYST2-R	51	30	60	120	50
KLF6-F & KLF6-R	60	30	60	120	50
***MRPL3*-F & *MRPL3*-R**	52	30	60	120	50
MYO1D-F & MYO1D-R		No Single PCR Product Produced			
PIP4K2A-F & PIP4K2A-R	56	30	60	120	50
PTER-F & PTER-R	53	30	60	120	50
RAB5AB-F & RAB5AB-R	54	30	30	120	50
RAB5A micro F & RAB5A micro R	58	15	30	30	45
RUNDC1-F & RUNDC1-R	54	30	60	120	45
SRI-F & SRI-R	55	30	60	120	50
TAXBP1-F & TAX1BP1-R	51	30	60	120	40
TUBG1-F & TUBG1-R	57	30	60	120	50
WACB-F & WACB-R	45	30	60	60	50

Subsequent DNA sequencing of the PCR products derived from the optimised PCR reactions, resulted in DNA sequences for 14 of the 20 genes targeted. No readable DNA sequence data for the PCR products representing the genes *KAT7*, *KLF6* and *TUBG1* were obtained, despite producing a single PCR product for these three genes. The PCR products for all three genes returned sequencing results consistent with a mixed sequencing template signal. This was despite redesigning the PCR primers targeting the three genes and cloning the PCR products and sequencing several clones. This result was consistent between sexes for all three genes. Because of limited sequence information available for these three genes, in snake species, identifying the genomic DNA sequence for these genes was not possible.

The sequencing results from the 14 genes are represented in Table 4-6. Geneious was used to align the DNA sequences identified for each gene target and evaluate their DNA features. Seven single nucleotide polymorphisms (SNPs) and three microsatellites were identified, one microsatellite was monomorphic and two were polymorphic. Consensus sequences were constructed, and repetitive sequences were identified using RepeatMasker. Two simple

repeats, three Short Interspersed Elements (SINEs), one ribosomal RNA (rRNA) repeat, one Long Interspersed Nuclear Element (LINE) and one Long Terminal Repeat (LTR) were identified across all the gene sequences.

Table 4-6 Gene target sequencing results

Gene Targets	Length (bp)	SNPs/Repeats	RepeatMasker Results	BLAST Results	Identities
AMPH	1446		rRNA - 681-723	PREDICTED: *P. bivittatus* GRAMD1A mRNA	54/68 (79%)
ATP6V0A1	384			PREDICTED: *P. bivittatus ATP6V0A1* mRNA	83/88 (94%) 41/44 (93%)
CTNNB1	2985 (in total)	SNP - Intron 3 - 11 (A/G), SNP - Intron 6 - 16 (C/T)	SINE - 226-315	PREDICTED: *P. bivittatus CTNNB1* mRNA	Refer to *Table 4-9*
EIF1	1049	SNP - 294 (C/G), Microsatellite - 694 (G)8,9,10	Simple Repeat - 961-1002 (CCCCTC)	PREDICTED: *P. bivittatus EIF1* mRNA	165/165 (100%)
GAD2	717	SNP - 104 (T/C), SNP - (625, A/C).		PREDICTED: *P. bivittatus GAD2* partial mRNA	89/89 (100%)
GH1	1093			*Mauremys reevesii* GH gene complete cds	82/90 (91%) 66/85 (78%)
KAT7	No DNA Sequence Produced				
KLF6	No DNA Sequence Produced				
MRPL3	884			PREDICTED: *P. bivittatus MRPL3* mRNA	121/124 (98%) 48/49 (98%)
PIP4K2A	1451		Simple Repeat – 1238-1267 (CTC)	PREDICTED: *Poecilia mexicana* uncharacterized ncRNA	36/42 (86%)
PTER	177			PREDICTED: *P. bivittatus PTER* mRNA	168/174 (97%)
***RAB5A* Large**	2123	Microsatellite - 311 (C)9,10,12,13	SINE - 1572-1643	PREDICTED: *P. bivittatus RAB5A* mRNA	44/45 (98%)
***RAB5A* Small**	338			PREDICTED: *P. bivittatus* MIER2 mRNA	313/316 (99%)
RUNDC1	353			PREDICTED: *P. bivittatus RUNDC1* mRNA	183/184 (99%) 23/23 (100%)
SRI	1488		LINE - 240-451, LTR - 143-174	PREDICTED: *P. bivittatus SRI* mRNA	59/59 (100%)
TAX1BP1	1300	Microsatellite - 744 (CT)4	SINE - 680-754	PREDICTED: *P. bivittatus* SUOX mRNA	744/770(97%) 319/334(96%)
TUBG1	No DNA Sequence Produced				
WAC	257	SNP - 173 (T/C), SNP - 236 (T/C)		PREDICTED: *P. bivittatus WAC* mRNA	256/257 (99%)

The ten genes (Table 4-6); *CTNNB1*, *EIF1*, *GAD2*, *MRPL3*, *PIP4K2A*, *PTER*, *RAB5A*, *RUNDC1*, *SRI* and *WAC*, of the 14 gene targets sequenced, all matched predicated mRNA sequences of the corresponding gene from *P. bivittatus*. With identity matches being 94% or higher. The gene *GH1*, matched a sequence belonging to the *GH1* gene from *Mauremys reevesii* (Chinese Pond Turtle). The BLAST results had a lower sequence identity of 91%, as would be expected from a more distantly related reptile species. This suggests that the *GH1* gene from *M. spilota* is the first genomic DNA sequence of the *GH1* gene, in any snake species. Three DNA sequences targeting the genes *AMPH*, *PIP4K2A* and *TAX1BP1* failed to produce a NCBI nucleotide BLAST search match that corresponded to their respective gene target.

Amphiphysin (*AMPH*) Gene

The BLAST match with the greatest sequence similarity to the DNA sequence produced resulting from the PCR of gene *AMPH*, was a predicted mRNA sequence derived from *P. bivittatus*, from a different gene than expected; GRAM Domain Containing 1A (*GRAMD1A*). The number of sequences returned from BLAST was diminutive and this result was only achieved using the BLASTn algorithm, which targets similar DNA sequences. The BLAST result showed a sequence identity of 79% across 68bp from a total of 1446bp of the query sequence. Other BLAST results also showed an alignment to this same small sequence, but with a lower identity, and were from a variety of different predicted *P. bivittatus* mRNA sequences, representing 25 other genes. A second smaller region 30bp downstream of this initial match also matched sequences using the BLASTn algorithm. This second DNA region has been previously identified using RepeatMasker as a ribosomal RNA repetitive sequence.

A search of this unidentified DNA sequence with the NCBI *P. bivittatus* Annotation Release 101 revealed a match with the *P. bivittatus* unplaced genomic scaffold number 958 (NW_006532972), at a 91% sequence identity across the entire length of the query sequence. Although this unplaced genomic scaffold contained several defined and undefined genes, the sequence matched an unbroken portion of non-coding DNA region located between the two genes Purkinje Cell Protein 4 (*PCP4*) and DS Cell Adhesion Molecule (*DSCAM*).

Phosphatidylinositol-5-Phosphate 4-Kinase Type 2 Alpha (*PIP4K2A*) Gene

Approximately 1451bp of *M. spilota* genomic DNA was sequenced for the gene *PIP4K2A*. NCBI's BLAST nucleotide search showed no significant matches to the expected gene target.

The top BLAST result achieved was a predicted uncharacterised RNA sequence from *Poecilia mexicana* (Shortfin Molly Fish), which aligned with 36bp of our query sequence at 86% sequence identity. This 36bp region was identified previously using RepeatMasker as a simple repeat containing a CTC motif. A search of the NCBI *P. bivittatus* Annotation Release 101 identified two matches within the *P. bivittatus* unplaced genomic scaffold number 3235 (NW_006535249). This unplaced genomic scaffold contained an incomplete assembly of the *PIP4K2A* gene and our sequence aligned a total of 933bp with 93% sequence identity and with a further 519bp having 98% sequence identify.

Tax1 Binding Protein (*TAX1BP1*) Gene

For the gene *TAX1BP1*, a DNA sequence 1300bp in length was obtained. There was no NCBI BLAST result that corresponded to our targeted gene. The two top BLAST results showed a sequence identity of 84% and 83% respectively for a predicted mRNA Sulfite Oxidase (*SUOX*) gene from *P. bivittatus* and a microsatellite sequence from *Agkistrodon contortrix* (Copperhead Snake). However, most BLAST hits that resulted from this analysis, aligned with a SINE repetitive element identified using RepeatMasker. This SINE also contains a short non-polymorphic dinucleotide microsatellite repeat at one end and the BLAST search returned a series of similar microsatellite sequences in other vertebrate species.

A BLAST search of the NCBI *P. bivittatus* Annotation Release 101 identified two regions of similarity across the *P. bivittatus* unplaced genomic scaffold number 598 (NW_006532612). This scaffold contains an uncharacterised gene; *LOC103060612* followed by the gene JAZF Zinc Finger 1 (*JAZF1*), followed by the last two exons and an intron of another uncharacterized gene; *LOC103060211*. The GenBank data from the genomic scaffold showed that the *JAZF1* gene and the uncharacterised *LOC103060211* gene are transcribed in opposite directions and both share the same DNA region as their terminal exons, overlapping by 1444bp. Our query DNA matched 770bp at 97% sequence identity and 334bp at 96% sequence identity at the location where these two genes overlapped. When an extracted DNA sequence from scaffold 598 representing the uncharacterised gene *LOC103060211* was used to search NCBI's nucleotide database, the top match was a predicted mRNA sequence from the *P. bivittatus* gene JAZF zinc finger 1 (*JAZF1*). However, the next BLAST hit showed a predicted mRNA sequence of the *TAX1BP1* gene, transcript variant X2 in *Protobothrops mucrosquamatus* (Brown Spotted Pit Viper). Further BLAST hits represented the same *TAX1BP1* gene in *Pogona vitticeps* (Central bearded dragon), *Gekko*

japonicus (Schlegel's Japanese gecko) and *Anolis carolinensis* (Carolina anole) as well as 28 other vertebrate species.

Genomic Sequencing Results

In total, DNA sequences totalling 16,045bp from the *M. spilota* genome were obtained (Table 4-7). At least one sequence representing a male and female genome for each gene investigated was identified, however in most cases this number was higher. No sex specific differences were observed between the DNA sequences studied. Consensus DNA sequences were used to identify DNA sequence contig from the Python_molurus_bivittatus-5.0.2 reference Annotation Release 101. A specific genomic scaffold representation for every DNA sequence obtained from *M. spilota* was identified (Table 4-7). None of these scaffolds overlapped. Twelve individual genomic scaffolds that represent the Z chromosome in *P. bivittatus* were identified, which together totals 4,383,299bp of DNA.

Table 4-7 P. bivittatus BLAST matches 2013 data

Gene Targets	Length (bp)	*P. bivittatus* Scaffold	Sequence BLAST Identity	Sequences Description	Scaffold Length (bp)
AMPH	1446	958 (NW_006532972)	1307/1429 (91%)	Non-coding region between PCP4 and DSCAM genes	289697
ATP6V0A1	384	1260 (NW_006533274)	368/379 (97%)	*ATP6V0A1* gene	286434
CTNNB1	2985	1101 (NW_006533115)	2891/3023 (96%)	*CTNNB1* gene	294141
EIF1	1049	971 (NW_006532985)	814/845 (96%)	*EIF1* gene (assembly gap)	313866
GAD2	717	8881 (NW_006540895)	693/716 (97%)	LOC103063935 gene (predicted mRNA *GAD2*)	26497
GH1	1093	462 (NW_006532476)	1061/1108 (96%)	LOC103053760 gene (predicted mRNA *GH1*)	435108
MRPL3	884	2270 (NW_006534284)	847/898 (94%)	*MRPL3* gene	194271
PIP4K2A	1451	3235 (NW_006535249)	1377/1454 (95%)	*PIP4K2A* gene	145982
PTER	177	901 (NW_006532915)	168/174 (97%)	*PTER* gene	331478
***RAB5A* Large**	2123	618 (NW_006532632)	1881/1994 (94%)	*RAB5A* gene	394516
***RAB5A* Small**	338	2972 (NW_006534986)	313/316 (99%)	MIER2 gene	159866
RUNDC1	353	485 (NW_006532499)	349/353 (99%)	*RUNDC1* gene	434177
SRI	1488	1017 (NW_006533031)	1433/1504 (95%)	*SRI* gene	308530
TAX1BP1	1300	598 (NW_006532612)	1215/1258 (96%)	LOC103060211 gene	400556
WAC	257	44 (NW_006532058)	256/257 (99%)	*WAC* gene	817743

Catenin Beta 1 (*CTNNB1*) Gene

In total 17 primer pairs were designed to amplify 13 intronic regions, as well as terminal exon 15 of the *CTNNB1* gene (Table 4-8). In some cases, where the introns length permitted, primer pairs that overlapped were designed, giving a sequence contig spanning more than one intron. In some examples, primer pairs were redesigned, when the initial primer pair failed to produce an adequate PCR result. One primer pair; CTN1B-F and CTN1B-R was designed, using sequence data obtained from *M. spilota* genomic DNA sequence, to produce a more robust PCR that was required when investing the presence of the *CTNNB1* gene in further experiments.

Table 4-8 CTNNB1 primer pairs, optimised PCR conditions

PCR Primer Pairs	Tm (°C)	Denaturation (s)	Annealing (s)	Extension (s)	PCR Cycles
CTN23F & CTN23R	No PCR Product Produced				
CTN34F & CTN34R	No PCR Product Produced				
CTN34B-F & CTN34B-R	62	60	90	120	35
CTN45F & CTN45R	No Single PCR Product Produced				
CTN56F & CTN56R	No Single PCR Product Produced				
CTN67F & CTN67R	59	60	90	120	35
CTN78F & CTN78R	59	60	90	120	35
CTN89F & CTN89R	55.8	60	90	120	35
CTN910F & CTN910R	No Single PCR Product Produced				
CTN910B-F & CTN910B-R	59	60	90	120	35
CTN1seq-F & CTN1seq-R	64	60	90	120	35
CTN1011F & CTN1011R	55	60	90	120	35
CTN1112F & CTN1112R	No PCR Product Produced				
CTN1213F & CTN1213R	No Single PCR Product Produced				
CTN1314F & CTN1314R	60.4	60	90	120	35
CTN1415F & CTN1415R	No Single PCR Product Produced				
CTN15F & CTN15R	61	60	90	120	40

Sequenceable PCR products were obtained for 9 of the 17 primer pairs that were designed for the *CTNNB1* gene (Table 4-9). In total, 2985bp of the *CTNNB1* gene sequence was obtained from the genome of *M. spilota*. This included 6 complete introns, 1 complete exon and the flanking regions of 10 other exons. DNA sequence data was obtained from at least 3 males and 3 females for every gene region sequenced, but in most cases the number of individuals sequenced was higher. No sex specific DNA was observed. One SINE repeat was discovered within intron 6 and a single SNP was found within intron 3. The BLAST results for all exonic DNA regions showed a sequence identity of 97% or higher with the predicted mRNA sequences of the *CTNNB1* gene from *P. bivittatus*. However, the DNA sequence representing terminal intron 15, only showed a sequence identity of 80%.

Table 4-9 CTNNB1 DNA sequencing results

***M. spilota* Contigs**	**Length (bp)**	**RepeatMasker**	**Genscan**	**SNPs/Repeats**	**nBLAST Match**	**Identities**
Contig #1 - Intron 3	463	No Repeats	No Result	SNP - (11, A/G)	PREDICTED: *P. bivittatus CTNNB1* mRNA	167/171(98%) 214/221(97%)
Contig #2 - Intron 6	630	SINE - (226-315)	No Result	SNP - (16, C/T)	PREDICTED: *P. bivittatus CTNNB1* mRNA	66/66(100%) 120/123(98%)
Contig #3 - Intron 9	698	No Repeats	No Result	None	PREDICTED: *P. bivittatus CTNNB1* mRNA	147/148(99%) 340/342(99%)
Contig #4 - Intron 13 and Intron 14	773	No Repeats	Intron - (149-270)	None	*P. bivittatus CTNNB1* partial cds	757/777(97%)
Contig #5 - Intron 15	421	No Repeats	No Result	None	*Elaphe quadrivirgata CTNNB1* mRNA	332/414(80%)
Total	**2985**					

Using the *P. bivittatus* genomic data available when this research was originally conducted, only 5 previously unidentified DNA sequence contigs encoding the *CTNNB1* gene from the *P. bivittatus* genome were identified (Castoe, de Koning, et al. 2011). This *P. bivittatus* DNA sequence data, published in 2011, allowed us to interpret some of the gaps between our *M. spilota* sequences (Table 4-10). The identified *M. spilota* sequence contigs and the sequence contigs identified from *P. bivittatus* were aligned to the annotated sequence of the *CTNNB1* gene from *A. carolinensis*, the most closely related complete *CTNNB1* sequence available at that time (Figure 4-4).

Table 4-10 CTNNB1 M. spilota and P. bivittatus contig assembly

<table>
<tr><th>M. spilota Contigs</th><th>P. bivittatus Contigs</th><th>Identity</th><th>Combined length (bp)</th><th>Start of Contig</th><th>End of Contig</th></tr>
<tr><td>Contig #1 - Intron 3</td><td rowspan="2">Contig #1 - AEQU010248189.1</td><td>452/462 (97%)</td><td>1768</td><td rowspan="2">Intron 2</td><td rowspan="2">Intron 7</td></tr>
<tr><td>Contig #2 - Intron 6</td><td>539/593 (91%)</td><td></td></tr>
<tr><td>Contig #3 - Intron 9</td><td>Contig #2 - AEQU010250816.1</td><td>686/698 (98%)</td><td>1782</td><td>Intron 8</td><td>Intron 10</td></tr>
<tr><td>Contig #4 - Intron 13 and Intron 14</td><td rowspan="2">Contig #3 - AEQU1011196067.1 AEQU010994310.1 AEQU010351695.1</td><td>759/779 (97%)</td><td>4757</td><td rowspan="2">Intron 12</td><td rowspan="2">Past terminal intron 15 into ULK4 Gene</td></tr>
<tr><td>Contig #5 - Intron 15</td><td>389/421 (92%)</td><td></td></tr>
<tr><td>Total</td><td></td><td>2825/2953 (97%)</td><td>8307</td><td></td><td></td></tr>
</table>

Figure 4-4 CTNNB1 sequence contigs from M. spilota and P. bivittatus, using an A. carolinensis reference assembly

Figure 4-4 depicts M. spilota sequence contigs (blue) and the sequence contigs identified from P. bivittatus (yellow), aligned to the annotated sequence of the CTNNB1 gene from A. carolinensis (green), with the exons from the gene (red).

With a more recent *P. bivittatus* genome assembly published in 2013, the assembly of all *M. spilota CTNNB1* contigs was possible (Castoe et al. 2013). An investigation of the similarities and differences between the genomic DNA sequence of *M. spilota* and *P. bivittatus* is shown in Table 4-11. The *M. spilota* genomic DNA representing the *CTNNB1* gene shares a combined sequence identity of 95.6% with the aligned *P. bivittatus* scaffold 1101 (NW_006533115). Within the coding regions of this gene there was only one single nucleotide difference observed between the two species, a nucleotide substitution located within exon 8. Differences between the DNA sequences of the two snakes were more commonly observed within the intronic DNA regions. The most common of these were single nucleotide substitutions and small DNA insertions or deletions no longer than 5bp in length.

There was one large difference observed between the DNA sequences of *M. spilota* and *P. bivittatus*, in the form of a 36bp indel located in intron 5 of *P. bivittatus* but absent from *M. spilota*. An investigation of this 36bp region in the *CTNNB1* orthologues for both *A. carolinensis* and *G. gallus* showed that the region was not present. Furthermore, a sequence alignment of this DNA region also showed the flanking DNA sequences to be highly divergent between the chicken, lizard and the two-snake species.

Table 4-11 CTNNB1 M. spilota and P. bivittatus genomic DNA sequence alignment

***M. spilota* Contigs**	**Length (bp)**	**Start of Contig**	**End of Contig**	**Sequence Difference Between M. spilota and P. bivittatus**	**Identical Sites**
Contig #1	463	Exon 2	Exon 3	6 Substitutions	451/463 (97.4%)
Contig #2	668	Exon 5	Exon 6	13 Substitutions Deletion 36bp Insertion 4bp Deletion 2bp	611/668 (91.5%)
Contig #3	698	Exon 8	Exon 9	9 Substitutions Insertion 3bp	685/698 (98.1%)
Contig #4	773	Exon 11	Intron 12	16 Substitutions Deletion 1bp	757/773 (97.4%)
Contig #5	421	Terminal Exon 14	Terminal Intron 15	24 Substitutions Deletion 5bp	387/421 (91.9%)
Total	**3023**				**2891/3023 (95.6%)**

BLAST analysis of the *M. spilota* contigs showed that contig 4 aligned for approximate 150bp with the DNA sequences from a study investigating the *CTNNB1* gene in 12 different snake species (Laopichienpong et al. 2017). A multiple sequences alignment of the sequences shows contig 4 from *M. spilota* is most like the DNA sequences produced from the snake species belonging to the families Cylindrophiidae, Xenopeltidae, and Pythonidae (Figure 4-5). As well as aligning well with the Z chromosomes sequences produced from more distantly relates snake species belonging to the families Viperidae, Elapidae, and Colubridae.

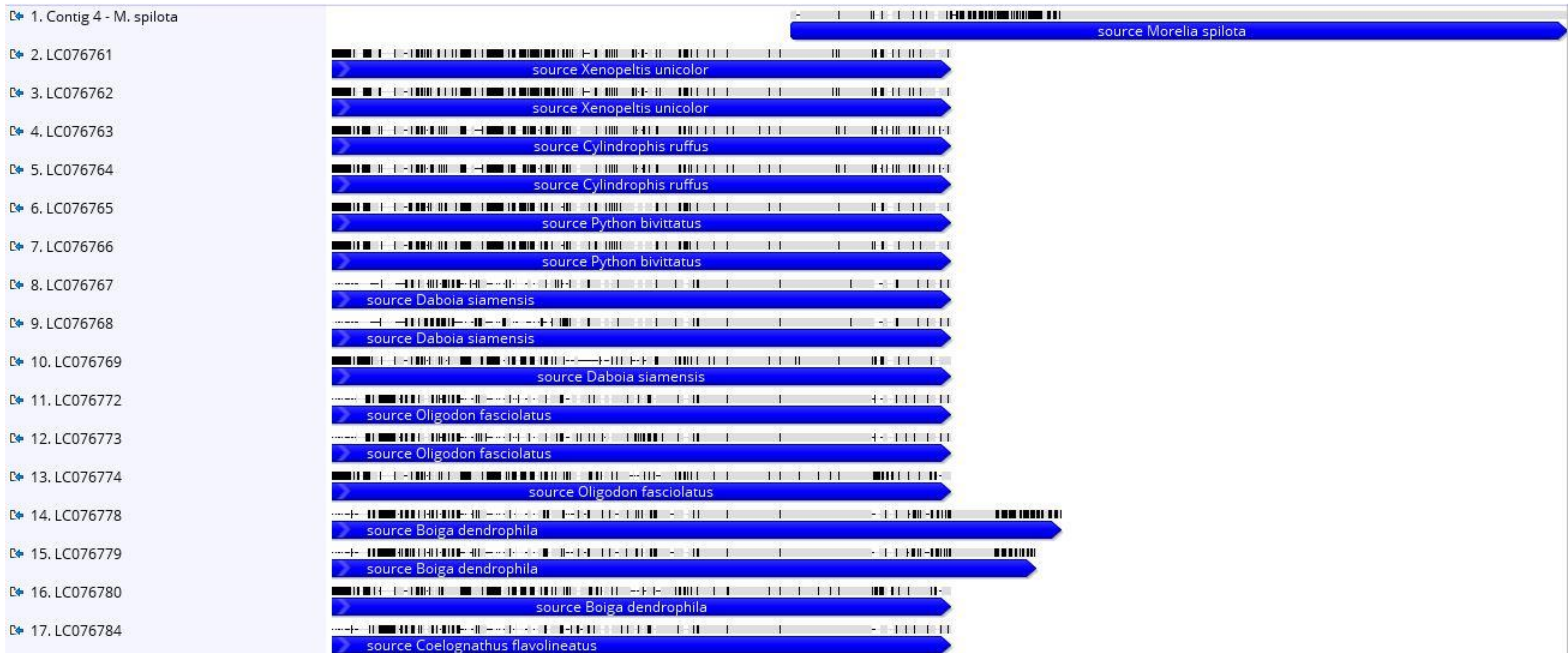

Figure 4-5 Multiple sequences alignment, contig 4 and CTNNB1 sequences

Figure 4-5 shows a multiple sequences alignment of the CTNNB1 gene – contig 4 from M. spilota, with 16 DNA sequences from the 7 different snakes species (Laopichienpong et al. 2017)*. The alignment shows contig 4 aligning to approximately 150bp of the DNA sequences and is most closely related to CTNNB1 gene sequences belonging to evolutionary older snake lineages.*

***RAB5A*, Member RAS Oncogene Family (*RAB5A*) Gene**

The primary PCR of the gene *RAB5A* produced two PCR products. A relatively small PCR product approximately 400bp in length and a large fragment approximately 2000bp in length. The two fragments were separated by molecular weight using agarose gel electrophoresis and each fragment extracted separately using a gel extraction kit. This was performed on two individuals representing one male and one female and the extracted PCR fragments were DNA sequenced. The sequencing results showed that the smaller DNA fragment was 338bp in length, with no observable differences identified between the two sexes and contained no repetitive elements. Using BLAST, the DNA sequence was shown to match a predicted mRNA sequence representing the gene MIER family member 2 (*MIER2*) in *P. bivittatus*, with a sequence identity of 99% across 316bp. These results showed that the smaller PCR product was not representative of the gene target *RAB5A* and thus further investigation of this PCR fragment was discontinued.

Sequencing results of the larger PCR fragment revealed that it was 2208bp in length. The sequence contained a mononucleotide microsatellite repeat spaning 9 cytosine bases and a SINE repeat 68bp in length. BLAST results revealed multiple matches for a 45bp region of the sequence, aligning with a predicted mRNA sequence from *P. bivittatus*, representing one transcriptional variant of the *RAB5A* gene (XM_025165194.1). This region showed 98% sequence identity.

In the one male and one female DNA sequence, a polymorphism appeared to be present, which was potentially sex-linked. It was located 311bp from the beginning of sequences and consisted of a 9-cytosine mononucleotide microsatellite. At this sequence position, sequencing result using the forward primer were obtained only, due to the PCR fragment's length. The male sequence clearly showed no polymorphism within the microsatellite ending after 9bp (Figure 4-5). However, after the 9 cytosine bases in the female sequence, the sequence signal halved and appeared to produce a mixed sequence signal.

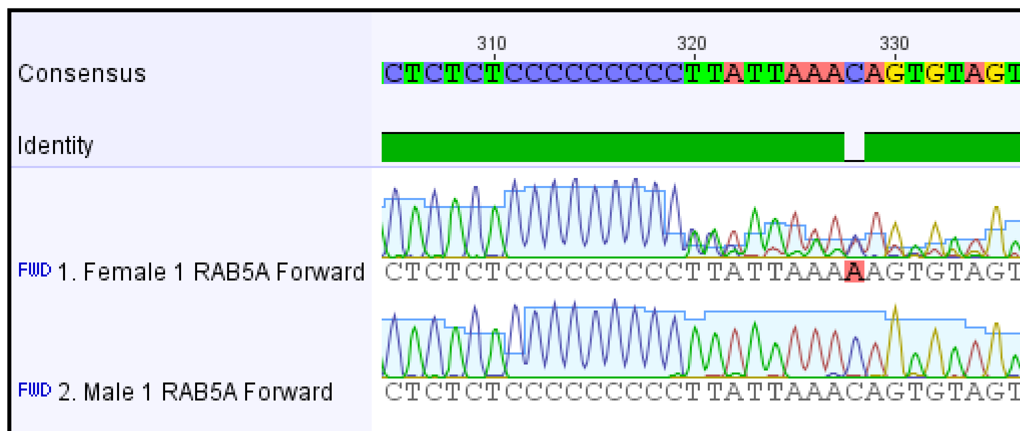

Figure 4-6 RAB5A mononucleotide microsatellite polymorphism, one male one female

Figure 4-5 shows two chromatograms, one male and one female for the potentially sex-linked microsatellite region identified in the RAB5A sequence.

Further investigation of this region was performed after designing a new set of PCR primers flanking either end of the polymorphic region; primers *RAB5A* micro F and *RAB5A* micro R as defined in Table 4-4. A larger cohort of individuals was sequenced, bringing the total to 4 males and 5 females (Figure 4-6). All sequence data produced using the reverse primer resulted in undifferentiable sequence nucleotides, beyond the polymorphic DNA region, despite having a shorter PCR product and different PCR primer. Some sequencing results, using the forward primer, produced a single DNA template signal, indicating that the individual being tested was homozygous for the microsatellite, with sequences containing 9 cytosine bases. These sequences continued past this region with little change to the sequencing signal intensity. While other sequencing results appear to be heterozygous at this position. For all sequences 9 cytosine bases were present, but in 5 out of the 9 animals tested at the 10th position a mutation resulted in a mixed template sequence signal (Figure 4-6). The results produced using a larger sample population showed that this microsatellite was unlikely to be sex-linked.

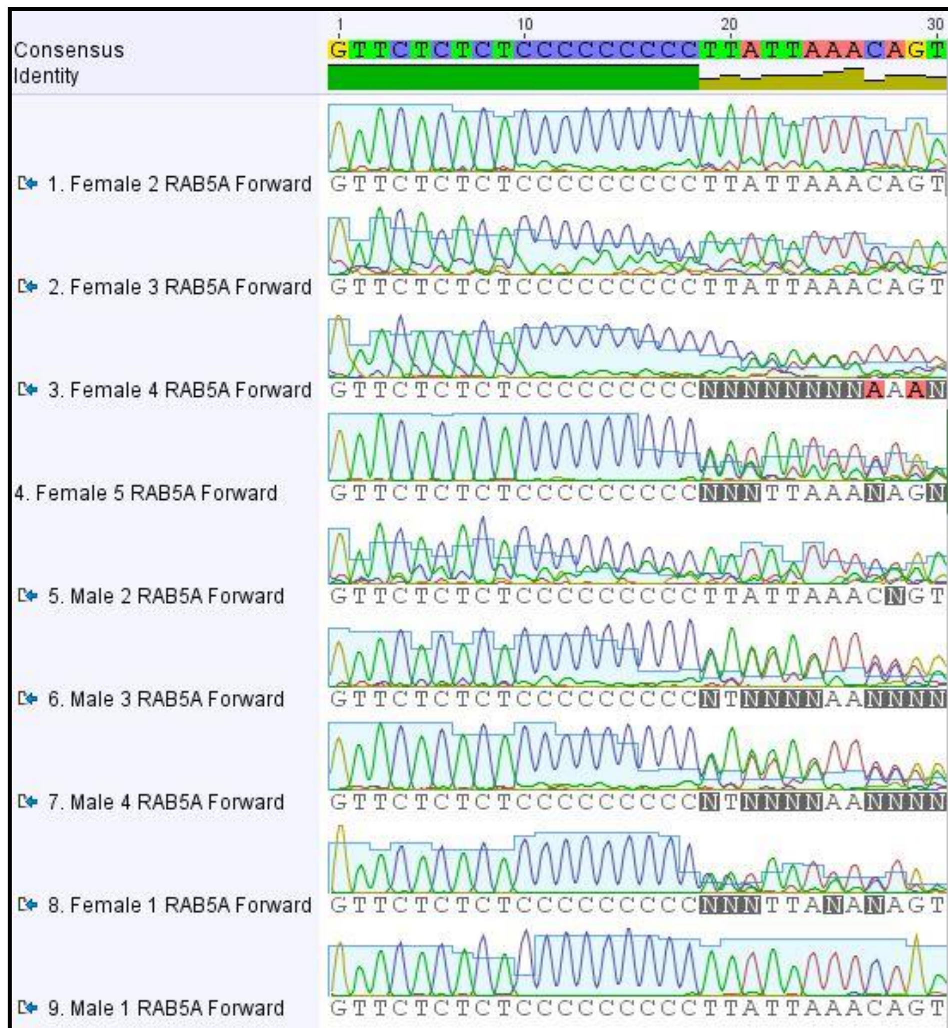

Figure 4-7 RAB5A mononucleotide microsatellite polymorphism, 7 animals

Figure 4-6 shows a chromatogram of four male and five females for the microsatellite region identified in the RAB5A sequence.

This polymorphic microsatellite DNA region was investigated further by using real-time PCR and melt curve analysis. The aim of this analysis was to differentiate male and female genomic DNA templates using melt-curve analysis of the PCR product produced from the *RAB5A* micro F and *RAB5A* micro R PCR. All 23 individuals of previously determined sex were tested. The results showed that the polymorphic microsatellite was not sex-linked. In general, female PCR samples started melting at a slightly higher temperature than males and generally males produced a more diverse range of melt curves within the population tested. However, from the results obtained, there was no obvious differentiation between male and female PCRs melt curves indicating that no sex associated exists for this DNA region.

Individual	Sequence	
Female 2	Record	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
Female 3	Record	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
Female 4	Record	G T T C T C T C T C C C C C C C C N N N N N N N N A A N N
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C C C C C T T A T T A A A
Female 5	Record	G T T C T C T C T C C C C C C C C N N N T T A A A N A G N
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C C C C T T A T T A A A C
Male 2	Record	G T T C T C T C T C C C C C C C C T T A T T A A A C N G T
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
Male 3	Record	G T T C T C T C T C C C C C C C C N T N N N N A A N N N N
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C C T T A T T A A A C A G
Male 4	Record	G T T C T C T C T C C C C C C C C N T N N N N A A N N N N
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C C T T A T T A A A C A G
Female 1	Record	G T T C T C T C T C C C C C C C C N N N T T A N A N A G T
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C C C C T T A T T A A A C
Male 1	Record	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 1	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T
	Copy 2	G T T C T C T C T C C C C C C C C T T A T T A A A C A G T

Figure 4-8 Representation polymorphic microsatellite sequences composition

Figure 4-7 is a representation of the sequence data produced from four males and five females sequenced for the microsatellite region of the RAB5A gene. The recorded sequence is the DNA nucleotides that were clearly represented in the chromatogram. While the two copies of each sequence are the most likely DNA sequences represented on each of the two chromosome copies in the genome.

Investigation of the sequencing chromatograms of the region, produced for the 9 individuals, showed the DNA microsatellite in most likely polymorphic. Using the information from each chromatogram, the most likely number of cytosine bases in both copies of the microsatellite for each animal sequenced was extrapolated (Figure 4-7). Interestingly, in all individuals sequenced at least one copy of the microsatellite region contained 9 cytosine bases.

A consensus sequence was created using the 30 DNA sequences obtained for the *M. spilota RAB5A* gene and this consensus sequence was used to search the *P. bivittatus* genome. From the original 2011 published *P. bivittatus* genome, two *P. bivittatus* contigs mapped to the *RAB5A* consensus sequence. Analysis of these two contigs showed that the *M. spilota RAB5A* sequence matched either end of the two *P. bivittatus* contigs, bridging the gap between them. With sequences matching 181bp and 120bp representing two exons either side of the targeted *RAB5A* intronic region.

Using the more recent 2013 published *P. bivittatus* Annotation Release 101 the *RAB5A* consensus DNA sequence matches scaffold 618 with 93% sequence identity across the query sequence's entire length. This included 24bp representing exon 3, 37bp of exon 4 and the entire 1905bp intron between the two exons of the *RAB5A* gene in *P. bivittatus*. The short exon regions appear to be highly conserved with no differences observed between the two snake sequences. However, the intronic region showed significant differences between the two-snake species. Approximately 30 SNPs were identified within the intronic region, as well as a series of short DNA deletions and insertions between the two-snake species. A large 163bp insertion in the DNA sequence of *M. spilota* was also identified. This insertion was not identified in in the NCBI's database using BLAST or RepeatMasker. However, within the NCBI *P. bivittatus* Annotation Release 101, 172 difference genomic scaffolds were identified that contained a similar DNA region. This may indicate that this region is a part of a previously undefined repetitive element that is present in the genomes of both *M. spilota* and *P. bivittatus*.

Discussion

When extracting the DNA from mammalian blood, most methods target white blood cells containing genomic DNA, discarding the erythrocytes. In reptiles, including snakes, all blood cells are nucleated and hence can also be used for genomic DNA extraction. This leads to an approximately 200-fold increase in the yield of DNA extracted compared with a comparable amount of mammalian blood. This advantage allowed the use a significantly smaller amount of blood to be sampled from each animal, ensuring the animal's wellbeing.

The Qiagen DNeasy Blood & Tissue Kit for DNA extract was used to extract the snake DNA from whole blood preserved in the anticoagulant EDTA. When using the recommended amount of 5 µl of whole blood, for a species with nucleated erythrocytes such as snakes, complete red blood cell lysis was unable to be achieved. Furthermore, initially these DNA extractions contained relatively low-quality DNA contaminated with protein and had poor DNA yields. It was observed that by reducing the starting material from 5µl to 2µl of whole blood these problems were rectified.

Comparative Genomics

Using the cDNA sequences generated from *E. quadrivirgata* as our reference, a comparative genomic approach was employed, which resulted in PCR primers and DNA sequenced PCR products, representing genomic DNA of the genes *ATP6V0A1*, *CTNNB1*, *EIF1*, *GAD2*, *GH1*, *MRPL3*, *PIP4K2A*, *PTER*, *RAB5A*, *RUNDC1*, *SRI*, *TAX1BP1* and *WAC* (Matsubara et al. 2006, 2012). Totalling 16,045bp from the *M. spilota* genome, representing the proposed Z and potentially the W chromosome. This is the first time that genomic DNA representing many of these genes has been sequence in a snake species and the first time any of these genes have been DNA sequenced in our model snake species *M. spilota*. Despite an extensive search no sex-linked DNA sequence were identified from the genomic DNA sequenced.

This comparative genomic approach was not successful for designing PCR primers specific for the three genes *CCDC126*, *CCDC127* and *MYO1D*. This outcome is not surprising considering these three genes showed the highest degree of sequence variability between the mRNA and cDNA sequences investigated in the model vertebrate species selected. It was found to be a difficult task to design specific PCR primers for these three genes, because there were limited regions of conservation between the different vertebrate species used in the multiple sequence alignment investigation. Therefore, the primer pairs designed to target these genes, did not satisfy all our primer specifications, and as a result failed to amplify DNA from *M. spilota*.

The DNA sequence data produced from the PCR products targeting the three genes *KAT7*, *KLF6* and *TUBG1* was unable to be determined. The PCR optimisation process for these three genes was more difficult than most and the sequencing results produced mixed sequencing signals. For all three genes this result was consistent for both male and female samples. Attempts to eliminate this problem through cloning of the PCR products and sequencing

individual plasmid clones, also did not produce consistent sequencing results. With most individual clones sequenced producing mixed sequencing signals and all usable sequencing data not appearing to correspond to the gene being targeted. It is hypothesised that the reason for this failure is the lack of prior sequence knowledge, making design of PCR primers specific for these three genes unreliable. It is suspected that the PCR-primers, designed to target these three genes, are binding to and amplifying multiple region, possibly repetitive elements, within the genome of *M. spilota*.

Initially the DNA sequences produced from targeting the genes *PIP4K2A* and *TAX1BP1* did not correspond to the correct target according to NCBI's BLAST database. This was because the comparative genomic approach involved using the conserved region of two exons, to design primers either side of an intronic region. This process results in a large proportion of the final DNA sequence being intronic. This was particularly true for the PCR products produced when targeting the two genes *PIP4K2A* and *TAX1BP1*. BLAST searching the NCBI *P. bivittatus* Annotation Release 101 resulted in a match of the *PIP4K2A* PCR generated gene sequence within an incomplete assembly of the same gene in the *P. bivittatus* genome. A match with the *TAX1BP1* gene sequence to the JAZF1 gene in the *P. bivittatus* genome was then used to infer that the PCR sequence matched a *TAX1BP1* transcript variant in *P. mucrosquamatus* as well as other reptile species.

The DNA sequence produced from targeting the gene *AMPH* could only be matched to a non-coding DNA region within the *P. bivittatus* genome assembly. This result indicates that the DNA sequence produced from targeting the gene *AMPH* most likely does not correspond to the correct gene target. These results show again that lack of genetic knowledge about our gene targets in snakes, makes initial PCR amplifying and subsequent DNA sequencing difficult. It also means that identifying sequences derived from this process that match our gene targets, is just as difficult.

CTNNB1, Catenin Beta 1

The vertebrate mRNA sequence alignment of the *CTNNB1* gene indicated that the expressed regions are well conserved amongst distantly related vertebrate species, except for the first and last exons. It was shown that exon number one, which is approximately 100bp in length and approximately 12,000bp away from the next nearest coding region, showed high inter species sequence variability. Due to its non-conserved nature and distance from exon 2 no

PCR primer specific for this region was able to be designed. Similarly, the terminal exon of the *CTNNB1* gene, is also not well conserved amongst the vertebrate species investigated. However, in this case cDNA information from *E. quadrivirgata* was also available, which allowed the design of primer pairs specific to this region that successfully amplified *M. spilota* DNA. In total, 3023bp from the *CTNNB1* gene of *M. spilota* was sequenced. Assuming the gene is approximately 21,000bp in total length, as is described in *G. gallus*, approximately 15% of the genomic DNA representing the *CTNNB1* gene in *M. spilota* was sequenced, without any prior genomic DNA sequence knowledge from any snake species (NCBI Resource Coordinators 2017).

No sex specific DNA regions within the *CTNNB1* gene of *M. spilota* were identified. However, two single nucleotides polymorphisms that occurred in at least two or more individuals and a SINE repetitive element within intron 6 were identified. Comparing the DNA sequences from the *CTNNB1* gene of *M. spilota* and *P. bivittatus* a 36bp region was found; present in *P. bivittatus* but absent from *M.* spilota within intron 5. This 36bp region was unable to be identified in *A. carolinensis* or *G. gallus*. This indicates that the 36bp region was most likely inserted into intron 5 of *CTNNB1* gene in *P. bivittatus*. This evidence shows that the *CTNNB1* gene has diverged from its ancestral sequence in *P. bivittatus*, sometime after the two species of python *M. spilota* and *P. bivittatus* have diverged. This DNA insertion is significant enough that it could be used to discriminate between DNA samples of the two species *M. spilota* and *P. bivittatus*, using molecular genetic techniques.

The most recent publication involving snake sex determination and the *CTNNB1* gene was published in 2017 and investigated the genetic differences between a small genomic DNA region of the *CTNNB1* gene, from both the proposed Z and W sex chromosome (Laopichienpong et al. 2017). Their analysis was performed on 12 snake species belonging to Henophidia snakes (Cylindrophiidae, Xenopeltidae, and Pythonidae) an older snake lineage with unidentified sex chromosomes or belonging to the HSU group and Caenophidia snakes (Viperidae, Elapidae, and Colubridae) a more modern snake lineage with identifiable female heterogametic sex chromosomes or the HFI group (Laopichienpong et al. 2017).

Analysis of the DNA sequences showed contig number 4 of the *M. spilota* DNA contigs produced, aligned with 150bp of the DNA sequences investigated in the published article (Laopichienpong et al. 2017). Our analysis showed that the DNA sequences produced from

M. spilota were most like the DNA sequences produced from the most closely related snake species belonging to the snake families Cylindrophiidae, Xenopeltidae, and Pythonidae, as would be expected. The evidence presented in this publication led the authors to predict "that *CTNNB1Z* and *CTNNB1W* sequences diverged independently in each lineage of Caenophidia after divergence from Henophidia around 100 MYA during the early Cretaceous era" (Laopichienpong et al. 2017).

The *CTNNB1* genomic DNA sequences and the PCR primers produced in our experiment could be used to advance the investigated of the 2017 published article by investigating more and larger regions of the *CTNNB1* gene (Laopichienpong et al. 2017). This information may give better insight into how the *CTNNB1* gene and the proposed Z and W chromosomes of snakes, have evolved in more closely related snake lineages and better define how the sex chromosomes have differentiated to different degrees in different lineages. We suggest that future investigations of the *CTNNB1* gene should focus on both the conserved exonic regions of the gene and the highly variable regions of the gene. Terminal exon 15, which we found to be highly variable between vertebrates investigated. As well as intron 5, which we showed to have a 36bp indel between the two closely related python species *M. spilota* and *P. bivittatus*.

Conclusion

When this project was started in 2010, the only DNA sequence information available for the snake Z or W chromosome was work described by Kazumi Matsubara research group's in 2006 (Matsubara et al. 2006). In this publication, the research group mapped 11 genes to the Z chromosomes of *E. quadrivirgata*. The present work commenced with using the cDNA sequences from these 11 genes that have been localised to the Z chromosome to identify these genes in another snake species. In 2012, Matsubara research group published a second journal article further mapping 9 additional genes to the Z chromosome of *E. quadrivirgata*, taking the total to 20 genes (Matsubara et al. 2012). This information was used to target the additional 9 genes in the same way as the first 11.

The laboratory component of the present research was completed in 2012. At that time, we were able to identify less than half of the DNA sequences produced from targeting these 20 genes using PCR. One issue is that snake and reptile DNA sequences in general within the NCBI's nucleotide database are relatively rare. Moreover, the comparative genomic

approach we used results in the sequences produced being made up largely of intronic DNA, further compounding the problem. In addition, in 2012, all the DNA sequences produced, except for *CTNNB1* sequences*,* were the only examples of genomic DNA sequences for these genes in a snake species. These three factors in context of knowledge and technology in 2012 resulted in less than half of the DNA sequenced representing the actual gene targets.

In 2013, NCBI *P. bivittatus* Annotation Release 101 was released. This genome assembly is the most comprehensive published genome of any snake species assembled to date and for the first time provided enough snake sequence information for cross species analysis. From this information most of the DNA sequences produced could be identified, and it was shown that these DNA sequences were from the gene regions being targeted.

Chapter 5 - Cosmid Library Construction and Second Generation Sequencing

Introduction

A cosmid is a hybrid vector that contains four distinct DNA regions: an origin of replication, which allows the cosmid DNA to be replicated independently of the rest of the host bacterial genome; one or more antibiotic resistance genes, which allows for identification of bacterial cells that contain a cosmid vector; a multiple cloning site made up of multiple endonuclease recognition sites that is used as a targeted point of DNA insertion; two cos sequences from lambda bacteriophage (Hohn and Murray 1977; Collins and Hohn 1978). The addition of two lambda bacteriophage cos sequences allows bacterial cells to be successfully transformed with cosmids of much larger total size than normally possible for a plasmid based vector and therefore contain a larger DNA insert (Hohn and Murray 1977; Collins and Hohn 1978). This is because these cos sequences are recognised and used by phage proteins allowing packaging of the cosmid containing a DNA insert into a phage capsid during a cosmid packaging reaction (Hohn and Murray 1977; Collins and Hohn 1978). Packaged cosmid and DNA insert can therefore be inserted into the host bacterial cell by transduction, rather than normal plasmid vector transformation (Hohn and Murray 1977; Collins and Hohn 1978).

This chapter describes the construction of two large genomic DNA insert cosmid based libraries, representing the genome of one male and one female *Morelia spilota.* The two libraries were constructed, with the aim of using them as a tool to investigate genes previously mapped to the sex chromosomes of snakes and for identifying potential sex-linked DNA regions described in the previous chapters within this thesis. For this experiment, the cosmid vector SuperCos 1 (Agilent Technologies) was used and the map is shown in Figure 5-1 (Evans, Lewis, and Rothenberg 1989).

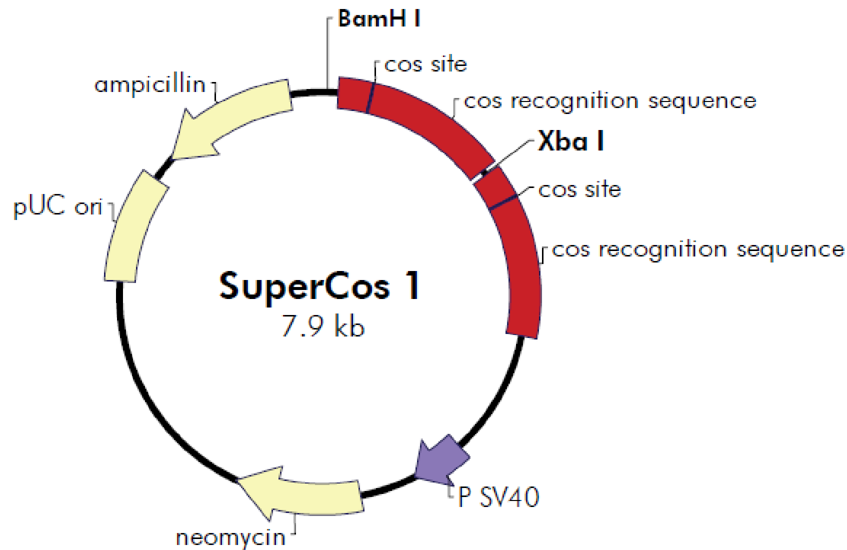

Figure 5-1 SuperCos 1 vector (Agilent Technologies)

Once the libraries were constructed and cells transformed, the libraries were targeted to identify, isolate, and analyse DNA regions that have been identified from our previous experiments, particularly the DNA regions identified in Chapter 4. In Chapter 4, DNA sequences from regions representing 13 genes that have previously been mapped to the Z chromosome of *Elaphe quadrivirgata* were obtained (Matsubara et al. 2006, 2012). The information obtained was then used to isolate cosmid clones representing primarily 3 of the 13 gene regions studied; Catenin Beta 1 (*CTNNB1*), RAS-Associated Protein *RAB5A*, Member RAS Oncogene Family (*RAB5A*) and WW Domain Containing Adaptor with Coiled-Coil (*WAC*). Facilitating an investigation into the genomic DNA sequences of these three genes, together with non-coding DNA sequences surrounding and within these genes. Using a male and a female genomic DNA library also allowed the investigation of potential sex specific DNA regions. With the female library representing the proposed chromosome Z and W complement and the male library representing two copies of the proposed Z chromosome (Matsubara et al. 2006, 2012).

Material and Methods

SuperCos 1 Cosmid Vector Kit from the manufacturer Agilent Technologies was used to construct two cosmid libraries, representing the genomes of one male and one female *M. spilota* respectively. Blood stored in EDTA at 4°C was extracted using a modified version of the phenol/chloroform DNA extraction protocol provided by the manufacturer in the instruction manual titled; SuperCos 1 Cosmid Vector Kit, Instruction Manual, Catalogue #251301, Revision B.

Two hundred microliters of whole-blood was pipetted into a 50 ml centrifuge tube containing 15 ml of lysis buffer (10 mM NaCl, 20 mM Tris.HCl (pH 8.0), 1 mM EDTA, 0.5% (w/v) SDS) and 100 µg/ml of proteinase K enzyme (Promega). The sample was very gently mixed through slow inversion of the tube and then the centrifuge tube was incubated in a heat block at 55°C for 24 hours and mixed by gentle inversion three times throughout the digestion process.

Upon completion of digestion (no red clumps being visible) the centrifuge tube was cooled to room temperature for approximately 20 minutes. Fifteen millilitres of 50:50 phenol/chloroform (Ajax) was added to the sample and the tube gentle inverted for approximately 5 minutes. The sample was centrifuged at 3000 g for 30 minutes. The upper aqueous layer was carefully removed using a 1 ml wide-bore pipette tip, being careful not to disturb or collect any of the interphase or lower layers. The wide-bore pipette tips were previously prepared by cutting the ends off regular 1 ml tips with scissors and sterilising through autoclaving. This process of phenol/chloroform addition, centrifugation, and aqueous phase extraction was repeated a further two times, totalling three phenol/chloroform separations.

Sodium acetate (3 M) was added to the third and final aqueous layer separation, bringing the final concentration to 0.3 M sodium acetate and the tube gently mixed. Two times the volume of ice cold 100% isopropanol was added and the tube gently mixed. The reaction was then incubated at -20°C for 24 hours.

The DNA was pelleted by centrifuging at 3000 g for 30 minutes at 4°C. The isopropanol/sodium acetate solution was decanted, and then 30 ml of room temperature 70% ethanol solution was added. The tube was incubated at room temperature for 15

minutes, before being centrifuged at 3000 g for 30 minutes at 4°C to pellet the DNA, followed by decanting of the ethanol. The 70% ethanol wash was repeated once more. The ethanol was decanted, and pellet dried at room temperature by placing upside down on a clean laboratory bench with the lid ajar. The DNA pellet was resuspended in 5 ml of TE buffer and stored at 4°C for at least 48 hours, to fully redissolve the DNA. The quality and quantity of the extracted DNA was evaluated using agarose gel electrophoresis, 0.7% (w/v) agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer and quantified using a NanoDrop spectrophotometer.

Genomic DNA preparation

The genomic DNA was required to be partially digested with the restriction endonuclease Sau3A I to a size range of between 30 to 42 kb in length. An initial experiment was used to determine the optimal final reaction conditions. The reaction contained 10 µg of genomic DNA, 10 µl of 10× reaction buffer, 0.5 units of Sau3A I (Promega), 1 µl of 100 µg/ml bovine serum albumin (BSA), in a final volume of 100 µl. The reaction was incubated at 37°C and at 5-minute intervals (0, 5, 10, 15, 20, 25, 30, 35, 40, 45 minutes), 10 µl of the reaction was removed and placed into a microfuge tube containing 10 µl of 6× agarose gel loading buffer.

The aliquots were analysed using agarose gel electrophoresis, 0.7% (w/v) agarose containing in TAE buffer stained with 0.5 µg/ml ethidium bromide in TAE buffer and the average DNA fragment size compared to controls of 1 µg of lambda phage DNA (Promega) and 1 µg lambda/Hind III DNA digest. The reaction conditions required to achieve the desired digested DNA size range of between 30 to 42 kb was; 100 µg of genomic DNA, 100 µl of 10× restriction buffer, 5 units of Sau3A I (Promega), 10 µl of 100 µg/ml BSA in a final volume of 1 ml and incubated for 30 minutes at 37°C. The reaction was stopped by heat inactivation for 30 minutes at 65°C. The DNA digestion was visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer, to confirm fragments of the desired size range.

The digested DNA was dephosphorylated, using 5 units of 1 U/µl Thermosensitive Alkaline Phosphatase (TSAP; Promega) added directly to the reaction tube. The Sau3A I reaction buffer is compatible with the TSAP and therefore no buffer change was required. The reaction was incubated at 37°C for 1 hour and the TSAP was then heat inactivated for 20 minutes at 65°C.

To determine if the digested genomic DNA had been dephosphorylated a test ligation reaction was performed. One microgram of dephosphorylated DNA was combined with 0.5 units of T4 DNA ligase (Promega), 1.5 µl of 10× ligase buffer, made to a total volume of 15 µl with high purity water, was then incubated at 16°C for 1 hour. The test ligation was visualised by agarose gel electrophoresis, using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE, along with 1 µg of the digested non-ligated genomic DNA control. No differences were observed between the two samples, indicating that the dephosphorylation was successful.

The dephosphorylated genomic DNA was precipitated by adding 3M sodium acetate to a final concentration of 0.3 M and then 2.5 volumes of ice-cold 100% ethanol. The DNA was pelleted by centrifugation at 3000 g for 30 minutes and the DNA pellet rinsed twice with 1 volume of 70% ethanol and allowed to air dry. The final DNA pellet was re-suspended in 50 µl of high purity water and the quality determined using agarose gel electrophoresis, using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE, and a NanoDrop spectrophotometer.

SuperCos1 Vector Preparation

Twenty-five micrograms of SuperCos 1 vector was digested with 10 units of the restriction endonuclease Xba I (Promega), 20 µl of 10× reaction buffer, 2 µl of 100 µg/ml BSA, in a total volume of 200 µl. The reaction was incubated at 37°C for 1 hour to ensure complete digestion of the vector. The reaction was heat inactivated for 30 minutes at 65°C. Ten microliters of the digestion reaction was visually assessed by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer. A single vector band at 8 kb was observed indicating complete digestion of the vector.

The remaining 190 µl of digested vector was dephosphorylated by adding 1 unit of TSAP, to the reaction tube and incubating for 1 hour at 37°C, followed by heat inactivation for 20 minutes at 65°C. To determine if the digested SuperCos 1 vector DNA had been successfully dephosphorylated a test ligation was performed. One microgram of dephosphorylated vector DNA (8 µl), 0.1 units of T4 DNA ligase (Promega), 1.5 µl of 10× ligase buffer, in a total volume of 15 µl was added to a microfuge tube and incubated for 1 hour at 16°C. The

ligation reaction was visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer, alongside an equivalent amount of digested, dephosphorylated vector DNA. No difference was seen between the two samples indicating that the dephosphorylation was complete.

The remaining 182 µl of SuperCos 1 vector was then again digested, by adding 10 units of the restriction endonuclease Bam HI (Promega) to the compatible buffered reaction. The reaction was incubated for 1 hour at 37°C. Ten microliters of the double digested vector was assessed visually by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer. The result was the presence of two expected vector bands at 1 kb and 6.5 kb and the absence of a band at 8 kb, indicating complete digestion. The concentration of the final product was determined by NanoDrop spectrophotometry. The double digested SuperCos 1 reaction was then ethanol precipitated as described previously.

Ligation and Packaging of Genomic DNA

The partially digested and dephosphorylated genomic DNA was ligated to the prepared vector DNA. The ligation reaction was performed using; 2.5 µg of prepared genomic DNA, 1 µg of prepared SuperCos 1 vector DNA, 2.0 µl of 10× ligase buffer and 1 µl of 10 U/µl T4 DNA ligase (Promega) in a final volume of 20 µl. The ligation reaction was prepared on ice and incubated overnight at 4°C. One microliter of the ligation reactions was visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer, along-side prepared vector DNA. The result was an almost complete absence of the two 1 kb and 6.5 kb vector bands. The concentration of the sample was determined by NanoDrop spectrophotometry and ethanol precipitated as previously described. The final pellet was resuspended in high purity water to a concentration of 0.5 µg/µl.

To package the prepared SuperCos I genomic library the Gigapack III Gold-4 packaging extract (Agilent Technology) was used. Two packaging extracts were recovered from storage (-80°C freezer) and placed on ice. One microgram of the DNA library was added to each packaging mix. The packaging reaction was held between the fingers to ensure it was completely thawed and then mixed gently, without creating bubbles. The sample was then briefly centrifuged and incubated at room temperature for 2 hours. Five-hundred microliters

of SM buffer was added to each packaging reaction. Twenty microliters of chloroform (Ajax) was added to the tube and the tube gentle mixed. The two layers were separated by centrifuging at 3000 g for 30 minutes and the supernatant removed carefully leaving any interface layer behind and stored at 4°C, ready for bacterial host cell infection.

Preparation of Host Strain

The host strain of *Escherichia coli* used was XL1-Blue MR strain. The XL1-Blue MR strain was revived by scrapping an ice sliver off the glycerol stock provided by the manufacturer (Agilent Technology) and plated onto a LB agar plate, which was then incubated at 37°C for 24 hours. One colony from the LB plate containing the revived host strain was picked and re-streaked onto another LB plate. The re-streaked LB plate was again incubated at 37°C overnight. One colony from the plate was picked and a 50 ml centrifuge tube containing 40 ml of LB broth containing 0.2% W/V maltose was inoculated. The inoculation was incubated at 37°C, with shaking for 4-6 hours. During this incubation period the cell concentration was measured by removing 1 ml of the growing culture and measuring its absorbance in a spectrophotometer at the wavelength 600 nm. This was done every hour, ensuring the bacterial cells had reached their log phase of growth before infection occurred. After 6 hours of growth the inoculation had reached an OD_{600} of 0.7. The inoculation was centrifuged at 3000 g for 10 min to pellet the cells. The supernatant was decanted from the 50 ml tube containing the cells and these were resuspended with 47.6 ml of sterile 10 mM $MgSO_4$ giving a calculated OD_{600} of 0.5.

Cosmid Library Titre, Amplification, and Storage

The phage titre of the packaging reaction was calculated by preparing two dilutions; a 1 in 10 and a 1 in 50 for each of the packaging reactions. In two separate tubes 25 µl of each dilution was added to 25 µl of the prepared host cells. The tubes were gently mixed and incubated for 30 minutes at room temperature. LB broth (200 µl) was added to each of the tubes and they were incubated for a further 60 min at 37°C, mixing the tubes by gentle inversion every 15 minutes. The tubes were centrifuged for 10 min at 3000 g, the supernatant removed, and the pellet of cells resuspended in 1 ml of LB broth. Using a disposable plastic sterile spreader, 10 µl, 50 µl, 100 µl and 200 µl of the two sets of dilutions were spread onto LBA plates containing 0.1 mg/ml ampicillin. The number of Colony Forming Units (CFUs) were counted for each dilution and determined to be approximately 4,200 CFU/µl for the undiluted packaging reaction.

From the estimated phage titre, it was determined that 12 µl of our library would give approximately 50,000 CFU. Eight one microfuge tubes each containing 12 µl of packaged DNA was added to 12 µl of freshly prepared host strain cells, at a calculated infection ratio of one phage per cell. The infection tubes were incubated at room temperature for 30 minutes. This was done for all packaged DNA. The two-packaging reactions yielded 81 cosmid amplification tubes for each male and female cosmid libraries.

After infection, 1 ml of LB broth was added to each of the microfuge tubes and further incubated, with shaking, for 1 hour at 37°C, allowing time for the antibiotic resistance genes to be expressed. The cells were then collected by centrifugation of the microfuge tubes at 3,000 g for 10 minutes. The supernatant was discarded and 200 µl of LB broth added to each tube. The cells in each tube were re-suspended using gentle pipetting of the LB broth up and down and 200 µl was spread out onto an LBA plate containing ampicillin (0.1 mg/ml), using a disposable plastic sterile spreader. The plates were incubated at 37°C overnight.

After overnight incubation, the colonies were clearly visible on the agar plates. Two millilitres of LB broth containing 0.1 mg/ml ampicillin was pipetted onto the agar plate surface. A disposable plastic sterile spreader was used to gentle rub the plate and detach and break up the colonies. The broth containing the colonies was then transferred into a 15 ml centrifuge tube and the process of flooding the plate and collecting the suspension was repeated to ensure efficient collection of all the bacterial colonies. A total of 81 plates were processed and 750 µl of each sample was aliquoted into two separate 1.5 ml microfuge tubes. To each tube 750 µl of 30% sterile glycerol was added and the tubes stored at -80°C. Thus, two copies of each pool of the 81 colony preparations were produced. The remaining suspension, containing approximately 1 ml of the bacterial suspension, was used for cosmid DNA extraction.

Genomic Library Assessment

One millilitre of each cosmid amplification was transferred into a new 1.5 ml microfuge tube ready for plasmid DNA extraction. The bacterial cells were pelleted by centrifuging at 15,000 g for 10 minutes and the cosmid DNA extracted using an alkaline lysis method (Feliciello and Chinali 1993). The purified cosmid DNA from each cosmid extraction was re-suspended in 1 ml of TE buffer. To assess the genomic depth of each cosmid library, the cosmid DNA

extracted from the 162 cosmid sub-populations for both the male and female cosmid libraries, were PCR amplified using primers specific for the genes Catenin Beta 1 (*CTNNB1*), RAS-Associated Protein *RAB5A*, Member RAS Oncogene Family (*RAB5A*) and WW Domain Containing Adaptor with Coiled-Coil (*WAC*; Table 5-1). These three genes were chosen because they have been previously mapped to the Z and W chromosome from the snake species *Python molurus* and *E. quadrivirgata* and to the W chromosome of *Trimeresurus flavoviridis* (Matsubara et al. 2006). Our previous investigations had already produced DNA sequences representative of these three genes in *M. spilota* for which PCR primers and optimised PCR reaction protocols specific for these gene regions were available.

Table 5-1 PCR primers of three reference genes

Target Genes	Primer Names	Primer Sequences
CTNNB1	CTN1B-F	AGGAAGGCATGGAAGGTCTTCTGGG
	CTN1B-R	AGGCCAGTGGGATGGTGGATGC
RAB5A	*RAB5A*B-F	AGTCCTAACATTGTAATAGCTTTA
	*RAB5A*B-R	ATTGCCATAAATATTTCATTTAC
WAC	*WAC*B-F	ACATAATGACAGAGACTACAGACTGGCC
	*WAC*B-R	GCTGATGATCTGACTGTAGAGAAAATGG

Three separate PCR reactions were performed using each of the three primer pairs, targeting the genes *CTNNB1*, *RAB5A* and *WAC*. One PCR reaction for each of the three gene regions being targeted was amplified using template DNA from each of the 162 (81 male and 81 female) cosmid amplification pools. The PCR reaction conditions were as follows; a total reaction volume of 10 µl, containing 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10× PCR buffer, 0.1 µM of each primer and 1 µl of template DNA (30 ng of cosmid pool DNA). The thermal profile for the PCR targeting the *CTNNB1* gene was 95°C for 10 min, followed by 40 cycles of: 95°C for 30 s, annealing temperature of 55°C for 60 s, 72°C for 120 s, final extension at 72°C for 10 min. The thermal profile for the PCR targeting the *RAB5A* gene was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 54°C for 30 s, 72°C for 120 s, final extension at 72°C for 10 min. And the thermal profile for the PCR targeting the *WAC* gene was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 45°C for 60 s, 72°C for 60 s, final extension at 72°C for 10 min. All PCR reactions were visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium

bromide in TAE buffer. These results were used to calculate the genome coverage of both the male and female cosmid libraries.

Identifying *CTNNB1*, *RAB5A* and *WAC* Positive Cosmid Clones

The cosmid sub-population PCRs that produced a DNA fragment consistent with the predicted DNA size of the region being targeted (from the genes *CTNNB1*, *RAB5A* or *WAC)*, were identified from the female cosmid library. The three cosmid pools that had the strongest PCR result for each of the three genes (gene; *CTNNB1*, *RAB5A* and *WAC)* were selected, and the corresponding glycerol stocks were removed from storage for further analysis. The second set of glycerol stocks remained in -80°C storage unaltered.

One hundred microliters of each of glycerol stock was aliquoted into a fresh 1.5 ml microfuge tube and the glycerol stock was then quickly returned to -80°C storage. Each 100 µl aliquot had 900 µL of LB broth containing 0.1 mg/ml ampicillin added to it producing a 1:10. Four dilutions were plated on to LBA plates containing 0.1 mg/ml ampicillin to determine the number of the CFUs present. The plates were left to incubate at 37°C overnight. It was determined that the 1:10 dilution of the original stock grew between 250-500 CFUs.

The remaining 640 µl of each diluted glycerol stock was then subsequently plated out evenly onto 7, LB-amp agar plates, using approximately 90 µl per plate. The plates were left to incubate at 37°C overnight. After incubation 2 ml of LB broth containing 0.1 mg/ml ampicillin was pipetted onto the top of each agar plate. A disposable plastic sterile spreader was used to gentle rub the plate and detach and break up all the colonies on the LB agar. The broth was collected with a pipette into a 15 ml centrifuge tube and the process of flooding of the plate and collecting the suspension was repeated to ensure all bacterial colonies had been collected. After all colonies from the 7 plates had been removed and aliquoted separately, 750 µl of each sample was aliquoted into two 1.5 ml microfuge tubes. Then 750 µl of 30% sterile glycerol was added to each tube and they were stored at -80°C, giving two glycerol stock copies of each of the 7 cosmid pool sub-populations. One millilitre of the remaining suspension was used for further DNA extraction and the remaining suspension was stored at 4°C.

The remaining 1 ml of each cosmid sub-populations was transferred into a new separate 1.5 ml microfuge tube ready for DNA extraction. The DNA was extraction from each sub-pool according to previously described protocols. A PCR was then performed using the freshly extracted DNA to determine which cosmid sub-population still contained the gene regions of interest. The PCR protocol for each of these three gene regions, specific for the genes *CTNNB1*, *RAB5A* and *WAC* were performed as previously described.

This process of dividing the cosmid populations into a series of sub-populations and identifying the sub-population that contained the gene regions from either the *CTNNB1*, *RAB5A* and *WAC* gene region, was repeated five times using the same protocol. The concentration of CFUs plated out from each sub-sampling was reduced by a factor of 2.

In the last round of sub-sampling, individual cosmid clones were investigated for the gene regions being targeted. Half of each colony from the LB agar plate were picked and placed into a 1.5 ml microfuge tube containing 100 µl of high purity water and mixed vigorously. Twenty-two colonies from each sub-population were investigated. A PCR on each of colony picks was performed using the following conditions; in a total reaction volume of 10 µl, containing 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10× PCR buffer, 0.1 µM of each primer (Table 5-1), 1 µl of template DNA (100 µl colony pick). The thermal profile for the *CTNNB1* PCRs was 95°C for 10 min, followed by 40 cycles of: 95°C for 30 s, annealing temperature of 55°C for 60 s, 72°C for 120 s, final extension at 72°C for 10 min. The thermal profile for the *RAB5A* PCRs was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 54°C for 30 s, 72°C for 120 s, final extension at 72°C for 10 min. And the thermal profile for the *WAC* PCRs was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 45°C for 60 s, 72°C for 60 s, final extension at 72°C for 10 min. All PCR reactions were visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer.

Roche 454 Sequencing

Ten cosmid clones were selected that were PCR positive for the gene regions being targeted; 5 clones representing the *CTNNB1* gene, 2 representing the *RAB5A* gene and 3 representing the *WAC* gene. Each remaining half colony located on the LB agar plate was picked and used to inoculate 10 ml of LB broth containing 0.1 mg/ml ampicillin within a 15 ml centrifuge tube

and incubated at 37°C overnight. The cosmid DNA was then extracted from the 10 inoculations using a Qiagen Plasmid Extraction Kit, following the manufacturer's instructions. The DNA extracted was visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer. The DNA concentrations of each sample were determined using a Nano-spectrophotometer.

Two sequencing reactions were performed on each of the positive clones using the T3 and T7 promoter primer sequences located within the vector, adjacent each of end of the DNA insert. Macrogen Korea, 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511, Rep. of Korea, who performed the DNA sequencing.

The ten positive cosmid clones were also sequenced using Roche 454 next-generation sequencing, using the sequencing platform Roche GS Junior. Ten DNA libraries were constructed representing each of the 10 cosmid clones using the protocol Sure Select Target Enrichment System for Roche 454 GS FLX and GS Junior Sequencing Platforms and the reagents provided from the manufacturer Agilent Technologies. Each cosmid clone was labelled with a unique identifying oligonucleotide adaptor during the library construction protocol, which was required for both DNA sequencing and later identification (Table 5-2). Use of the Roche GS Junior platform was contracted to Dr Mark Castalanelli from the WA Museum Collections and Research Centre; 49 Kew Street, Welshpool Western Australia 6106.

Table 5-2 Roche 454 library oligonucleotide adaptors

Cosmid Clones	Library Oligonucleotide Adaptors	Oligonucleotide Sequence
Cosmid C1	BC1	TCTCTGTGATGAT
Cosmid C2	BC2	TGTACGTGATGAT
Cosmid C3	BC3	ATCGTCTGATGAT
Cosmid C4	BC4	TAGCTATGATGAT
Cosmid C5	BC5	AGTATCTGATGAT
Cosmid W1	BC6	TCGAGCTGATGAT
Cosmid W2	BC7	TCATACTGATGAT
Cosmid W3	BC8	TCATACTGATGAT
Cosmid R1	BC9	ACTCACTGATGAT
Cosmid R2	BC10	TACGACTGATGAT

Ion Torrent Sequencing

In preparation of the Ion Torrent sequencing experiment, 4 cosmid clones representing the *CTNNB1* gene, 2 representing the *RAB5A* gene and 2 representing the *WAC* gene were isolated. The cosmid DNA was extracted and purified from these 8 selected cosmid clones as previously described. Cosmid DNA extracted from the 8 positive cosmid clones was sequenced on a 314 chip, Ion Torrent Personnel Genome Machine (PGM) platform. This service was contracted to Dr Richard Allcock at the University of Western Australian and Lotterywest State Biomedical Facility Genomics.

The cosmid DNA extractions in conjunction with sequencing were digested with the endonuclease enzyme EcoR1 under the following reaction conditions; 1 µg of cosmid DNA, 5 units of EcoR1 (Promega), 1 µl of 10× reaction buffer, made up to a final volume of 10 µl with high purity water. Each reaction was incubated at 37°C for 1 hour. After digestion, 5 µl of 6x loading buffer containing, with 0.1% SDS, was added to each reaction to stop the digestion continuing and fragment pattern produced from each digestion reaction was visualised by agarose gel electrophoresis using a 0.7% (w/v) agarose gel in TAE, stained with 0.5 µg/ml ethidium bromide in TAE buffer.

The cosmid DNA isolated from the cosmids clones was also PCR amplified for the purposes of Sanger DNA sequencing of the PCR products produced. Each cosmid PCR was replicated 5 times according to the following PCR condition; each PCR reaction had a total volume of 10 µl and contained 0.5 units of Taq DNA polymerase (Roche), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 1 µl of 10× PCR buffer, 0.1 µM of each primer (Table 5-1), 1 µl of template DNA (cosmid DNA extraction). The thermal profile for the *CTNNB1* PCRs was 95°C for 10 min, followed by 40 cycles of: 95°C for 30 s, annealing temperature of 55°C for 60 s, 72°C for 120 s, final extension at 72°C for 10 min. The thermal profile for the *RAB5A* PCRs was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 54°C for 30 s, 72°C for 120 s, final extension at 72°C for 10 min. And the thermal profile for the *WAC* PCRs was 95°C for 10 min, followed by 50 cycles of: 95°C for 30 s, annealing temperature of 45°C for 60 s, 72°C for 60 s, final extension at 72°C for 10 min. Each PCR replicate was combined into one microfuge tube and a PCR clean-up performed using a QIAquick PCR Purification Kit following the manufacturers protocol. The samples were then sent to Macrogen Korea, 10F, 254 Beotkkot-ro, Geumcheon-gu, Seoul, 08511, Rep. of Korea, who performed all sequencing reactions, sequencing clean-ups, and DNA sequencing.

Sequence Analysis

All sequence analysis was performed using Geneious (Biomatters Ltd, Auckland, NZ). DNA that had sequence similarity to the SuperCos 1 vector was identified using the plug-in BLAST algorithm and subsequently trimmed. All sequence sets were investigated visually and all DNA sequences that were shorter than 20bp in length were discarded. A *de novo* sequence assembly was performed for each sequence data set produced. All individual sequences produced using Sanger DNA sequencing and the contigs produced from the *de novo* sequence assembly were investigated using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) in a general search of all nucleotide sequences in the NCBI's database and a more specific search using the Python_molurus_bivittatus-5.0.2 reference Annotation Release 101 (Castoe et al. 2013). Sequence date sets that produced recurring BLAST sequence matches to specific *P. bivittatus* genomic scaffolds were investigated further. The identified *P. bivittatus* genomic scaffolds were retested against the NCBI's database and the sequence data set was assembled using the *P. bivittatus* scaffold as the reference sequence.

Results

Two cosmid library representing one male and one female genome of the snake species *M. spilota* were produced. It was predicted that the genome size of *M. spilota* is comparable to *Python reticulatus*, which is the most closely related snake species that has had its genome size investigated and is estimated to be approximately 1.44Gb (17,18). With the number of phage particles produced during the cosmid packaging reaction and the minimum cosmid DNA insert size (between 30kb and 42kb), it was calculated that the minimum coverage of our two cosmid libraries was between 8.4 and 11.8 times the male and female genome respectively. Moreover, the average number of positive PCR results obtained from PCR amplifying the three reference genes *CTNNB1*, *RAB5A* and *WAC*, from the 81-female and 81-male cosmid amplification pools, was 9.33 and 10 respectively, further supporting the calculated genomic coverage of these libraries (Table 5-3).

Table 5-3 PCR results from cosmid amplification pools

PCR Gene Targets	Female Cosmid Library Positive Pools	Male Cosmid Library Positive Pools
CTNNB1	7	8
RAB5A	9	9
WAC	12	13
Calculated Average	**9.33**	**10.00**

Roche 454 Sequencing

To identify and isolate individual cosmid clones that contained the three gene regions being investigated (*CTNNB1*, *RAB5A* and *WAC*), three positive cosmid pools, those which produced the most intense PCR fragment when visualised on an agarose gel, were studied further. Sub-populations of each cosmid pool were prepared and the process repeated, until 10 cosmid clones that produced the most intense positive PCR result, of the correct fragment size were isolated. DNA sequencing results derived from sequencing each end of the cosmid clones using the T3-promoter and T7-promotor primers, present in the cosmid vector SuperCos 1, are shown in Table 5-4.

Table 5-4 Cosmid clones T3 and T7 sequencing

Cosmid Clones	**PCR Results**			**T3 Sequences**		**T7 sequences**	
	CTNNB1	***RAB5A***	***WAC***	**BLAST Results**	***P. bivittatus* Scaffold**	**BLAST Results**	***P. bivittatus* Scaffold**
Clone 1 - C1	+			PREDICTED: *P. bivittatus* TBRG4 mRNA	1926 (NW_006533940)	PREDICTED: *P. bivittatus* NACAD mRNA	1926 (NW_006533940)
Clone 2 - C2	+			PREDICTED: *P. bivittatus* SEPT9 mRNA	743 (NW_006532757)	No similarities found	743 (NW_006532757)
Clone 3 - C3	+			No significant similarities found	No similarities found	PREDICTED: *P. bivittatus* CHEK2 mRNA	7594 (NW_006539608)
Clone 4 - C4	+			PREDICTED: *P. bivittatus* ACAP2 mRNA	5151 (NW_006537165)	PREDICTED: *P. bivittatus* ETV7 mRNA	252 (NW_006532266)
Clone 5 - C5	+			No similarities found	7960 (NW_006539974)	PREDICTED: *P. bivittatus* LOC103063316 mRNA	9586 (NW_006541600)
Clone 6 - W1			+	No similarities found	747(NW_006532761)	No similarities found	4430 (NW_006536444)
Clone 7 - W2			+	PREDICTED: *P. bivittatus* U2AF2 mRNA	2083 (NW_006534097)	No similarities found	2083 (NW_006534097)
Clone 8 - W3			+	PREDICTED: *P. bivittatus* U2AF2 mRNA	2083 (NW_006534097)	No similarities found	2083 (NW_006534097)
Clone 9 - R1		+		No similarities found	No similarities found	PREDICTED*: P. bivittatus* PFKP mRNA	7272 (NW_006539286)
Clone 10 - R2		+		No similarities found	No similarities found	PREDICTED*: P. bivittatus* PFKP mRNA	7272 (NW_006539286)

The T3 and T7 sequencing results showed that the majority, 9 out of 10 of the inserts contained within the isolated cosmids, had identifiable sequence similarities with the *P. bivittatus* genome. The sequencing also showed that it is likely some cosmid duplicates were present. Clones 7 - W2 and 8 - W3 produced similar DNA sequencing results, as did Clones 9 - R1 and 10 - R2. The sequencing chromatograms for these probable duplicates, were almost identical, indicating that these two cosmid duplicates possibly contained identical DNA inserts from the same DNA region. However, this is not enough to rule out the possibility that these duplicate cosmid clones represent two different alleles from the same DNA region and for this reason it was decided all 10 cosmid clones isolated would be sequenced.

The Roche 454 sequencing using the Roche GS Junior platform, produced 162,795 individual sequences. The average sequence length was 92.7bp with a standard deviation of 176.1bp. The experiment produced a total a 15.09Mb of sequence. The software program Geneious and the 10 library identification tags presented in Table 5-2, were used to separate the DNA sequencing data produced for each of the 10 cosmid clones sequenced. *De novo* sequence assembly for each cosmid clone, using the software Geneious, failed to produce any adequate sequence assemblies. This was attributed to a lack of raw sequence data and therefore limited sequence coverage. However, the data produced from this analysis and using BLAST allowed for the identification of the most likely scaffolds from NCBI *P. bivittatus* Annotation Release 101 that were representative of each cosmid clone. Using these scaffolds as reference sequences, sequence assemblies of each DNA sequence data set was performed (Table 5-5).

Table 5-5 Roche 454 sequence data

Roche 454 Data	Cosmid - C1	Cosmid - C2	Cosmid - C3	Cosmid - C4	Cosmid - C5	Cosmid - W1	Cosmid - W2	Cosmid - W3	Cosmid - R1	Cosmid - R2
Number of Sequences	1621	1754	19176	No Data	14770	105	658	No Data	No Data	344
Minimum size (bp)	4	1	1		1	3	4			6
Maximum size (bp)	557	561	565		595	530	568			558
Mean (bp)	392.5	390.7	376.8		387	406.7	423.8			385.3
Standard Deviation (bp)	134.7	132.5	142.8		144.4	112.7	121.3			134.8
Reference Assembly										
Reference *P. bivittatus* Scaffold	1926	743	7594		7960/9586		2083			7272
Number of Sequences Mapped	1177	1186	1086		604/1157		597			206
Estimated Cosmid Length (bp)	46687	over 28961	over 12089		over 41105		38956			over 27727
Cosmid Coverage (%)	44.6	90.3	95.1		52.0/53.9		66.0			65.7
Pairwise Identity (%)	93.4	92.4	97.3		86.4/79.0		85.9			86.3
Reference Sequence Description	*TBRG4* and *NACAD* genes	*SEPT9* gene	*CHEK2* gene and beyond		*CDK5* and three uncharacterised genes		*CCDC106* and *U2AF2* genes			Unknown

Cosmid Clone - C1

A total of 1,621 individual sequences reads, averaging 392.5bp in length representing cosmid clones C1. One thousand, one hundred and seventy-seven of these sequences were successfully mapped to the *P. bivittatus* scaffold 1926 (NW_006533940; Figure 5-2). The cosmid DNA sequences mapped to the reference scaffold spanned a region of 46kb in length with an overall 44.6% coverage. This cosmid DNA insert length is above the maximum insert size of 42kb. However, within the reference scaffold region mapped, 14 separate assembly gaps exist, possibly giving a false prediction of the expected cosmid insert length. Using this reference sequence, it was identified that the cosmid clones most likely represented the entirety of the Transforming Growth Factor Beta Regulator 4 (*TBRG4*) gene and about half of the NAC Alpha Domain Containing (*NACAD*) gene in *M. spilota*.

Figure 5-2 - Cosmid Clone - C1 Reference Assembly

Cosmid Clone - C2

For the cosmid clone C2, 1,754 individual sequence reads were produced, of which 1,186 reads were mapped to the *P. bivittatus* scaffold 743 (NW_006532757; Figure 5-3). The sequences from the C2 cosmid were well distributed across the cosmid's entire length and as a result, the overall coverage of the reference sequence was approximately 90.3%, for a 28kb region. However, within this reference sequence region a 1,339bp assembly gap existed and there was not enough sequence data to bridge this assembly gap. This indicates that the cosmid insert is most likely larger than 28kb in length. The sequences produced from the cosmid clone C2 likely represented approximately one quarter of the Septin 9 (*SEPT9*) gene, approximately 176kb in length.

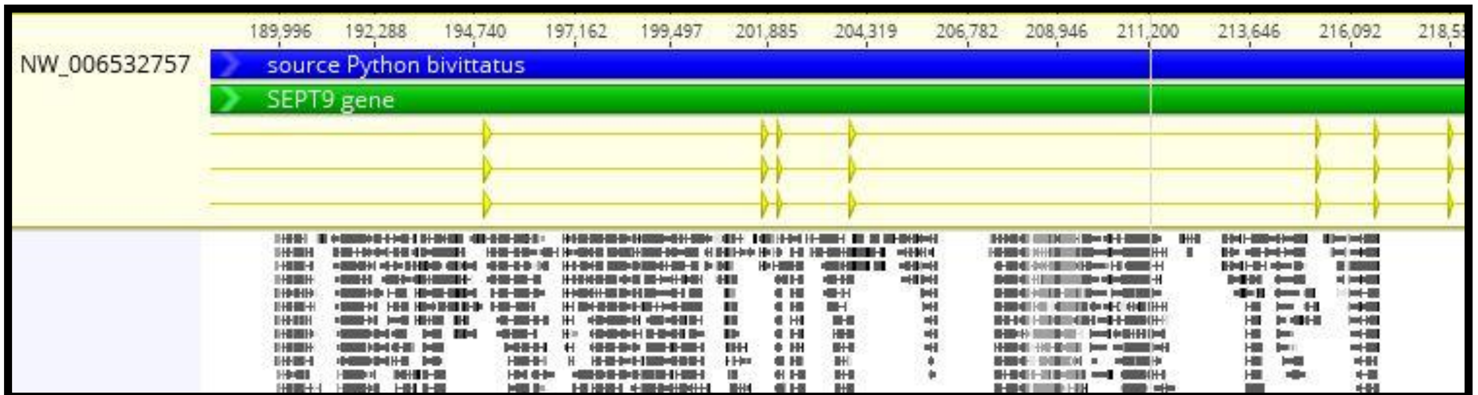

Figure 5-3 - Cosmid Clone - C2 Reference Assembly

Cosmid Clone - C3

Comparing the Roche 454 sequencing results between cosmid clones, cosmid C3 was the most successful with 19,176 individual sequence reads produced. However, only 1,086 sequences were mapped to the *P. bivittatus* reference scaffold 7594 (NW_006539608), leaving 17,702 individual sequences unannotated (Figure 5-4). The 17,702 sequence reads that were unused in the reference sequences assembly were investigated further and it was found that 14,859, approximately 77% of the total sequences produced, matched the SuperCos 1 vector sequence. The reference assembly comparison identified a 12kb region, with an overall sequence coverage of approximately 95.1%. Furthermore, our cosmid sequence assembly appears to bridge an assembly gap within the region mapped to the *P. bivittatus* reference scaffold 7594.

Cosmid clone C3, represents part of the last intron and the terminal exon for the gene Checkpoint Kinase 2 (*CHEK2*). Cosmid 3 extends 9kb downstream of the gene, to where the cosmid contig ends because the *P. bivittatus* scaffold is missing. The location of the T7 primer sequences was determined to be located at one end of the cosmid clone reference sequence assembly, but the location of the T3 primer sequence was unable to be determined. This provides evidence suggesting that the cosmid sequences extend past the reference sequence.

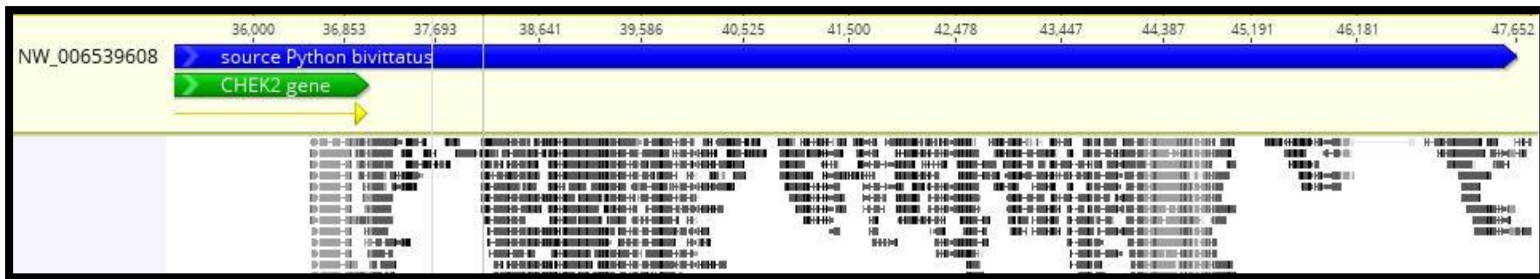

Figure 5-4 - Cosmid Clone - C3 Reference Assembly

Cosmid Clone - C5

Cosmid clone C5, represents a collection of genes from two different *P. bivittatus* scaffolds 7960 (NW_006539974) and 9586 (NW_006541600). Six hundred and four sequences were mapped to scaffold 7960, representing approximate 31.5kb of the uncharacterised gene *LOC103062025* and the entirety of the uncharacterised gene *LOC103062264* (Figure 5-5). The cosmid sequences map to scaffold 7960 right up to the 3' end, indicating that the cosmid insert most likely extends past the *P. bivittatus* scaffold assembly.

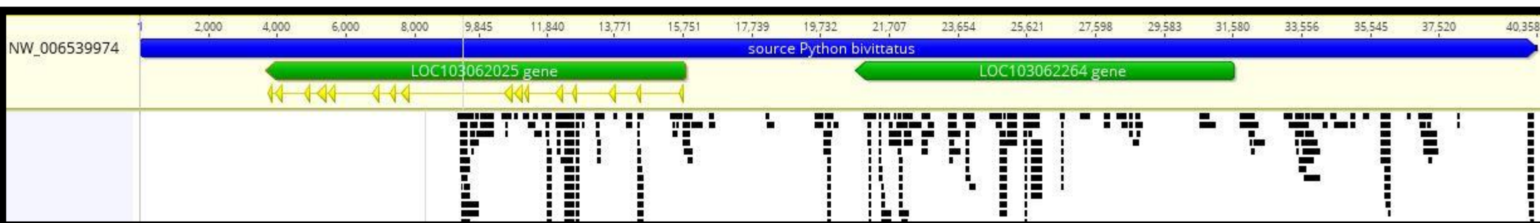

Figure 5-5 - Cosmid Clone - C5 Reference Assembly P. bivittatus scaffolds 7960 (NW_006539974)

A further 1157 sequences where mapped to *P. bivittatus* scaffolds assembly 9586, over a region approximately 9kb in length (Figure 5-6). Over this region two genes exist, the Cyclin Dependent Kinase 5 (*CDK5*) gene and the uncharacterised gene *LOC103063316*. Interestingly the T3 derived cosmid sequences were found at the start of the cosmid assembly represented by the scaffold 7960, but the T7 derived cosmid sequences were lacking in either scaffold assembly. This indicates that the two *P. bivittatus* scaffold assemblies 7960 and 9586 are possibly not adjacent to one another within the genome but the sequences from scaffold 7960 could reside within scaffold 9586.

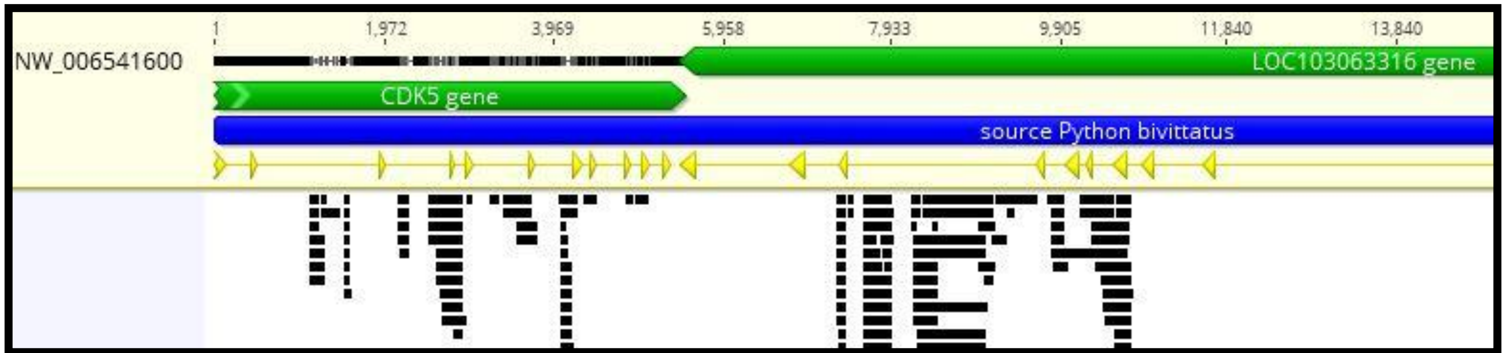

Figure 5-6 - Cosmid Clone - C5 Reference Assembly P. bivittatus *scaffolds assembly 9586 (NW_006541600)*

Cosmid Clone - W1

The Roche 454 sequencing results produced for cosmid clone - W1 were poor, with only 105 sequences produced that did not represent the SuperCos 1 vector sequence and were greater than 20bp in length. As a result, trying to map the 105 cosmid sequences to a reference genome assembly was very difficult. The T3 and T7 cosmid sequences were localised to two *P. bivittatus* scaffolds assemblies; 747 (NW_006532761) and 4430 (NW_006536444), which are likely to represent parts of the cosmid clone. However, the lack of sequence data resulted in an ambiguous reference assembly.

Cosmid Clone - W2

For the cosmid clone – W2 both the T3 and T7 cosmid sequences were mapped to the *P. bivittatus* scaffolds assembly 7954 (NW_006539608). This allowed us to determine the length of the DNA insert of W2 as approximately 38,956bp. In addition, 597 individual sequences were mapped to reference scaffold 7954, which gave us an overall sequence coverage of 66%. Within scaffold 7954 there are 5 individual assembly gaps, of which our sequences data bridges 4. This cosmid clone represents two genes; the entirety of the Coiled-coil Domain Containing 106 (*CDCC106*) gene and approximately half of the U2 Small Nuclear RNA Auxiliary Factor 2 (*U2AF2*) gene.

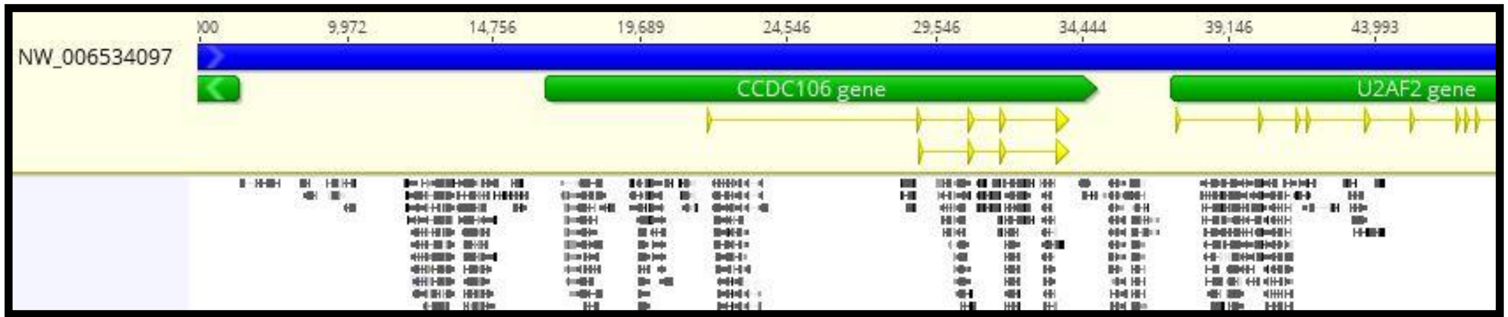

Figure 5-7 - Cosmid Clone – W2 Reference Assembly

Cosmid Clone - R2

Three hundred and forty-four independent sequence reads for the cosmid clone – R2 of which 206 of were mapped to the *P. bivittatus* scaffolds assembly 7272 (NW_006539286; Figure 5-8). This represented a region approximately 27kb in length, with an overall sequence coverage of approximately 65%. The reference scaffold 7272 lacks gene annotations and has several assembly gaps. Using the consensus sequence produced from the reference assembly it was determine that the DNA region has sequence similarity to PREDICTED: *P. bivittatus* adaptor related protein complex 4 epsilon 1 subunit (*AP4E1*), mRNA and PREDICTED: *P. bivittatus* phosphofructokinase, platelet (*PFKP*), transcript variant X3, mRNA, as was highlighted by the BLAST results of the T7 cosmid derived sequence. The *P. bivittatus* scaffold assembly 155 (NW_006532169) alignment was investigated further and it represents the *PFKP* gene. However only 36 sequences could be mapped to this reference sequence region, with only 53.3% sequence pairwise identity.

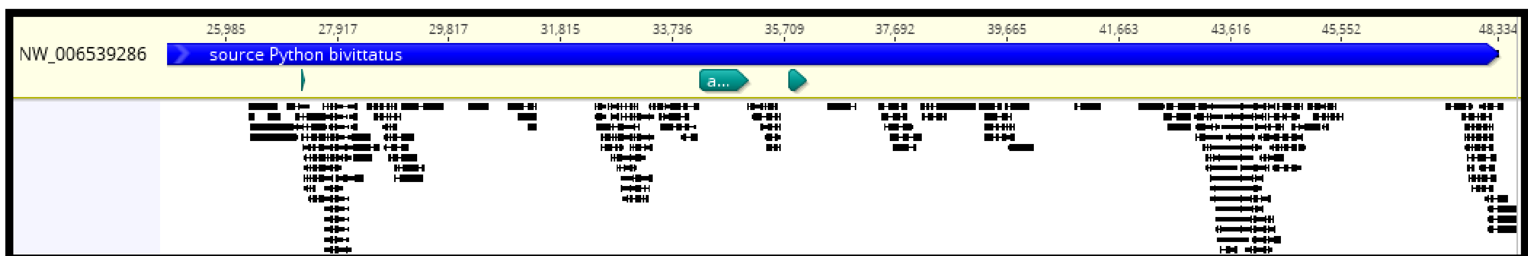

Figure 5-8 - Cosmid Clone - R2 Reference Assembly

Ion Torrent Sequencing

Eight cosmid clones were isolated and sequenced using the Ion Torrent sequencing platform; four representing the gene *CTNNB1*, two representing the gene *RAB5A* and two representing the gene *WAC*. Cosmid DNA extractions were used to perform restriction endonuclease mapping, which were visualised using agarose gel electrophoresis (Figure 5-

9). These results show that each cosmid clones isolated, 1 through 8, have a defined band at position 7.9kb, which is most likely representative of the cosmid vector SuperCos 1. This band was followed by a variety of smaller DNA fragments, which represents the inserted genomic DNA. Using these results, it is probable that cosmid clones 2 and 3 contain the same genomic DNA insert, whilst all other cosmid clones exhibit different DNA fragment profiles, indicating that they represented different inserted genomic DNA region.

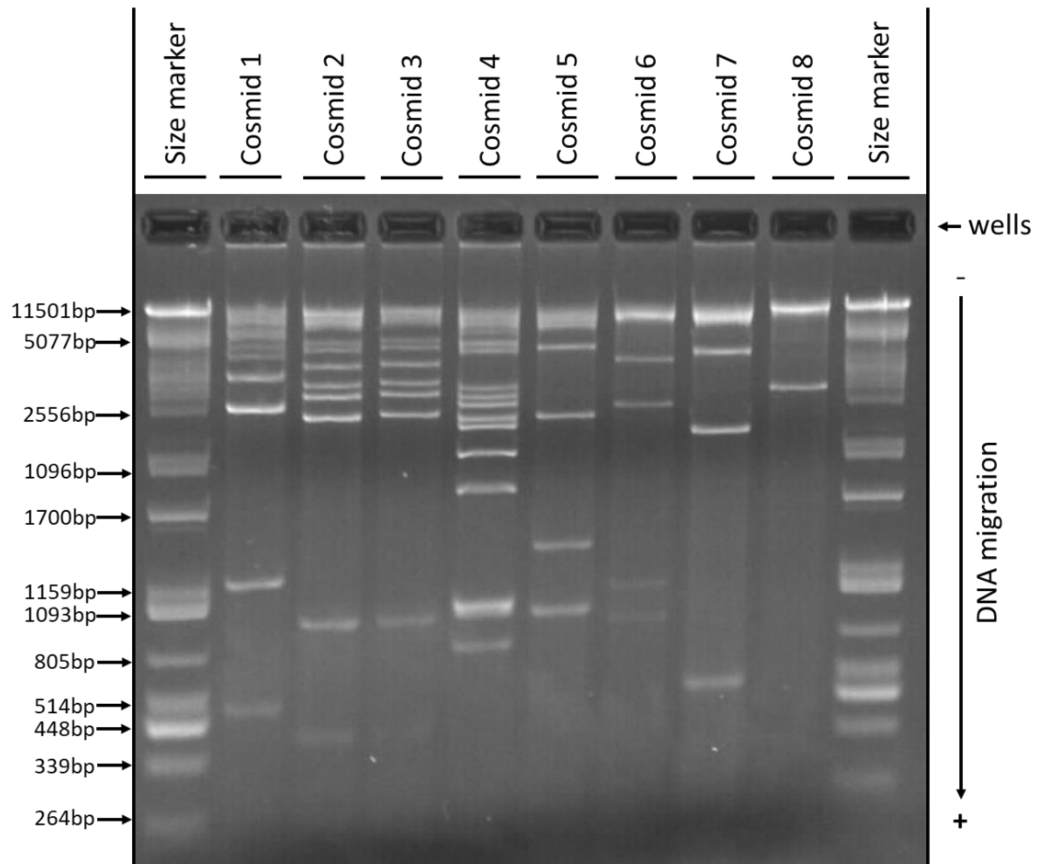

Figure 5-9 ECOR1 digest of cosmid clones 1 through 8

Figure 5-9 is an agarose gel electrophoresis image of the cosmid clones 1 through 8 that have been digested with the endonuclease ECOR1.

DNA sequencing of the products from PCR amplified each cosmid clone using the primers specific for *CTNNB1*, *RAB5A* and *WAC* was performed. However, all DNA sequencing results returned what appeared to be a mixed template sequencing signals and none of the DNA sequencing results could be reliably determined. In all cases, the DNA sequencing result indicated a PCR product size that was consistent to the DNA region being targeted.

The Ion Torrent next-generation sequencing experiment produced 734,729 individual sequences, with an average sequence length of 262.9bp and a standard deviation of 92.2bp. This generated approximately 193Mb of DNA sequences representing the 8 cosmid clones (Table 5-6).

Table 5-6 Ion Torrent sequence data

Ion Torrent Data	Cosmid Clone 1	Cosmid Clone 2	Cosmid Clone 3	Cosmid Clone 4	Cosmid Clone 5	Cosmid Clone 6	Cosmid Clone 7	Cosmid Clone 8
Number of Sequences	75,234	102,807	69,925	120,919	96,791	80,035	64,917	124,101
Minimum size (bp)	8	8	8	8	8	8	8	8
Maximum size (bp)	484	605	526	572	562	609	542	533
Mean (bp)	262.7	267.6	262.5	259.1	262	265.7	263.6	261.9
Standard Deviation (bp)	93.4	90.1	93.7	93	92.8	89.1	90.8	93.4

Prior to any bioinformatics analysis, the SuperCos 1 vector sequences were trimmed and DNA sequences less than 20bp in length removed. This pre-analysis was performed for all 8 data sets. However, the number of sequences remaining after this process, were variable (Table 5-7). The resulting sequences after processing from cosmid clones 4 and 6, represented 68% and 80% respectively, of the total DNA sequences produced for each cosmid clone. However, the sequences resulting for the cosmid clones 1, 2, 3, 5, 7, and 8 were all less than 2% of the DNA sequences from the initial sequence data sets.

Table 5-7 Comparison after sequence trimming

Ion Torrent Cosmid Clones	Initial number of sequences produced	Remaining number of sequences after trimming	Percentage Difference (%)
Cosmid Clone 1 (*CTNNB1*)	75,234	1249	1.66%
Cosmid Clone 2 (*CTNNB1*)	102,807	1642	1.60%
Cosmid Clone 3 (*CTNNB1*)	69,925	1388	1.98%
Cosmid Clone 4 (*CTNNB1*)	120,919	81944	67.77%
Cosmid Clone 5 (*RAB5A*)	96,791	1681	1.74%
Cosmid Clone 6 (*RAB5A*)	80,035	64743	80.89%
Cosmid Clone 7 (*WAC*)	64,917	1022	1.57%
Cosmid Clone 8 (*WAC*)	124,101	2173	1.75%

This finding was investigated further by using the SuperCos 1 vector DNA sequence as a reference sequence and assembling the initial unaltered sequences from each data set. This analysis showed that for the cosmid clones 1, 2, 3, 5, 7, and 8, approximately 90% or more of the DNA sequences assembled to the SuperCos 1 vector sequence. Unfortunately, the small number of DNA sequences that represented the DNA inserts for cosmid clones 1, 2, 3, 5, 7, and 8 were not enough to identify any scaffold within the NCBI *P. bivittatus* Annotation Release 101. All *de novo* assemblies of these sequence sets produced no analysable results and subsequently no BLAST analysis could be performed accordingly.

Cosmid Clone - 4

Approximately 120,919 sequence reads were obtained from cosmid clone 4. After the sequences representing the SuperCos 1 vector and sequences that were below 20bp in length were removed from this sequence data set, 81,994 sequence reads remained. A large proportion of these sequence reads had sequence similarity to the *P. bivittatus* scaffold 2195 (NW_006534209). Using the *P. bivittatus* scaffold 2195 as a reference sequence approximately 80,571 (98%) sequences were assembled, which gave 99.8% coverage of a 12,813bp DNA region (Figure 5-10).

The assembled DNA region represents the initial exon from the Sine Oculis Binding Protein Homolog (*SOBP*) gene and a small proportion of the corresponding intron 1. Approximately 10,000bp of this cosmid represented the non-coding DNA region, upstream of the *SOBP* gene. The 12,491bp contig from *M. spilota* has a pairwise sequence alignment of 92.9%

with the *P. bivittatus* scaffold 2195. Interestingly, despite having 80,571 sequence reads over a 12,491bp region there was a proportion of the reference scaffold that was hardly represented by the sequence reads. This region spans approximately 200bp, with only three sequences representative of it. This result is also reflected in the reference *P. bivittatus* scaffold 2195 sequence, with the region identified as a sequence assembly gap.

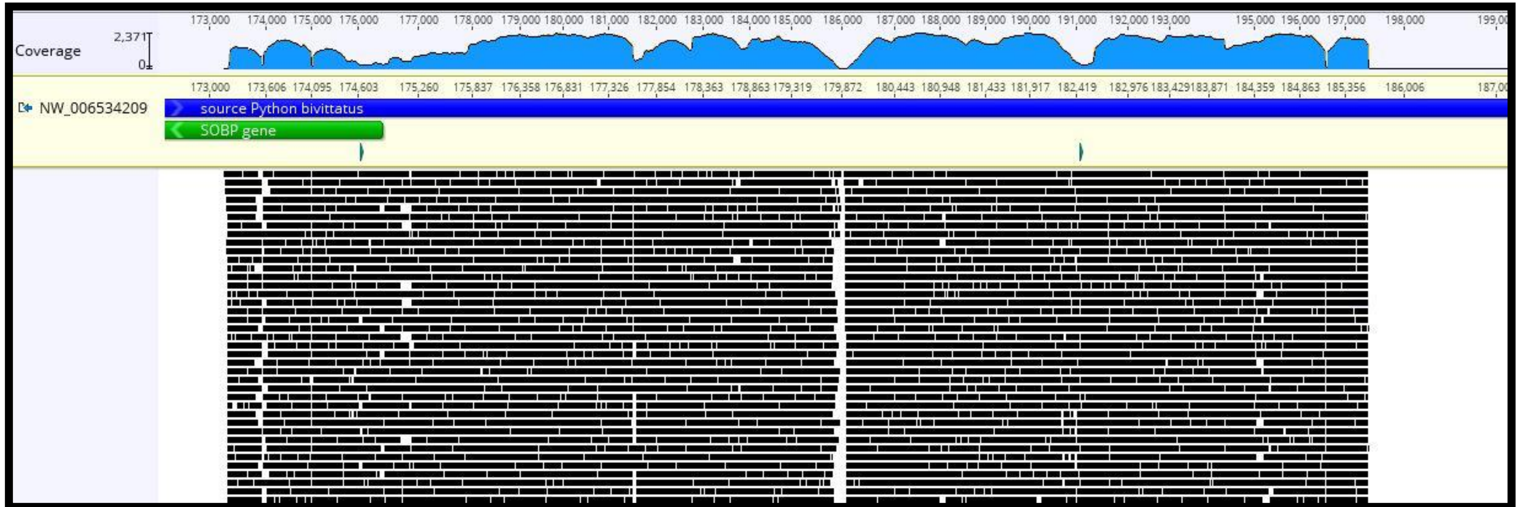

Figure 5-10 - Ion Torrent Cosmid Clone 4

Cosmid Clone - 6

The Ion Torrent sequencing produced 80,035 sequence reads from cosmid clone 6. After processing 64,743 sequence reads remained and 62,759 (97%) of these were assembled to the *P. bivittatus* scaffold 2195 (NW_006535725), with 98.8% coverage of the reference sequence. Using these data, a consensus sequence of cosmid - 6 was constructed, which was 32,617bp in length and had a pairwise sequence identity of 88.2% with the reference scaffold 2195. This sequence represented almost the entirety of two genes Neurotrophic Receptor Tyrosine Kinase 1 (*NTRK1*) and Regulation of Nuclear Pre-MRNA Domain Containing 2 (*RPRD2*).

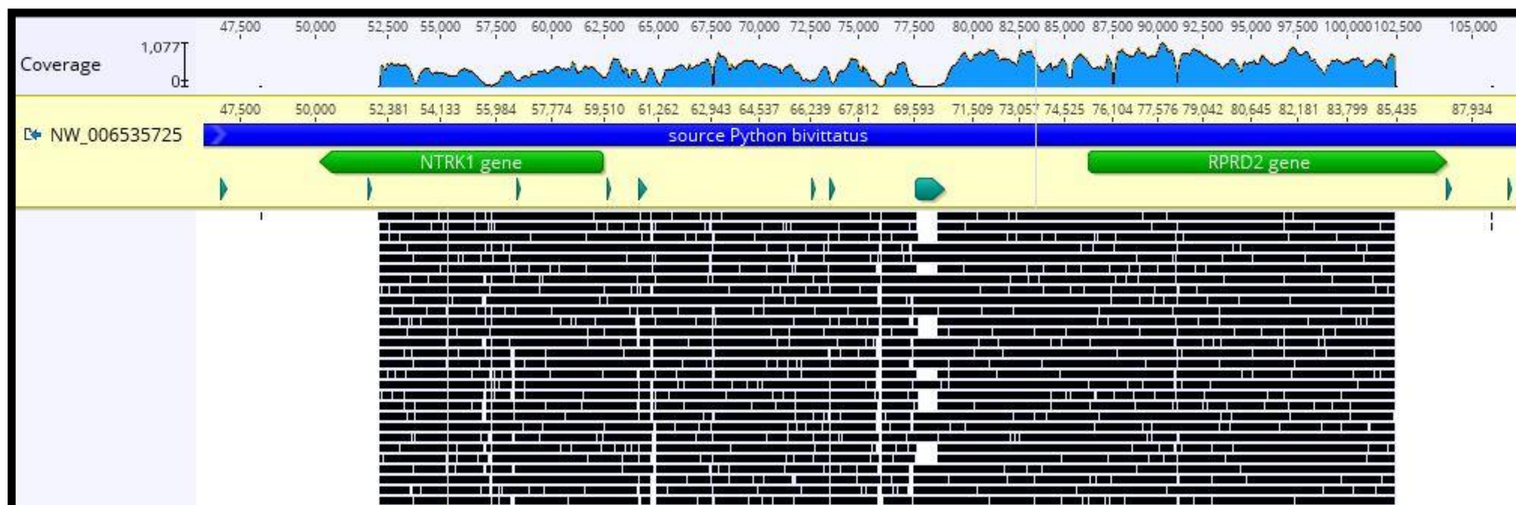

Figure 5-11 - Ion Torrent Cosmid Clone 6

Discussion

Using the SuperCos 1 vector system and the Gigapack III Gold-4 packaging reaction two cosmid libraries were produced, representing one male and one female genome for the snake species *M. spilota*. It was found that the efficiency of the procedure was reduced by small changes in DNA concentration, correct DNA size selection and DNA quality. It took 9 months to produce these two final high-quality libraries. It was calculated that the coverage of each cosmid libraries was at least 8 genomes and the experimental data further supported this calculation.

Cosmid Library Screen and Clone Selection

A PCR based screening method was employed in this study using previously developed PCR primers and reactions conditions that had been optimised in experiments from Chapter 4 of this thesis. This approach identified amplification positive pools of cosmids containing three DNA regions of interest. Using this approach, the initial investigation identified both positive and negative cosmid pools and the PCR fragments were shown to be the expected fragment size. However, the PCR results from further sub-populations of the cosmid pools produced multiple PCR fragments with a variety of different sizes, known as PCR false positives. This phenomenon occurred with all three PCR reactions used to target the three genes *CTNNB1*, *RAB5A* and *WAC*.

The production of a greater number and diversity of false positive PCR products with each round of sequential sub-sampling of the cosmid library, can be explained by two reasons. This first is that the PCR primers and optimised PCR reactions, for the three genes *CTNNB1*, *RAB5A* and *WAC*, were borrowed from Chapter 4 - Comparative Genomic study. This studied used mRNA and genomic DNA from a collection of distantly related vertebrate species to design PCR primers specific for conserved DNA regions within the gene targets of *M. spilota*. This meant that the PCR primers that were designed and the subsequent optimised PCR reaction needed to be broad and not very specific to ensure amplification of the target region, which in most cases did not match the PCR primers designed specifically. Specific PCR primers are important during screen methodologies to ensure that false PCR positives are limited (Israel 1993). Furthermore, most modern PCR-based screening methods employ a secondary screening method to validate PCR positives, either during the screening process or at the final stages of individual clone selection, illuminating almost all false positive PCR products (Israel 1993; Asakawa et al. 1997; Campbell and Choy 2002).

Secondly, after each round of sub-sampling the complexity of the template DNA being used in the screening PCR was reduced (Israel 1993; Asakawa et al. 1997; Campbell and Choy 2002). This process increased the likelihood of a DNA region with non-specific primer binding being PCR amplified, by increasing its DNA concentration in some sub-populations. Furthermore, by sub-sampling the competition associated with primer binding was also reduced. During normal genomic PCR amplification non-specific primer binding and amplification is usually outcompeted by specific primer binding of the DNA region being targeted, under optimised PCR conditions. By chance some of these sub-populations contained DNA regions of non-specific primer binding and lacked DNA region with competing specific primer binding. Resulting in a greater chance of PCR amplification of these non-specific DNA regions, otherwise known as or false PCR positives, as was observed (Israel 1993; Asakawa et al. 1997; Campbell and Choy 2002).

The occurrence of false positives made identifying cosmid sub-populations containing the three gene regions of interest extremely difficult. It was reasoned that selecting sub-populations of cosmid clones that produced a PCR product of the expected fragment size was sufficient. Because the likelihood of a false positive PCR product being a similar size to the actual PCR product being targeted would be remote.

We are unable to show that any of the final 18 cosmid clones sequenced contain the expected gene targets *CTNNB1*, *RAB5A* and *WAC*. On review, relying principally on the size of the PCR product being produced for validity of the screening results was flawed and a secondary test when screening the cosmid sub-populations would be suitable. A more robust method to validate that the PCR products produced were amplicons from the correct DNA target, is using some type of DNA fingerprinting, otherwise known as restriction mapping (Campbell and Choy 2002). Using digestion with a specific endonuclease enzyme PCR products could have been used to produce a restriction fragment length pattern capable of resolving false positives (Campbell and Choy 2002). By comparing the fingerprint produced from positive cosmid sub-populations with previously determined DNA fingerprint from a control PCR product (Campbell and Choy 2002). The addition of this technique would have provided an accurate validation test to ensure the cosmid clones being targeted contained the gene regions of interest (Campbell and Choy 2002). Alternatively, a combination of PCR and colony hybridisation would ensure a greater degree of specificity (Campbell and Choy 2002).

Roche 454 Sequencing

The results produced from sequencing each end of the 10 cosmid clones isolated, using vector T3 and T7 primers, revealed very little specific information regarding the gene contents of each cosmid clone. When this experiment was conducted in 2012 the NCBI *P. bivittatus* Annotation Release 101 had not yet been published and there were limited snake genomic DNA sequences available in the NCBI's nucleotide database. As a result, it was difficult to determine much from the sequence information produced, except that no DNA sequences showed similarity with the three genes *CTNNB1*, *RAB5A* and *WAC*. However, this result was not considered discouraging, as each cosmid insert was between 30kb and 42kb in length and therefore it was considered likely that the gene sequences were internal to the approximate 800bp of sequence generated from DNA sequencing each end of the cosmids. As a result, the decision was made to proceed with Roche 454 sequencing for all 10 cosmid clones isolated.

The Roche 454 sequencing technology was the first commercial available next-generation sequencing system, launched in 2005 (L. Liu et al. 2012; Hoffmann-La 2010). The technology uses a complex enzymatic cascade that generates light, during pyrophosphate release, when an individual nucleotide is incorporated into a new DNA strand (L. Liu et al. 2012). The DNA preparation process requires the construction of a DNA library and the addition of

specific oligonucleotide adaptors to the DNA being sequenced (L. Liu et al. 2012). This allows for multiple libraries to be run together across a single sequencing run, as the oligonucleotide adaptors can be used to later identify the DNA sequences produced bioinformatically (L. Liu et al. 2012). To construct a DNA library a DNA sample is prepared by fragmenting the DNA either by enzyme or mechanical fragmentation and the unique oligonucleotide adaptor are ligated. The adapted DNA is then capture onto beads and the bound DNA is amplified using an emulsion PCR, ready for the DNA sequencing procedure. All steps involved in the library construction and preparation are critical in producing a DNA library of a high quality, which directly impacts the quality and quantity of sequencing results produced using the Roche 454 sequencing platform.

The Roche 454 GS Junior platform was released in 2010, with an expected output of 35Mb of sequence data, with an average read length of 400bp, from a single sequencing experiment (Hoffmann-La 2010; Loman et al. 2012). However, our Roche 454 sequencing experiment produced 162,795 individual sequences with an average length of 92.7bp, totalling 15Mb of sequence data. The main reason for this reduced data yield could be attributed to the DNA sequence read length being one quarter of what was expected. It was calculated that 74% of the total sequences produced were less than 20bp in length.

This result most likely occurred due to an inadequate removal of the short DNA sequences, remaining from the DNA preparation protocol. During library preparation, the DNA is fragmented to produce a DNA fragment range of between 400bp to 2,000bp ("Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series" 2012). The following processes result in the purification of the DNA and the removal of DNA fragments with a size less than 400bp, which are then discarded from the library preparation ("Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series" 2012). Failure to remove these small DNA fragments can result in them preferentially binding to the capture beads during the following steps of the library construction (Riggio 2017). As these smaller fragments bind to the capture beads more efficiently and outcompete the binding of larger DNA fragments and hence a reduced DNA sequence read length is observed as a result (Riggio 2017).

The Roche 454 sequencing study involved the construction of 10 individual DNA libraries, representing the 10 cosmid clones, with each having a unique oligonucleotide adaptor. However, 3 of the 10 unique oligonucleotide adaptors were not identified in the final

sequencing results, representing the cosmid clones C4, W3 and R1. This result was most likely caused by a failure of the ligation of the oligonucleotide adaptor to the fragmented DNA for the cosmid clones C4, W3 and R1 ("Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series" 2012; Riggio 2017). In this step of the library construction protocol, the oligonucleotide adaptor is ligated to the fragmented DNA and they are a critical component required to capture the DNA using the beads ("Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series" 2012). If the oligonucleotide adaptor has failed to ligate onto the fragmented DNA then no DNA will be captured onto the beads, resulting in no DNA being sequenced, as was observed for these three DNA libraries constructed ("Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series" 2012; Riggio 2017).

Sequence analysis showed that over two thirds of the sequences produced represented the SuperCos 1 vector and not the DNA inserts. Proportionally, it is expected that the DNA sequences produced, representing the cosmid vector (7.9kb), should have been between 20-26% of the total sequences. A possible cause for this finding is that when the cosmid clones were being amplified in their *E. coli* host strain, some bacterial cells managed to eject or rearrange a proportion or the genomic DNA insert within their cosmid vectors (Al-Allaf et al. 2013). Under stressful or strong selection conditions, *E. coli* has been known to rearrange vector DNA inserts and eject large DNA inserts to become more biologically efficient (Al-Allaf et al. 2013). With a smaller, more efficient cosmid, these bacterial cells have the ability to reproduce faster than their large cosmid laden counterparts, eventually outcompeting them (Al-Allaf et al. 2013). This process could have occurred during our *E. coli* preparation of cells for cosmid DNA extraction leading to a mixed population of cosmids. This would theoretically have resulted in higher than expected proportion of DNA sequences representing the SuperCos 1 vector, as was observed. To eliminate this problem in the future, we suggest an investigation for the presence of the 7.9kb SuperCos 1 vector minus insert be conducted on the cosmid DNA extraction prior to sequencing and as a precaution size selecting the cosmid DNA extracted from *E. coli* inoculations to ensure any unwanted vector minus insert is removed prior to DNA sequencing.

In total, 38,428 sequences were identified that contained one of the 10 oligonucleotide adaptors. This is less than one quarter of the total DNA sequences produced (162,795). Of the 124,367 sequences that did not contain an oligonucleotide adaptor approximately 117,820 (95%) of these sequences were less than 13bp in length. Considering each

oligonucleotide adaptor is 13bp in length, it is not surprising that our bioinformatic software could not distinguish any of the oligonucleotide adaptors from these sequences.

The number of sequences recovered for the remaining 7 oligonucleotide adaptors were extremely variable. For example, 19,176 sequence reads representing the cosmid clone C3, were identified but only 344 sequences for the cosmid clone R2. This variability might be attributed to the difference in the quality of the 10 libraries that were constructed. Approximately 38,428 sequences were identified that contained an oligonucleotide adaptor and were used to identify similar DNA sequence scaffolds in the NCBI *P. bivittatus* Annotation Release 101, for 5 of the 10 cosmid clones. The DNA sequences and sequence assemblies for cosmid clone C3 was represented with the greatest number of DNA sequences in the data set. As a result, this cosmid clone had the greatest percentage of coverage for the representative *P. bivittatus* DNA regions identified; 95% coverage with a sequence pairwise identity of 97%, representing a small proportion of the gene *CHEK2* and the DNA downstream. This sequence information could prove useful in other research studies comparing the genomic DNA from different snake species or other reptile species. Although, the DNA regions that the 4 other cosmid clones most likely represent were identified, the sequence coverage was considerably lower, and more DNA sequence information is required before these regions would be useful in further research. From the 162,795 DNA sequences produced, no DNA sequences that likely represented the three genes being targeted *CTNNB1*, *RAB5A* and *WAC* were identified.

Ion Torrent Sequencing

Considering the relatively poor result produced using the Roche 454 GS Junior platform, and the technical difficulties identified during library construction, the decision was made to re-isolate new cosmid clones and use a different second generation sequencing platform. The Ion Torrent PGM platform launched at the end of 2010, and uses semiconductor sequencing technology to read DNA sequences as they are being replicated within individual microwells of a chip (L. Liu et al. 2012). The DNA molecules are replicated using a single species of nucleotide at a time, with the addition of a single nucleotide polymerisation resulting in a measurable change in pH (L. Liu et al. 2012).

The 8 cosmid clones were isolated using the same PCR identification and sub-population culturing technique as was used in the Roche 454 sequencing experiment. The results are comparable, with false positive PCR results becoming more common with every round of

sub-population culturing performed. Due to this issue the 8 cosmid clones isolated were analysed in greater detail than for previous experiments, by generating an EcoR1 DNA fragment profile of the purified cosmid clones. The endonuclease EcoR1 was used because the SuperCos 1 vector contains two EcoR1 restriction sites at either end of the DNA insert (Evans, Lewis, and Rothenberg 1989). It was observed that 7 of the 8 cosmid clones purified showed a different DNA fragment profile (Figure 5-9). Cosmid clones 2 and 3, both positive for the gene *CTNNB1*, produced a similar DNA fragment profile, indicating that these two cosmid clones most likely contain the same or similar genomic DNA insert. This observation indicates that our PCR screening method resulted in selectively targeting the same false positive allele from the cosmid library twice. This result shows that our PCR-based screening method was used successfully to isolate a positive cosmid clone, although a false positive PCR targeting an incorrect DNA region.

DNA sequencing of the PCR products from the 8 cosmid clones was performed to confirm that the cosmids were true positives. All sequencing results produced a mixed DNA sequence signal and as a result were unreadable. However, the DNA sequence length of all DNA sequences produced was comparable to the expected size of the PCR for the DNA regions being targeted (from the genes *CTNNB1*, *RAB5A* and *WAC)*. We theorised that despite the DNA sequencing results failing to be readable, the PCR products produced were the correct size and therefore most likely from the correct DNA region.

DNA sequences for all the 8 cosmid clones were generated using the Ion Torrent PGM sequencing platform, producing a total of 193Mb of DNA sequences, almost four times the expected of 30-50Mb (“Ion PGM™ System Specifications,” n.d.). The average sequence length was 262.9bp, which was within the expected average sequence length of 200-400bp (“Ion PGM™ System Specifications,” n.d.). However, the investigation of the DNA sequences did not find result in any sequences that showed sequence similarity to the intended three gene targets *CTNNB1*, *RAB5A* and *WAC*. The DNA sequences produced representing the two cosmid clones 4 and 6 contained many high-quality reads and from these data sets a complete sequence assembly of both cosmid 4 and 6 clones were constructed. The sequence assembly for cosmid 4 was only 12,491bp in length, which was unexpected, considering the process of constructing and packaging our cosmid clones was strongly selective for DNA inserts that are larger than 30kb in length. The sequence assembly representing cosmid 6 was 32,617bp in length, as would be anticipated. Both cosmid clone assemblies included coding and noncoding DNA regions and both assemblies

bridged multiple assembly gaps within the two reference scaffolds from the NCBI *P. bivittatus* Annotation Release 101.

In contrast, the sequence analysis of the data sets from cosmid clones 1,2,3,5,7 and 8 showed that 98% or more of these data sets were made up of DNA sequences representing the SuperCos 1 vector. This leaves less than 2%, or approximately 1500-2000 individual sequence reads, in each data set representing the genomic DNA insert within each cosmid. This result supports the idea that the colonies used to generated the cosmid DNA extractions for the cosmid clones 1,2,3,5,7 and 8 contained were contaminated with *E. coli* that with only SuperCos 1 vector DNA and no insert, as previously discussed (Al-Allaf et al. 2013). This hypothesis is also supported by the observation that cosmid clone 4 only containing a 12kb insert approximately, which is possibly indicative of a cosmid clone that has undergone rearrangement of the DNA insert (Al-Allaf et al. 2013). However, the reasons why the sequencing results produced for cosmid clones 4 and 6 were successful, while all others failed is unclear, remains indefinable.

Using the DNA sequences produced from Cosmid Clone 4 a small proportion of the SOBP gene was identified. The protein translated from this gene is a nuclear zinc finger protein that is involved in development of the cochlea (Z. Chen et al. 2008). An autosomal recessive mutation in the SOBP gene has been identified to cause severe mental retardation in humans, including syndromic and non-syndromic ID with psychosis (Birk et al. 2010). In mice, two different mutations in the orthologue gene have a major impact on hearing loss as well as inducing erratic circling behaviour (Calderon et al. 2006). There are DNA sequences representing this gene in over 300 different species in the NCBI database, including three crocodilian species, three turtle species and a lizard species. However, to our knowledge there is only one snake species from which the gene has been sequenced previously; *P. bivittatus*. In humans, the gene is located on chromosome 6, in mouse chromosome 10 and in chicken chromosome 3 (NCBI Resource Coordinators 2017). The location of this gene within the snake genome has not been determined, however it is most likely located on chromosome 1 in snakes, a chromosome shown to share homology with chicken chromosome 3 (Matsubara et al. 2012).

Within the sequences produced from cosmid clone 6, two genes, NTRK1 and RPRD2, were identified. The NTRK1 gene encodes a protein that is a member of the neurotrophic tyrosine kinase receptor (NTKR) family and is a membrane-bound kinase receptor

(NCBI Resource Coordinators 2017). The NTRK1 protein is thought to play an important role in pain sensitivity, with mutations in the gene observed to be associated with decreased sensitivity to pain in humans (N. Li et al. 2018). The RPRD2 gene is also protein coding, but its function is not well understood. It has been suggested the gene product acts to regulate the synthesis of a specific microRNA molecule known as mir-205 and its associated biological functions (Son et al. 2017). While mutations in the gene have been found in Endemic Burkitt Lymphoma (Kaymaz et al. 2017). Both genes have been localised to Chromosome 1 in humans, chromosome 3 in mouse and chromosome 25 in chicken (NCBI Resource Coordinators 2017). The location of both gene has not yet been determined in snakes, with only a single gene example, from chicken chromosome 25, localised to a snake microchromosome, suggesting these genes might also be located on a snake microchromosome (Matsubara et al. 2012).

Conclusion

Our primary aim of this thesis chapter was to construct two cosmid libraries, representing the genome of one male and female *M. spilota*. This primary aim was successfully achieved. The secondary aim was to use these two cosmid libraries to located clones containing the gene regions observed in our previous experiments, specifically the DNA regions representing the three genes *CTNNB1*, *RAB5A* and *WAC,* was not achieved.

To eliminate the problems associated with the cosmid screening strategy it would have been useful to design new more specific PCR primers and optimised PCR reactions for the three gene targets, using the *M. spilota* DNA sequences obtained from previous experiments. This would have reduced the chances of producing false PCR positives during cosmid screening. In addition to the PCR-based screening technique a combination of two or more cosmid screening techniques paralleled would have been beneficial. Using a combination of two screening methods would significantly reduce the possibility of selectively targeting a false positive cosmid clone during cosmid screening.

To eliminate the problems associated with our next-generation sequencing experiment we suggest investigating the presence of the 7.9kb SuperCos 1 vector minus insert in cosmid DNA extractions before sequencing and as a precaution size select the cosmid DNA extracted, excluding the vector containing no inserts. Furthermore, reducing the concentration of ampicillin used in the sub-culturing of the transformed *E. coli* from 0.1

mg/ml to 0.05 mg/ml, would alleviate some of the selection pressure on the host cells and reduce the likelihood of cosmid DNA insert rearrangement.

Chapter 6 - Synopsis

The hypothesis for this thesis was that *Morelia spilota* uses a female heterochromatic sex chromosomes system with homogametic (ZZ) males and heterogametic (ZW) females, most likely common in all snake species. *M. spilota*; a non-traditional model snake species, was selected because it belongs to the Pythonidae snake family and has been shown to have cytologically indistinguishable sex chromosomes (Olmo and Signorino 2010). In addition, the proximity of this snake species natural habitat and its popularity among snake keepers and breeders made obtaining the required DNA samples relatively easy. *M. spilota* therefore was a perfect candidate as a member of our defined Heterogametic Sex Unidentified (HSU) group, to test the hypothesis. It was the primary objective to investigate *M. spilota* for a genetic sex determination mechanism, which is initiated by a sex specific DNA region or regions, which had evolved independently of other amniote sex determination systems. To achieve this objective, the molecular genetic techniques used needed to be conducted independently of previously discoveries of both snake and other amniote sex determination systems. Therefore, it was decided not to restrict the methods of investigation to sex specific differences in females only, but to males and females equally. It was believed that this approach would provide the greatest chance of discovering any new, undescribed sex determination system in *M. spilota* and the tool to either prove or disproving the hypothesis.

Unlike investigating the genetics of birds or mammals, investigating the genetics of reptiles and the genetics of snakes is and remains challenging. This is because there is a scarcity of genetic information representing different snake species. Since this thesis began in 2009, the first and second draft of *Python bivittatus* (Burmese python) genome and the first draft of *Ophiophagus hannah* (King cobra) genome have been added to NCBI's genome database (Vonk et al. 2013; N. Chen and Lai 2010; Castoe, de Koning, et al. 2011; Castoe et al. 2013). This information has greatly expanded our knowledge of snake genetics in general and has helped to understand the DNA sequence findings from experiments presented in this thesis. It is important to note that the first and second drafts of the *P. bivittatus* genome, which have been referenced extensively in this thesis, were published in 2011 and 2013 respectively and the research component of this thesis was completed in 2012.

To test the thesis hypothesis, four different molecular genetic techniques were employed, and these comprise the four experimental chapters in this thesis. Random Amplified Polymorphic DNA (RAPD) analysis and Representational Difference Analysis (RDA), were two techniques chosen for following reasons. These techniques do not require previous genetic knowledge, which at the time of commencement of the study were unavailable. In addition, these two techniques have been successful in identifying sex specific DNA regions in a multitude of other species (O'Neill et al. 2000; Leon Huynen, Craig D. Millar, and David M. Lambert 2002; Banerjee, Manoj, and Das 1999; Semerikov et al. 2003; Martinez E. A. et al. 2003; Agrawal et al. 2007; Urasaki et al. 2002; Bello and Sanchez 1999; Hormaza, Dollo, and Polito 1994). The third technique applied in this study was a comparative genomics approach, which was designed to build upon published research involving genes previously mapped to the Z and W chromosome of *Elaphe quadrivirgata* (Matsubara et al. 2006, 2012). While the fourth technique involved the construction of two genomic libraries, with the aim of using these libraries to investigate the genomic DNA regions from both the male and female snakes, which were identified to be sex specific or DNA regions of interest, identified from our previous experiments.

Random Amplified Polymorphic DNA Analysis

For all RAPD analyses performed, an unconventional source of DNA was used, in the form of discarded snake slough. To our knowledge, snake slough has never been used as a primary source of DNA for this type of molecular genetic research. A modification of an existing DNA extraction method was used to produce DNA of high quality and yield from these sloughs (Fetzner 1999). In this study, it was shown that DNA extracted from snake slough is quite suitable for RAPD analysis and therefore other PCR-based molecular genetic applications. It was also shown that, snake slough, when stored correctly, can produce high quantity DNA that is only slightly degraded even after three years of storage. From the personal interactions when collecting sloughs from reptile keepers and breeders and the quality of the resulting RAPD patterns indicates that using sloughs, as a source of DNA, has some advantages over most other methods, both invasive and non-invasive, provided the DNA fragment length required is not larger than approximately 10,000bp.

Of the 102 RAPD primers investigated, one potentially sex-associated RAPD marker was identified. This DNA region was shown to be 440bp in length and was produced using the RAPD primer OPA-17. DNA sequence analysis of the PCR fragment identified features similar to a retrotransposon element, and it contained three separate microsatellites, some

of which were shown to be polymorphic. The DNA marker was PCR amplified to a moderate intensity in 6 out of 7 female samples tested, while moderate and low intensity equivalent fragments were observed in male samples equally.

This suggests three possible reasons. The result produced was artifactual and simply caused by slight differences in template DNA concentrations between individual DNA samples (Handrys, Balick, and Schierwater 1992). Alternatively, the results could indicate that the DNA region is present in all individuals tested, but small genetic differences in the priming site of the DNA fragment or the copy number has resulted in different fragment profiles being produced from different individual's DNA templates tested (Mandolino et al. 1999; Sakamoto et al. 1995; Kafkas et al. 2015). This difference, which is most likely to be a single nucleotide polymorphism (SNP), may be female-associated and potentially linked to the W chromosome of *M. spilota*. Finally, the results could indicate the DNA region is X chromosome linked and the snake species *M. spilota* utilises a male heterogametic sex determination system. This result is in direct violation of our experimental hypothesis, but is supported by a recent published article, which gives evidence of two snake species belonging to the snake families Boidae and Pythonidae that most likely exhibit male heterogametic sex determination systems (Gamble et al. 2017).

The retrotransposon features observed in the DNA sequence of the 440bp DNA region suggests that the region could be present within the genome of *M. spilota* in multiple copies, further complicating the task of defining the region's association with sex (Finnegan 2012). An investigation of this region using fluorescent *in situ* hybridisation and comparing the patterns observed in male and female karyotypes would determine the region's chromosomal location and give an estimation of its genomic copy number and prevalence throughout the genome, including both autosomes and the proposed W chromosome. This technique has been applied previously to define W chromosomes specific repetitive elements identified in *Notechis scutatus* (O'Meally et al. 2010).

Representational Difference Analysis

The RDA experiments worked successfully to enrich DNA regions for both the male and female *M. spilota* genomes. Approximately 20 DNA regions were identified that showed potential sex-linked characteristics, 6 for the female genome and 14 for the male genome. PCR protocols were developed for DNA subsets of all 20 DNA regions identified but no

apparent differences were observed in the PCR patterns produced from male and female individuals.

DNA sequence analysis showed 7 of the 20 DNA regions identified contained retrotransposon related repetitive elements, having sequence similarity to multiple regions within the *P. bivittatus* reference annotation release 101. This suggested that these repetitive DNA regions may occur multiple times within the *M. spilota* genome and therefore have most likely been enriched within the RDA protocol because of either their polymorphic nature, their copy number difference or their ability to form complex secondary structures. Using DNA secondary structure predictive tools, we showed that at least one of these DNA regions, Female Hind III – Insert 1, likely formed a secondary structure during the RDA hybridisation protocol and may explain why it was selectively enriched by this process.

DNA sequence analysis of the remaining 13 DNA regions, identified 10 DNA regions most likely represented exonic and intronic DNA from 10 different genes. With the remaining 3 regions being representing non-coding DNA regions. Investigation into the genes that these DNA fragments represented, identified 5 of these genes had functions associated with sex differentiation, fertility and sex cell maturation, as summarised in Table 6-1.

Table 6-1 Gene with sexual function

DNA Region	**Representative Gene**	**Function**	**Gene Expression Levels**
Female Bam HI - Insert 1	*FSTL1*	Linked to sexual differentiation and sex reversal in fish	Highest in placenta
Male Hind III - Insert 1	*DNAH2*	Component of sperm flagella and has direct impact on male fertility	Highest in testis
Male Bgl II - Insert 4	*CCDC141*	Function unknown, mutation in gene linked to Kallmann syndrome, preventing puberty in humans	Highest in heart
Male Bam HI - Insert 2	*MBTD1*	Function unknown, stabilisation of mouse oocyte meiotic maturation	Highest in testis
Male Hind III - Insert 2	*SLAIN1*	Differentiation of embryonic stem cells in mice	Highest in testis and brain

Interestingly, the DNA regions Male Bam HI – Insert 2, which was proven to represent an intronic region of the Mbt Domain Containing 1 (*MBTD1*) gene, is located on chromosome 17 in humans at position 17q21.33. This chromosomal location is the same as the gene Lysine Acetyltransferase 7 (*KAT7*), which has previously been mapped to the Z chromosome in three snake species (Matsubara et al. 2012, 2006). In humans, the *MBTD1* and *KAT7* genes are separated by a relatively small distance of approximately 1.4Mb. Although the

presented PCR results showed that a subset of the DNA region was present in both male and female genomes, PCR products derived from the entire genomic region were not obtained for either sex.

The evidence suggests that the *MBTD1* gene is potentially located on the proposed Z chromosome in *M. spilota*. While, the identification of a potentially sex specific region of the gene using the RDA protocol gives support suggesting that this gene may be sex-linked in *M. spilota* and may be involved in snake sex determination. However, this gene region along with three of the five RDA regions, were identified only in the male RDA experiment. This experiment was designed to isolate and identified DNA regions that were unique to the male genome, suggesting that these regions are most likely male-specific or represent a male-specific allele. This finding is in direct opposition to our original thesis hypothesis and supports a suggestion that a male heterogametic sex chromosome system exists in *M. spilota*, as has been discovered in two other snake species (Gamble et al. 2017). Further research is required to elucidate this theory.

DNA sequencing was performed on 6 of the 7 DNA regions identified in our female-specific RDA protocol and the PCR products produced from both male and female genomes compared. The sequences produced from PCR amplifying a subset of the DNA region Female Bam HI - Insert 1, were consistent between all males and females tested. However, the sequences produced were different from the Female Bam HI - Insert 1 cloned and isolated RDA DNA fragment. This result suggests that the DNA region is represented in the female genome by two copies that differ slightly with respect to their DNA sequence, supporting our original thesis hypothesis. Moreover, this DNA region was found to represent both exonic and intronic regions of the Follistatin-Like Protein 1 (*FSTL1*) gene and this gene has been shown to have strong female bias expression in fish exhibiting sex reversal behaviour, indicating the gene may play a role in modifying ovarian functions (Casas et al. 2016). Giving support that the *FSTL1* gene is not simply sex-linked but might play a role in the sex determination system of *M. spilota*.

The DNA sequence produced for the DNA region Female Bgl II - Insert 1, showed strong sequence similarity between male and female genomes. However, a single SNP was identified, which was observed in all female sequences, in addition to the RDA sequence, as an adenine nucleotide, whilst all male sequences contained either cytosine or a thymine. Although this DNA region is most likely a non-coding DNA segment between the two genes

ADAM Metallopeptidase with Thrombospondin Type 1 Motif 17 (*ADAMTS17*) and Ceramide Synthase 3 (*CERS3*) the data supports that this SNP is female sex-linked and the DNA region may be located on a female specific chromosome in *M. spilota*.

Although, we were unable to absolutely conclude that the DNA regions identified from our RDA experiments were sex-linked. Many of the DNA regions that were enriched, were identified as representative of genes shown to have functions involved in sexual development, sex reversal and fertility. While two of the most promising regions identified appear to be female-linked and therefore support the original thesis hypothesis. One of the most promising DNA regions identified, representing a proportion of *MBTD1*, was found to be likely male-linked and located on the proposed Z chromosome of *M. spilota*. This observation is in direct opposition to our original thesis hypothesis.

Comparative Genomics

Using a comparative genomic approach, DNA representing regions for the 13 genes (*ATP6V0A1*, *CTNNB1*, *EIF1*, *GAD2*, *GH1*, *MRPL3*, *PIP4K2A*, *PTER*, *RAB5A*, *RUNDC1*, *SRI*, *TAX1BP1* and *WAC*) of the original 20 genes targeted, were identified and sequenced (Matsubara et al. 2006, 2012). The DNA sequence obtained totalled 16,045bp of the *M. spilota* genome, representing potential Z and W chromosomes. The genomic sequences, from many of the genes targeted, have never been previously sequenced in a snake species. All 13 gene regions are the first examples of these genes in our non-traditional model snake species *M. spilota*. Despite efforts, no sex-linked genomic DNA sequences were identified.

The Catenin beta-1 (*CTNNB1*) gene was investigated in the greatest depth of the 20 genes targeted and resulted in approximately 3023bp of DNA sequence, accounting for approximately 15% of the estimated gene size. It was observed that the *CTNNB1* gene is highly conserved between *M. spilota* and *P. bivittatus*, having a sequence identity of 95.6%. However, a 36bp DNA region located within intron 5 of the *CTNNB1* gene was also identified, which was found to be absent in *M. spilota*, but present in *P. bivittatus*. Further investigation showed that this 36bp region was also unable to be identified in *A. carolinensis* and *G. gallus*. This suggests that the 36bp region was most likely inserted into intron 5 of *CTNNB1* gene in *P. bivittatus*, relatively recently in evolutionary terms, sometime after *M. spilota* and *P. bivittatus* diverged.

These findings did not provide any support to our original hypothesis. However, the aim of this investigation was to produce genomic DNA sequence data that represented the 20 genes that had previously been localised to the Z chromosome of snakes. It was our intention to apply this DNA sequence information to identify, isolate and sequence large portions of the proposed snake Z and W chromosome using our Cosmid library.

Cosmid Genomic Library

We successfully created two large insert genomic libraries using a cosmid vector, representing multiple coverage of one female and one male *M. spilota* genome. Each cosmid library had at least 8-time coverage of each genome. The aim of this study was to screen these cosmid libraries to isolate and sequence large sections of the snake Z and W chromosome, using the sequence information obtained from the comparative genomic experiments. In addition, it was intended that any DNA regions of interest identified during our RAPD and RDA experiments were also to be used to screen the cosmid libraries.

Using a PCR based method, we isolated cosmid clones that were originally thought to contain parts of the *CTNNB1*, RAS-Associated Protein *RAB5A*, Member RAS Oncogene Family (*RAB5A*) and WW Domain Containing Adaptor with Coiled-Coil (*WAC*) genes. The production of false positive PCR products resulted in the cosmid clones being identified and isolated which did not actually contain the three regions being targeted. However, in hindsight, this PCR-based protocol applied to identify cosmid clones that contained the target genes should have included either endonuclease fragment mapping or some type of probe hybridisation, such as southern blotting, to provide an independent validation of the positive results.

In total 18 cosmid clones were sent for second generation sequencing, using either the Roche 454® and Ion Torrent sequencing platforms. The sequence data produced from the Roche 454® sequencing experiment was of low number and short length and 3 of the 10 library tags used were absent from the data produced. All results indicate that the Roche 454 ® sequencing preparation and particularly the construction of the sequencing libraries was poor.

The sequencing data produced from the Ion Torrent experiment was of a high quality with the longer than expected sequence lengths and higher than expected total sequence number. However, 98% of the sequencing results returned for 6 of the 8 cosmid clones in

the experiment was representative of only the SuperCos 1 vector and approximately 2% was representative of the actual cosmid insert. The resulting 2% sequence data was too sparse to identify the DNA regions or allow further investigation. In contrast, two of the eight cosmid clones returned good sequence coverage. Superficially, the poor sequencing result produced from 6 of the 8 cosmid clones possibly indicates that the original DNA extractions were contaminated with Supercos1 vector containing no insert.

Unfortunately, the false positive produced during the cosmid identification protocol together with issues relating to isolation meant that the DNA sequencing of DNA regions of interested were, in fact, DNA regions resulting from false positive PCR results. This result was disappointing because it adds very little to our final experimental hypothesis conclusions.

Limitation of the Thesis

When this project started in 2009, the greatest limitation of the study was the sheer lack of DNA sequence information known about snakes. In 2009 the most closely related reptile species to have its genome sequenced was *Anolis carolinensis* with its genome published in 2007. The initial proposal to sequence the genome of the first snake species, *P. bivittatus*, was published in 2010. The only DNA sequences representative of the snake sex chromosome system prior to then was the cDNA sequences published from Matsubara et al. 2006 study (Matsubara et al. 2006). This presented some unique challenges to the study, which shaped the way the thesis experiments were designed and lead to the original decision to use both the RAPD and RDA methods to investigate sex specific differences in *M. spilota* and to build on previous research and investigate the genomic DNA of all genes mapped to the snake sex chromosomes.

In the beginning, it was anticipated that the thesis project would yield multiple sex specific DNA regions from the genome of *M. spilota* and, because *M. spilota* does not have cytologically distinguishable sex chromosomes, any sex specific difference would likely be involved with, or near, the genes involved in snake sex determination. It was believed at the time that these genes would likely be conserved among other snake species. This approach made the thesis experiments extremely challenging when all potentially sex specific differences identified were unable to be proven to exist in only one sex from *M. spilota*.

In hindsight, it would have been useful to have used at least one additional snake species being concurrently tested in all experiments described in the thesis. The snake species *Notechis scutatus* (Tiger snake) is the most obvious candidate. *N. scutatus* is native to Western Australia and hence DNA samples would be readily available locally, but more importantly, it is a member of the Elapidae family and has been previously shown to have highly differentiated sex chromosomes and identified sex specific DNA regions (O'Meally et al. 2010). The results produced from investigating the genome of *N. scutatus* would have provided both validation of the experimental protocols and provided a direct comparison between the results produced from the two-snake species. This comparison would have provided a greater insight into the sex determination system of both snakes, representative of the two distinct snake groups, those with cytologically distinguishable sex chromosomes and those without and how the sex determination systems of these two groups of snakes relate to one another.

Impact of Findings and Recommendations for Future Research

The identification of a sex-linked marker is conditional of a genetic sex determination mechanism. In theory, not being able to prove the identification of a sex-linked marker provides evidence to suggest that there is no genetic difference between male and female *M. spilota*. This idea is supported by previous cytogenetic and limited genetic research, which shows that the potential Z and W sex chromosome of pythons and boas are indistinguishable, including *M. spilota* (Vicoso et al. 2013; Matsubara et al. 2012, 2006; O'Meally et al. 2010; Hoff, M. 2013). However, the absence of identified sex specific DNA does not prove that it does not exists in the *M. spilota* genome. Therefore, the evidence provided in this thesis suggests that sex specific DNA region in *M. spilota* probably represents a very small proportion of the genome.

Although this study was unable to prove or disprove the thesis hypothesis through the experimental data described in this thesis, our results have provided critical knowledge about the genomic DNA sequence of the proposed Z and W chromosome of *M. spilota* that is intended to be expand in the future. This research has provided some of the first genomic DNA sequences for the 13 different gene regions located on the proposed Z and W chromosome of *M. spilota*. It has investigated and identified one potentially sex-linked RAPD DNA marker. A new and improved RDA protocol has been developed, which has resulted in the identification of 20 DNA regions that are of interest. Six of which have been shown to be representative of genes that have function related to sexual development, sex

reversal and fertility, as well as one region very likely to be located on the Z chromosomes of snakes, making these six DNA regions likely to be sex-linked and potentially involved in the sex determination mechanisms of *M. spilota* and other snake species. It was unfortunate that we were unable to use our newly constructed male and female cosmid libraries to investigate the genome of *M. spilota* in greater depth. However, the construction of this new resource provides a valuable molecular genetics tool.

The advent of second and now third generation sequencing technology has had a major impact on how this type of research is conducted now and into the future. The process of identifying sex specific differences in any species, that has not previously been studied genetically, could be conducted using *de novo* whole genome sequencing and assembly followed by comparing the DNA sequences produced from male and female individuals. In the same way the RDA technique was used to enrich and isolate DNA regions that are sex specific from the genome of *M. spilota*, in-silico analysis of genomic sequencing could be used to isolate and identify DNA regions that are only contained in a single sex. This process would be advantageous because it would be relatively cheaper, easier, and faster than traditional molecular genetic methods and generate other sequence information that can be further mined. However, it would also inherit some of the same problems as older techniques, like differentiating between individual polymorphisms and sex specific differences and investigating the DNA of species without adequate reference genomes. Ultimately, all sex specific differences identified using in-silico analysis would have to be investigated and confirmed using more traditional molecular genetics techniques, but the largest most challenging step of the process, identifying unknown sex specific differences, would become much simpler with this type of approach.

References

Adams, Derek, Barry Larman, and Leif Oxburgh. 2007. "Developmental Expression of Mouse Follistatin-like 1 (Fstl1): Dynamic Regulation during Organogenesis of the Kidney and Lung." *Gene Expression Patterns* 7 (4): 491–500. https://doi.org/10.1016/j.modgep.2006.10.009.

Adelson, David L., Joy M. Raison, and Robert C. Edgar. 2009. "Characterization and Distribution of Retrotransposons and Simple Sequence Repeats in the Bovine Genome." *Proceedings of the National Academy of Sciences* 106 (31): 12855–60.

Agrawal, Veena, Kuldeep Sharma, Sarika Gupta, Ravindra Kumar, and Manoj Prasad. 2007. *Identification of Sex in Simmondsia Chinensis (Jojoba) Using PADP Markers*. Vol. 1. https://doi.org/10.1007/s11816-007-0031-6.

Al-Allaf, Faisal A, Oleg E Tolmachov, Lia Paola Zambetti, Viktoria Tchetchelnitski, and Huseyin Mehmet. 2013. "Remarkable Stability of an Instability-Prone Lentiviral Vector Plasmid in Escherichia Coli Stbl3." *3 Biotech* 3 (1): 61–70. https://doi.org/10.1007/s13205-012-0070-8.

Aloor, Jim, Kymberly Gowdy, Kathleen Azzam, Jennifer Madenspacher, David Draper, John Guardiola-Bright, Alex Merrick, Suraj Dhungana, and Michael Fessler. 2012. "Leucine Rich Repeats and Calponin Homology Containing Protein 4 (Lrch4) Is a Novel Regulator of Toll-like Receptor Signaling (68.8)." *The Journal of Immunology* 188 (1 Supplement): 68.8.

Alsop, Amber E., Pat Miethke, Ruth Rofe, Edda Koina, Natasha Sankovic, Janine E. Deakin, Helen Haines, Robert W. Rapkins, and Jennifer A. Marshall Graves. 2005. "Characterizing the Chromosomes of the Australian Model Marsupial Macropus Eugenii (Tammar Wallaby)." *Chromosome Research* 13: 627–36.

Arakawa, Y., C. Nishida-Umehara, Y.Matsuda, S.Sutou, and H.Suzuki. 2002. "X-Chromosomal Localization of Mammalian Y-Linked Genes in Two XO Species of the Ryukyu Spiny Rat." *Cytogenetic and Genome Research* 99: 303–9.

Asakawa, Shuichi, Izumi Abe, Yoshiki Kudoh, Noriyuki Kishi, Yimin Wang, Ryo Kubota, Jun Kudoh, Kazuhiko Kawasaki, Shinsei Minoshima, and Nobuyoshi Shimizu. 1997. "Human BAC Library: Construction and Rapid Screening." *Gene* 191 (1): 69–79. https://doi.org/10.1016/S0378-1119(97)00044-9.

Bachtrog, Doris, Emily Hom, Karen M. Wong, Xulio Maside, and Pieter de Jong. 2008. "Genomic Degradation of a Young Y Chromosome in Drosophila Miranda." *Genome Biology* 9 (2): R30. https://doi.org/10.1186/gb-2008-9-2-r30.

Bahy A. Ali, Tian-Hua Huang, Da-Nian Qin, and Xiao-Mei Wang. 2004. "A Review of Random Amplified Polymorphic DNA (RAPD) Markers in Fish Research." *Reviews in Fish Biology and Fisheries* 14: 443–53.

Baird, Nathan A., Paul D. Etter, Tressa S. Atwood, Mark C. Currey, Anthony L. Shiver, Zachary A. Lewis, Eric U. Selker, William A. Cresko, and Eric A. Johnson. 2008. "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers." *PLOS ONE* 3 (10): e3376. https://doi.org/10.1371/journal.pone.0003376.

Baker, R.J., James J. Bull, and G.A. Mengden. 1971. "Chromosomes of Elaphe Subocularis (Reptilia: Serpentes), with the Description of an in Vivo Technique for Preparation of Snake Chromosomes." *Experientia* 27 (10): 1228–29.

Banerjee, N. S., P. Manoj, and M. R. Das. 1999. "Male-Sex-Associated RAPD Markers in Piper Longum L." *Current Science* 77 (5): 693–95.

Barbieri, Christopher E., Sylvan C. Baca, Michael S. Lawrence, Francesca Demichelis, Mirjam Blattner, Jean-Philippe Theurillat, Thomas A. White, et al. 2012. "Exome Sequencing Identifies Recurrent SPOP, FOXA1 and MED12 Mutations in Prostate Cancer." *Nature Genetics* 44 (6): 685–89. https://doi.org/10.1038/ng.2279.

Bardakci, F, and D O F Skibinski. 1994. “Application of the RAPD Technique in Tilapia Fish: Species and Subspecies Identification.” *Heredity* 73 (July): 117.
Barker, David, and Tracy Barker. 2006. “Determining the Sex of Snakes.” 2006. http://www.vpi.com/publications/determining_the_sex_of_snakes.
Barton, N. H., and B. Charlesworth. 1998. “Why Sex and Recombination?” *Science* 281: 1986–90.
Becak, Willy, Marria Luiza Becak, H.R.S. Nazareth, and Susumu Ohno. 1964. “Close Karyological Kinship between the Reptilian Suborder Serpentes and the Class Aves.” *Chromosoma* 15: 606–17.
Bello, N., and A. Sanchez. 1999. “The Identification of Sex-Specific DNA Markers in the Ostrich Using Random Amplified Polymorphic DNA (RAPD) Assay.” *Molecular Ecology* 8: 667–69.
Belshaw, Robert, Vini Pereira, Aris Katzourakis, Gillian Talbot, Jan Pačes, Austin Burt, and Michael Tristem. 2004. “Long-Term Reinfection of the Human Genome by Endogenous Retroviruses.” *Proceedings of the National Academy of Sciences* 101 (14): 4894–99. https://doi.org/10.1073/pnas.0307800101.
Birk, Efrat, Adi Har-Zahav, Chiara M. Manzini, Metsada Pasmanik-Chor, Liora Kornreich, Christopher A. Walsh, Konrad Noben-Trauth, et al. 2010. “SOBP Is Mutated in Syndromic and Nonsyndromic Intellectual Disability and Is Highly Expressed in the Brain Limbic System.” *American Journal of Human Genetics* 87 (5): 694–700. https://doi.org/10.1016/j.ajhg.2010.10.005.
Brauner, Ian, Douglas B Spicer, Catherine E Krull, and Judith M Venuti. 2010. “Identification of Responsive Cells in the Developing Somite Supports a Role for B-Catenin-Dependent Wnt Signaling in Maintaining the DML Myogenic Progenitor Pool.” *Development Dynamics* 239: 222–36.
Bricker, J., L.M. Bushar, H.K. Reinert, and L. Gelbert. 1996. “Purification of High Quality DNA from Shed Skins.” *Herpetological Review* 27: 133–34.
Bull, James J. 1980. “Sex Determination in Reptiles.” *The Quarterly Review of Biology* 55: 3–21.
Bull, James J. 1983. *Evolution of Sex Determining Mechanisms*. The Benjamin/Cummings Publishing Company, Inc.
Burge, C., and S. Karlin. 1997. “Prediction of Complete Gene Structures in Human Genomic DNA.” *Journal of Molecular Biology* 268 (1): 78–94. https://doi.org/10.1006/jmbi.1997.0951.
Calderon, Alfredo, Adam Derr, Barden B. Stagner, Kenneth R. Johnson, Glen Martin, and Konrad Noben-Trauth. 2006. “Cochlear Developmental Defect and Background-Dependent Hearing Thresholds in the Jackson Circler (Jc) Mutant Mouse.” *Hearing Research* 221 (1): 44–58. https://doi.org/10.1016/j.heares.2006.07.008.
Campbell, Tessa N., and Francis Y.M. Choy. 2002. “Approaches to Library Screening.” *Journal of Molecular Microbiology and Biotechnology* 4 (6): 551–54.
Camper, Jeffrey D., and Brian G Hanks. 1995. “Variation in the Nucleolus Organizer Region among New World Snakes.” *Journal of Herpetology* 29 (3): 468–71.
Casas, Laura, Fran Saborido-Rey, Ryu Taewoo, Craig Michell, Timothy Ravasi, and Xabier Irigoien. 2016. “Sex Change in Clownfish: Molecular Insights from Transcriptome Analysis.” *Scientific Reports*, 1–19.
Castoe, Todd A., Kathryn T. Hall, Marcel L Guibotsy Mboulas, Wanjun Gu, Jason A.P. Koning, Samuel E. Fox, Alexander W. Poole, et al. 2011. “Discovery of Highly Divergent Repeat Landscapes in Snake Genomes Using High-Throughput Sequencing.” *Genome Biology Evolution* 3: 641–53.
Castoe, Todd A., A. P. Jason de Koning, Kathryn T. Hall, Daren C. Card, Drew R. Schield, Matthew K. Fujita, Robert P. Ruggiero, et al. 2013. “The Burmese Python Genome Reveals the Molecular Basis for Extreme Adaptation in Snakes.” *Proceedings of the*

National Academy of Sciences 110 (51): 20645–50. https://doi.org/10.1073/pnas.1314475110.
Castoe, Todd A, AP Jason de Koning, Kathryn T Hall, Ken D Yokoyama, Wanjun Gu, Eric N Smith, Cédric Feschotte, et al. 2011. "Sequencing the Genome of the Burmese Python (Python Molurus Bivittatus) as a Model for Studying Extreme Adaptations in Snakes." *Genome Biology* 12 (7): 406–406. https://doi.org/10.1186/gb-2011-12-7-406.
Ceplitis, Helene, and Hans Ellegren. 2004. "Adaptive Molecular Evolution of HINTW, a Female-Specific Gene in Birds." *Society of Molecular Biology and Evolution* 21: 249–54.
Chaly, Yury, Anthony D. Marinov, Leif Oxburgh, Daniel S. Bushnell, and Raphael Hirsch. 2012. "FSTL1 Promotes Arthritis in Mice by Enhancing Inflammatory Cytokine/Chemokine Expression." *Arthritis & Rheumatism* 64 (4): 1082–88. https://doi.org/10.1002/art.33422.
Chapelin, Catherine, Bénédicte Duriez, Fabrice Magnino, Michel Goossens, Estelle Escudier, and Serge Amselem. 1997. "Isolation of Several Human Axonemal Dynein Heavy Chain Genes: Genomic Structure of the Catalytic Site, Phylogenetic Analysis and Chromosomal Assignment." *FEBS Letters* 412 (2): 325–30. https://doi.org/10.1016/S0014-5793(97)00800-4.
Charlesworth, Brian. 1991. "The Evolution of Sex Chromosomes." *American Association for the Advancement of Science* 251: 1030–33.
Chassot, Anne-Amandine, Isabelle Gillot, and Marie-Christine Chaboissier. 2014. "R-Spondin1, WNT4, and the CTNNB1 Signaling Pathway: Strict Control over Ovarian Differentiation." *Reproduction (Cambridge, England)* 148 (6): R97-110. https://doi.org/10.1530/REP-14-0177.
Chen, Nian, and Xiao-Ping Lai. 2010. "Sequencing and analysis of the complete mitochondrial genome of the King Cobra, Ophiophagus hannah (Serpents: Elapidae)." *Yi chuan = Hereditas* 32 (7): 719–25.
Chen, Zheng, Mireille Montcouquiol, Rene Calderon, Nancy A. Jenkins, Neal G. Copeland, Matthew W. Kelley, and Konrad Noben-Trauth. 2008. "Jxc1/Sobp, Encoding a Nuclear Zinc Finger Protein, Is Critical for Cochlear Growth, Cell Fate, and Patterning of the Organ of Corti." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 28 (26): 6633–41. https://doi.org/10.1523/JNEUROSCI.1280-08.2008.
Chia-Feng, Liu, Nathan Bingham, Keith Parker, and Humphrey H.-C. Yao. 2008. "Sex-Specific Roles of b-Catenin in Mouse Gonadal Development." *Human Molecular Genetics* 18 (3): 405–17.
Clark, A. 1998. "Reptile Sheds Yield Quality DNA." *Herpetological Review* 29 (1): 17.
Clinton, Michael. 1998. "Sex Determination and Gonadal Development: A Bird's Eye View." *The Journal of Experimental Zoology* 281: 457–65.
Collins, J., and B. Hohn. 1978. "Cosmids: A Type of Plasmid Gene-Cloning Vector That Is Packageable in Vitro in Bacteriophage Lambda Heads." *Proceedings of the National Academy of Sciences* 75 (9): 4242–46.
Crew, F.A.E. 1933. "A Case of Non-Disjunction in the Fowl." *Proceedings of the Royal Society of Edinburgh* 53: 89–105.
Daniels, Danette, Michael Ford, Marie Schwinn, Hélène Benink, Matthew Galbraith, Ravi Amunugama, Richard Jones, et al. 2013. "Mutual Exclusivity of MED12/MED12L, MED13/13L, and CDK8/19 Paralogs Revealed within the CDK-Mediator Kinase Module." *Journal of Proteomics & Bioinformatics* S2 (July): 1–7. https://doi.org/10.4172/jpb.S2-004.

Dellinger, Thomas, and Gisela von Hegel. 1990. “Sex Identification through Cloacal Probing in Juvenile Marine Iguanas (Amblyrhynchus Cristatus).” *Journal of Herpetology* 24: 424–26.

Dewaele, Barbara, Joanna Przybyl, Anna Quattrone, Julio Finalet Ferreiro, Vanessa Vanspauwen, Ellen Geerdens, Valentina Gianfelici, et al. 2014. “Identification of a Novel, Recurrent MBTD1-CXorf67 Fusion in Low-Grade Endometrial Stromal Sarcoma.” *International Journal of Cancer* 134 (5): 1112–22. https://doi.org/10.1002/ijc.28440.

Donnison, Ian S, Jiri Siroky, Boris Vyskot, Heinz Saedler, and Sarah R Grant. 1996. “Isolation of Y Chromosome-Specific Sequences from Silene Latifolia and Mapping of Male Sex-Determining Genes Using Representational Difference Analysis.” *Genetics Society of America* 144: 1893–1901.

Eckert, Kristin A., and Thomas A. Kunkel. 1991. “DNA Polymerase Fidelity and the Polymerase Chain Reaction.” *Laboratory of Molecular Genetics* 1: 17–24.

Eguchi, Tomoko, and Yukinori Eguchi. 2000. “High Yield DNA Extraction from the Snake Cast-off Skin or Bird Feathers Using Collagenase.” *Biotechnology Letters* 22: 1097–1100.

Ellegren, Hans. 2001. “Hens, Cocks and Avian Sex Determination.” *European Molecular Biology Organization* 2: 192–96.

El-Mogharbel, Nisrine, Matthew Wakefield, Janine E. Deakin, Enkhjargal Tsend-Ayush, Frank Grützner, Amber Alsop, Tariq Ezaz, and Jennifer A. Marshall Graves. 2007. “DMRT Gene Cluster Analysis in the Platypus: New Insights into Genomic Organization and Regulatory Regions.” *Genomics* 89: 10–21.

Epplen, Joerg Thomas, John R. McCarry, Shizuyo Sutou, and Susumu Ohno. 1982. “Base Sequence of a Cloned Snake W-Chromosome DNA Fragment and Identification of a Male-Specific Putative MRNA in the Mouse.” *Genetics* 79: 3798–3802.

Eryilmaz, Jitka, Patricia Pan, Maria F. Amaya, Abdellah Allali-Hassani, Aiping Dong, Melanie A. Adams-Cioaba, Farrell MacKenzie, Masoud Vedadi, and Jinrong Min. 2009. “Structural Studies of a Four-MBT Repeat Protein MBTD1.” *PLOS ONE* 4 (10): e7274. https://doi.org/10.1371/journal.pone.0007274.

Evans, Glen A., Kathy Lewis, and Barry E. Rothenberg. 1989. *High Efficiency Vectors for Cosmid Microcloning and Genomic Analysis. Gene 79, 9-20*. Vol. 79. https://doi.org/10.1016/0378-1119(89)90088-7.

Evgen’ev, Michael B., Helena Zelentsova, Natalia Shostak, Marina Kozitsina, Victor Barskyi, Dirk-Henner Lankenau, and Victor G. Corces. 1997. “Penelope, a New Family of Transposable Elements and Its Possible Role in Hybrid Dysgenesis in Drosophila Virilis.” *Proceedings of the National Academy of Sciences* 94 (1): 196. https://doi.org/10.1073/pnas.94.1.196.

Ewert, Michael A., Chael A. Ewert, and Craig E. Nelson. 1991. “Sex Determination in Turtles: Diverse Patterns and Some Possible Adaptive Values.” *Copeia* 1991: 50–69.

Ezaz, Tariq, Benjamin Moritz, Paul Waters, Jennifer A. Marshall Graves, Arthur Georges, and Stephen D. Sarre. 2009. “The ZW Sex Microchromosomes of an Australian Dragon Lizard Share No Homology with Those of Other Reptiles or Birds.” *Chromosome Research* 17: 965–73.

Ezaz, Tariq, Alexander E. Quinn, Ikuo Miura, Stephen D. Sarre, Arthur Georges, and Jennifer A. Marshall Graves. 2005. “The Dragon Lizard Pogona Vitticeps Has ZZ/ZW Micro-Sex Chromosomes.” *Chromosome Research* 13: 763–76.

Ezaz, Tariq, Rami Stiglec, Fredreic Veyrunes, and Jennifer A. Marshall Graves. 2006. “Relationships between Vertebrate ZW and XY Sex Chromosome Systems.” *Current Biology* 16: 736–43.

Ezaz, Tariq, Nicole Valenzuela, Frank Grützner, Ikuo Miura, Arthur Georges, Russell L. Burke, and Jennifer A. Marshall Graves. 2006. “An XX/XY Sex Microchromosome System in

a Freshwater Turtle, Chelodina Longicollis (Testudines: Chelidae) with Genetic Sex Determination.” *Chromosome Research* 14 (2): 139–50.

Fagerberg, Linn, Bjorn M. Hallstrom, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. 2014. “Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics.” *Molecular & Cellular Proteomics: MCP* 13 (2): 397–406. https://doi.org/10.1074/mcp.M113.035600.

Feliciello, I., and G. Chinali. 1993. “A Modified Alkaline Lysis Method for the Preparation of Highly Purified Plasmid DNA from Escherichia Coli.” *Analytical Biochemistry* 212 (2): 394–401. https://doi.org/10.1006/abio.1993.1346.

Ferguson-Smith, M. 2006. “The Evolution of Sex Chromosomes and Sex Determination in Vertebrates and the Key Role of DMRT1.” *Sexual Development* 1: 2–11.

Ferrari, Simona, Vincent R.Harley, Andrea Pontiggia, Peter N.Goodfellow, Robin Lovell-Badge, and Marco E.Bianchi. 1992. “SRY, like HMG1, Recognizes Sharp Angles in DNA.” *The EMBO Journal* 11: 4497–4506.

Fetzner, James W. Jr. 1999. “Extracting High-Quality DNA from Shed Reptile Skins: A Simplified Method.” *BioTechniques* 26: 1052–54.

Finnegan, David J. 2012. “Retrotransposons.” *Current Biology* 22 (11): R432–37. https://doi.org/10.1016/j.cub.2012.04.025.

Fitzgibbon, J., M.J. Neat, L. Jones, N. Foot, T.A. Lister, and R.K. Gupta. 2000. “Assignment1 of B-Cell Lymphoma 6, Member B (Zinc Finger Protein) Gene (BCL6B) to Human Chromosome 17p13.1 by in Situ Hybridization.” *Cytogenetic and Genome Research* 89 (3–4): 218–19.

Foster, Jamie W., Francine E. Brennan, Gregory K. Hampikian, Peter N. Goodfellow, Andrew H. Sinclair, Robin Lovell-Badge, Lynne Selwood, Marilyn B. Renfree, Desmond W. Cooper, and Jennifer A. Marshall Graves. 1992. “Evolution of Sex Determination and the Y Chromosome: SRY-Related Sequences in Marsupials.” *Nature* 359: 531–33.

Foster, Jamie W., and Jennifer A. Marshall Graves. 1994. “An SRY-Related Sequence on the Marsupial X Chromosome: Implications for the Evolution of the Mammalian Testis-Determining Gene.” *Genetics* 91: 1927–31.

Franck A. Atienzar, Frank A., and Awadhesh N. Jha. 2006. “The Random Amplified Polymorphic DNA (RAPD) Assay and Related Techniques Applied to Genotoxicity and Carcinogenesis Studies: A Critical Review.” *Mutation Research* 613: 76–102.

Fridolfsson, Anna-Karin, Hans Cheng, Neal G. Copeland, Nancy A. Jenkins, Hsiao-Ching Liu, Trerje Raudsepp, Trevor Woodage, Bhanu Chowdhary, Joy Halverson, and Hans Ellegren. 1998. “Evolution of the Avian Sex Chromosomes from an Ancestral Pair of Autosomes.” *Genetics* 95: 8417–8152.

Fuchs, J, J-B Fiche, A Buhot, R Calemczuk, and T Livache. 2010. “Salt Concentration Effects on Equilibrium Melting Curves from DNA Microarrays.” *Biophysical Journal* 99 (6): 1886–95. https://doi.org/10.1016/j.bpj.2010.07.002.

Fulton, Lucinda A., Elaine R. Mardis, Richard K. Wilson, and International Chicken Genome Sequencing Consortium. 2004. “Sequence and Comparative Analysis of the Chicken Genome Provide Unique Perspectives on Vertebrate Evolution.” *Nature* 432: 695–716.

Gamble, Tony, Todd Castoe, Stuart Nielsen, Jaison L. Banks, Daren Card, Drew Schield, Gordon Schuett, and Warren Booth. 2017. *The Discovery of XY Sex Chromosomes in a Boa and Python*. Vol. 27. https://doi.org/10.1016/j.cub.2017.06.010.

Geng, Yan, Yingying Dong, Mingyan Yu, Long Zhang, Xiaohua Yan, Jingxia Sun, Long Qiao, et al. 2011. “Follistatin-like 1 (Fstl1) Is a Bone Morphogenetic Protein (BMP) 4 Signaling Antagonist in Controlling Mouse Lung Development.” *Proceedings of the*

National Academy of Sciences 108 (17): 7058. https://doi.org/10.1073/pnas.1007293108.

Geraldes, Armando, Teri Rambo, Rod A. Wing, Nuno Ferrand, and Michael W. Nachman. 2010. “Extensive Gene Conversion Drives the Concerted Evolution of Paralogous Copies of the SRY Gene in European Rabbits.” *Molecular Biology and Evolution* 27 (11): 2437–40.

Gill, G. P., C. F. Harvey, and L. G. Fraser. 1998. “Development of Sex-Linked PCR Markers for Gender Identification in Actinidia.” *Theoretical and Applied Genetics* 97: 439–45.

Gise, Alexander von, Bin Zhou, Leah B. Honor, Qing Ma, Anna Petryk, and William T. Pu. 2011. “WT1 Regulates Epicardial Epithelial to Mesenchymal Transition through β-Catenin and Retinoic Acid Signaling Pathways.” *Developmental Biology* 356 (2): 421–31. https://doi.org/10.1016/j.ydbio.2011.05.668.

Gorinsek, Benjamin, Franc Gubensek, and Dusan Kordis. 2004. “Evolutionary Genomics of Chromoviruses in Eukaryotes.” *Molecular Biology and Evolution* 21 (5): 781–98.

Grafodatskaya, Daria, Willem Rens, Mary C. Wallis, Vladimir Trifonov, Patricia C. M. O’Brien, Oliver Clarke, Jennifer A. M. Graves, and Malcolm A. Ferguson-Smith. 2007. “Search for the Sex-Determining Switch in Monotremes: Mapping WT1, SF1, LHX1, LHX2, FGF9, WNT4, RSPO1 and GATA4 in Platypus.” *Chromosome Research* 15: 777–85.

Graves, Jennifer A. Marshall. 1998. “Interactions between SRY and SOX Genes in Mammalian Sex Determination.” *BioEssays* 20: 264–69.

Graves, Jennifer A. Marshall. 2003. “Sex and Death in Birds: A Model of Dosage Compensation That Predicts Lethality of Sex Chromosome Aneuploids.” *Cytogenetic and Genome Research* 101: 278–82.

Graves, Jennifer A. Marshall. 2006. “Sex Chromosome Specialization and Degeneration in Mammals.” *Cell* 124: 901–14.

Graves, Jennifer A. Marshall. 2008. “Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes.” *The Annual Review of Genetics* 42: 565–86.

Gregory, Ryan T, James A Nicol, Heidi Tamm, Bellis Kullman, Kaur Kullman, I.J. Leitch, B.G. Murray, D.F. Kapraun, J. Greilhuber, and M.D. Bennett. 2007. “Eukaryotic Genome Size Databases.” *Nucleic Acids Research* 35: 332–38.

Griffiths, Richard, Mike C. Double, Kate Orr, and Robert J.G. Dawson. 1998. “A DNA Test to Sex Most Birds.” *Molecular Ecology* 7: 1071–75.

Griffiths, Richard, and Bela Tiwari. 1993. “The Isolation of Molecular Genetic Markers for the Identification of Sex.” *National Academy of Science* 90: 8324–26.

Groot, TVM, E Bruins, and JAJ Breeuwer. 2003. “Molecular Genetic Evidence for Parthenogenesis in the Burmese Python, Python Molurus Bivittatus.” *Nature* 90: 130–35.

Grutzner, Frank, Willern Rens, Enkhjargal Tsend-Ayush, Nisrine El-Mogharbel, Patricia C. M. O’Brien, Russel C. Jones, Malcolm A. Ferguson-Smith, and Jennifer A. Marshall Graves. 2004. “In the Platypus a Meiotic Chain of Ten Sex Chromosomes Shares Genes with the Bird Z and Mammalian X Chromosomes.” *Nature* 432: 931–37.

Grzebelus, Dariusz, Slawomir Lasota, Tomasz Gambin, Gregory Kucherov, and Anna Gambin. 2007. “Diversity and Structure of PIF/Harbinger-like Elements in the Genome of Medicago Truncatula.” *BMC Genomics* 8 (November): 409. https://doi.org/10.1186/1471-2164-8-409.

Hadany, Lilach, and Tuvik Beker. 2007. “Sexual Selection and the Evolution of Obligatory Sex.” *BMC Evolutionary Biology* 7: 1–7.

Handrys, H, M Balick, and B Schierwater. 1992. “Applications of Random Amplified Polymorphic DNA (RAPD) in Molecular Ecology.” *Molecular Ecology* 1 (1): 55–63.

Harley, Vincent R., David I. Jackson, Patrick J. Hextall, J. Ross Hawkins, Gary D. Berkovitz, Shanthini Sockanthan, Robin Lovell-Badge, and Peter N. Goodfellow. 1992. “DNA

Binding Activity of Recombinant SRY from Normal Males and XY Females." *Science* 255: 453–56.

Haseltine FP, DePonte KK, Breg WR, and Genel M. 1982. "Presence of H-Y Antigen in Patients with Ullrich-Turner Syndrome and X-Chromosome Rearrangements." *American Journal of Medical Genetics* 11 (1): 97–107.

Hawkins, J. R., A. Taylor, P. Berta, J. Levilliers, B. Van der Auwera, and P. N. Goodfellow. 1992. "Mutational Analysis of SRY: Nonsense and Missense Mutations in XY Sex Reversal." *Human Genetics* 88: 471–74.

Hirst, Claire E., Elizabeth S. Ng, Lisa Azzola, Anne K. Voss, Tim Thomas, Edouard G. Stanley, and Andrew G. Elefanty. 2006. "Transcriptional Profiling of Mouse and Human ES Cells Identifies SLAIN1, a Novel Stem Cell Gene." *Developmental Biology* 293 (1): 90–103. https://doi.org/10.1016/j.ydbio.2006.01.023.

Hoff, M. 2013. "Slithering Toward Clarity: Snakes Shed New Light on the Evolution and Function of Sex Chromosomes." *PLOS Biology* 11 (8): 1–2.

Hoffmann-La, F. 2010. "Roche's Newly Launched Next-Generation DNA Sequencer Opens Innovation in Genomic Research to Broad Scientific Community." Investor Update. https://www.roche.com/investors/updates/inv-update-2010-05-27.htm.

Hohn, B, and K Murray. 1977. "Packaging Recombinant DNA Molecules into Bacteriophage Particles in Vitro." *Proceedings of the National Academy of Sciences* 74 (8): 3259–63.

Hori, Tetsuya, Shuichi Asakawa, Yuichiro Itoh, Nobuyoshi Shimizu, and Shigeki Mizuno. 2000. "Wpkci, Encoding an Altered Form of PKCI, Is Conserved Widely on the Avian W Chromosome and Expressed in Early Female Embryos: Implication of Its Role in Female Sex Determination." *Molecular Biology of the Cell* 11: 3645–60.

Hormaza, J. I, L Dollo, and V. S Polito. 1994. "Identification of a RAPD Marker Linked to Sex Determination in Pistacia Vera Using Bulked Segregant Analysis." *Theoretical and Applied Genetics* 89 (9): 9–13.

Hoser, Raymond T. 1999. "Hybridisation in Carpet Snakes Genus: Morelia (Serpentes: Pythoninae) and Other Australian Pythons." *Herptile* 24 (2): 61–67.

Hu, Fangzhou, Kang Xu, Yunfan Zhou, Chang Wu, Shi Wang, Jun Xiao, Min Wen, et al. 2017. "Different Expression Patterns of Sperm Motility-Related Genes in Testis of Diploid and Tetraploid Cyprinid Fish." *Biology of Reproduction* 96 (4): 907–20. https://doi.org/10.1093/biolre/iox010.

Hutchins, B. Ian, L. Damla Kotan, Carol Taylor-Burds, Yusuf Ozkan, Paul J. Cheng, Fatih Gurbuz, Jean D. R. Tiong, et al. 2016. "CCDC141 Mutation Identified in Anosmic Hypogonadotropic Hypogonadism (Kallmann Syndrome) Alters GnRH Neuronal Migration." *Endocrinology* 157 (5): 1956–66. https://doi.org/10.1210/en.2015-1846.

"Ion PGM^{TM} System Specifications." n.d. ThermoFisher Scientific. https://www.thermofisher.com/au/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-workflow/ion-torrent-next-generation-sequencing-run-sequence/ion-pgm-system-for-next-generation-sequencing/ion-pgm-system-specifications.html.

Irving, Julie, Cheng-Han Lee, Stephen Yip, Oliva Esther, Glen. W McCluggage, and Robert Young. 2015. "Microcystic Stromal Tumor: A Distinctive Ovarian Sex Cord-Stromal Neoplasm Characterized by FOXL2, SF-1, WT-1, Cyclin D1, and β-Catenin Nuclear Expression and CTNNB1 Mutations." *American Journal of Surgical Pathology* 39 (10): 1420–26.

Israel, D I. 1993. "A PCR-Based Method for High Stringency Screening of DNA Libraries." *Nucleic Acids Research* 21 (11): 2627–31.

Iturra, P., J. F. Medrano, M. Bagley, N. Lam, N. Vergara, and J. C. Marin. 1997. "Identification of Sex Chromosome Molecular Markers Using RAPDs and

Fluorescent in Situ Hybridization in Rainbow Trout." *Genetica* 101 (3): 209–13. https://doi.org/10.1023/A:1018371623919.
IUCN. 2010. "IUCN Red List of Threatened Species." 2010. http://www.iucnredlist.org.
Jagar, Ralf J., Maria Anvret, Kerstin Hall, and Gerd Scherer. 1990. "A Human XY Female with a Frame Shift Mutation in the Candidate Gene SRY." *Nature* 348: 452–54.
Janzen, F. J., and P. C. Phillips. 2006. "Exploring the Evolution of Environmental Sex Determination Especially in Reptiles." *European Society for Evolutionary Biology* 19: 1775–84.
Janzen, Fredric J, and Gary L Paukstis. 1991. "Environmental Sex Determination in Reptiles: Ecology, Evolution, and Experimental Design." *The Quarterly Review of Biology* 66: 149–79.
Jones, PH, L Wallace, and HB Bitten. 2009. "Isolation and Characterization of 11 Microsatellite Loci from Oropsylla Hirsuta, a Vector of Sylvatic Plague." *Molecular Ecology Resources* 9 (3): 1041–44.
Jose L. Horreo, Maria L. Pelaez, and Patrick S. Fitze. 2015. "Skin Sheds as a Useful DNA Source for Lizard Conservation." *Phyllomedusa* 14 (1): 73–77.
Just, W., A. Baumstark, A. Süß, A. Graphodatsky, W. Rens, N. Schäfer, I. Bakloushinskaya, H. Hameister, and W. Vogel. 2007. "Ellobius Lutescens: Sex Determination and Sex Chromosome." *Sexual Development* 1: 211–21.
Just, Walter, Wolfgang Rau, Walther Vogel, Mikhail Akhverdian, Karl Fredga, Jennifer A. Marshall Graves, and Elena Lyapunova. 1995. "Absence of Sry in Species of the Vole Ellobius." *Nature Genetics* 11: 117–18.
Kafkas, Salih S. K., Selim G. D. Cetiner, Rafael Perl-Treves, and Ada A. N. Nissim-Levi. 2015. "Development of Sex-Associated RAPD Markers in Wild Pistacia Species." *The Journal of Horticultural Science and Biotechnology* 76 (2): 242–46.
Kaymaz, Yasin, Cliff I. Oduor, Hongbo Yu, Juliana A. Otieno, John Michael Ong'echa, Ann M. Moormann, and Jeffrey A. Bailey. 2017. "Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences." *Molecular Cancer Research: MCR* 15 (5): 563–76. https://doi.org/10.1158/1541-7786.MCR-16-0305.
Keyvanshokooh, S., M. Pourkazemi, and M. R. Kalbassi. 2007. "The RAPD Technique Failed to Identify Sex-Specific Sequences in Beluga (Huso Huso)." *Journal of Applied Ichthyology* 23: 1–2.
Kiel-Metzger, Karen, Gwynedd Warren, Golder N. Wilson, and Robert P. Erickson. 1985. "Evidence That the Human Y Chromosome Does Not Contain Clustered DNA Sequences (BKM) Associated with Heterogametic Sex Determination in Other Vertebrates." *The New England Journal of Medicine* 313 (4): 242–45.
Knisbacher, Binyamin A., and Erez Y. Levanon. 2016. "DNA Editing of LTR Retrotransposons Reveals the Impact of APOBECs on Vertebrate Genomes." *Molecular Biology and Evolution* 33 (2): 554–67. https://doi.org/10.1093/molbev/msv239.
Koopman, Peter, John Gubbay, Nigel Vivian, Peter Goodfellow, and Robin Lovell-Badge. 1991. "Male Development of Chromosomally Female Mice Transgenic for Sry." *Nature* 351: 117–21.
Koopman, Peter, Andrea Musterberg, Blanche Capel, Nigel Vivian, and Robin Lovell-Badge. 1990. "Expression of a Candidate Sex-Determining Gene during Mouse Testis Differentiation." *Nature* 348: 450–52.
Kudo-Saito, Chie, Takafumi Fuwa, Kouichi Murakami, and Yutaka Kawakami. 2013. "Targeting FSTL1 Prevents Tumor Bone Metastasis and Consequent Immune Dysfunction." *Cancer Research* 73 (20): 6185. https://doi.org/10.1158/0008-5472.CAN-13-1364.
Kumar, Sudhir, and S. Blair Hedges. 1998. "A Molecular Timescale for Vertebrate Evolution." *Nature* 392: 917–20.

Lamar, Ellen E., and Edward Palmer. 1984. "Y-Encoded, Species-Specific DNA in Mice: Evidence That the Y Chromosome Exists in Two Polymorphic Forms in Inbred Strains." *Cell* 37: 171–77.

Lang, Jeffrey W., and Harry V. Andrews. 1994. "Temperature-Dependent Sex Determination in Crocodilians." *The Journal of Experimental Zoology* 270: 28–44.

Laopichienpong, Nararat, Narongrit Muangmai, Lawan Chanhome, Sunutcha Suntrarachun, Panupon Twilprawat, Surin Peyachoknagul, and Kornsorn Srikulnath. 2017. "Evolutionary Dynamics of the Gametologous CTNNB1 Gene on the Z and W Chromosomes of Snakes." *The Journal of Heredity* 108 (2): 142–51. https://doi.org/10.1093/jhered/esw074.

Leon Huynen, Craig D. Millar, and David M. Lambert. 2002. "A DNA Test to Sex Ratite Birds." *Molecular Ecology* 11: 851–56.

Li, Ningbo, Guangyou Duan, Jiaoli Sun, Shanna Guo, Qingli Wang, Hua Zheng, Cong Wang, Changmao Zhu, Yi Liu, and Xianwei Zhang. 2018. "Associations between Single-Nucleotide Polymorphisms in the NTRK1 Gene and Basal Pain Sensitivity in Young Han Chinese Women." *Neuroscience Letters* 662 (January): 312–17. https://doi.org/10.1016/j.neulet.2017.10.029.

Li, Y., J. A. Hill, G. H. Yue, F. Chen, and L. Orban. 2002. "Extensive Search Does Not Identify Genomic Sex Markers in Tetraodon Nigroviridis." *Journal of Fish Biology* 60: 1–4.

Lisitsyn, N, N Lisitsyn, and M Wigler. 1993. "Cloning the Differences between Two Complex Genomes." *Science* 259: 946–951.

Liu, Jie, Qi Zheng, Yiqun Deng, Chao-Sheng Cheng, Neville R Kallenbach, and Min Lu. 2006. "A Seven-Helix Coiled Coil." *Proceedings of the National Academy of Sciences* 103 (42): 15457–62. https://doi.org/10.1073/pnas.0604871103.

Liu, Lin, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law. 2012. "Comparison of Next-Generation Sequencing Systems." *Journal of Biomedicine and Biotechnology* 2012: 11.

Llorens, C, R Futami, L Covelli, L Dominguez-Escriba, J. M. Viu, D Tamarit, J Aguilar-Rodriguez, et al. 2011. "The Gypsy Database (GyDB) of Mobile Genetic Elements." GyDB Project. 2011. http://gydb.org/index.php/Intro.

Loman, Nicholas J, Raju V Misra, Timothy J Dallman, Chrystala Constantinidou, Saheer E Gharbia, John Wain, and Mark J Pallen. 2012. "Performance Comparison of Benchtop High-Throughput Sequencing Platforms." *Nature Biotechnology* 30 (April): 434.

Lovsin, N., F. Gubensek, and D. Kordi. 2001. "Evolutionary Dynamics in a Novel L2 Clade of Non-LTR Retrotransposons in Deuterostomia." *Molecular Biology and Evolution* 18 (12): 2213–24. https://doi.org/10.1093/oxfordjournals.molbev.a003768.

Lundberg, K. S., D. D. Shoemaker, M. W. Adams, J. M. Short, J. A. Sorge, and E. J. Mathur. 1991. "High-Fidelity Amplification Using a Thermostable DNA Polymerase Isolated from Pyrococcus Furiosus." *Gene* 108 (1): 1–6.

Luo, Yi-Bo, Jun-Yu Ma, Qing-Hua Zhang, Fei Lin, Zhong-Wei Wang, Lin Huang, Heide Schatten, and Qing-Yuan Sun. 2013. "MBTD1 Is Associated with Pr-Set7 to Stabilize H4K20me1 in Mouse Oocyte Meiotic Maturation." *Cell Cycle* 12 (7): 1142–50. https://doi.org/10.4161/cc.24216.

Macdonald, Stewart. 2017. "The Australian Reptile Online Database." 2017. http://www.arod.com.au/arod/.

Maiti, Amit K, Marie-Geneviève Mattéi, Mark Jorissen, Armin Volz, Andreas Zeigler, and Patrice Bouvagnet. 2000. "Identification, Tissue Specific Expression, and Chromosomal Localisation of Several Human Dynein Heavy Chain Genes." *European Journal of Human Genetics* 8 (December): 923.

Mandolino, G., A. Carboni, S. Forapani, V. Faeti, and P. Ranalli. 1999. "Identification of DNA Markers Linked to the Male Sex in Dioecious Hemp (Cannabis Sativa L.)."

Theoretical and Applied Genetics 98 (1): 86–92. https://doi.org/10.1007/s001220051043.

Martinez E. A., Destombe C., Quillet M. C., and Valero M. 2003. "Identification of Random Amplified Polymorphic DNA (RAPD) Markers Highly Linked to Sex Determination in the Red Alga Gracilaria Gracilis." *Molecular Ecology* 8 (9): 1533–38. https://doi.org/10.1046/j.1365-294x.1999.00721.x.

Mashima, Jun, Yuichi Kodama, Takatomo Fujisawa, Toshiaki Katayama, Yoshihiro Okuda, Eli Kaminuma, Osamu Ogasawara, Kousaku Okubo, Yasukazu Nakamura, and Toshihisa Takagi. 2017. "DNA Data Bank of Japan." *Nucleic Acids Research* 45 (D1): D25–31. https://doi.org/10.1093/nar/gkw1001.

Matsubara, Kazumi, Shigehiro Kuraku, Hiroshi Tarui, Osamu Nishmura, Chizuko Nishida, Kiyokaza Agata, Yoshinori Kumazawa, and Yoichi Matuda. 2012. "Intra-Genomic GC Heterogeneity in Sauropsids: Evolutionary Insights from CDNA Mapping and GC3 Profiling in Snake." *BMC Genomics* 13 (604): 1–14.

Matsubara, Kazumi, Hiroshi Tarui, Michihisa Toriba, Kazuhiko Yamada, Chizuko Nishida-Umehara, Kiyokazu Agata, and Yoichi Matsuda. 2006. "Evidence for Different Origin of Sex Chromosomes in Snakes, Birds, and Mammals and Step-Wise Differentiation of Snake Sex Chromosomes." *Proceedings of the National Academy of Sciences* 103: 18190–95.

Matsuda, Yoichi, Chizuko Nishida-Umehara, Hiroshi Tarui, Asato Kuroiwa, Kazuhiko Yamada, Taku Isobe, Junko Ando, et al. 2005. "Highly Conserved Linkage Homology between Birds and Turtles: Bird and Turtle Chromosomes Are Precise Counterparts of Each Other." *Chromosome Research* 13: 601–15.

Mattey, R., and J.M. van Brick. 1957. "Sex Chromosomes in Aminiota." *Evolution* 11: 163–65.

Maxine P. Piggott, and Andrea C. Taylor. n.d. "Remote Collection of Animal DNA and Its Application in Conservation Management and Understanding the Population Biology of Rare and Cryptic Species." *Wildlife Research* 30 (1): 1–13.

McElreavy, Ken, Eric Vilain, Nacer Abbas, Ira Herskowitz, and Marc Fellous. 1993. "A Regulatory Cascade Hypothesis for Mammalian Sex Determination: SRY Represses a Negative Regulator of Male Development." *Genetics* 90: 3368–72.

McGowan, C., and W. S. Davidson. 1998. "The RAPD Technique Fails to Detect a Male-Specific Genetic Marker in Atlantic Salmon." *Journal of Fish Biology* 53 (5): 1134–36.

McQueen, Heather A., and Michael Clinton. 2009. "Avian Sex Chromosomes: Dosage Compensation Matters." *Chromosome Research* 17: 687–97.

Mengden, Gregory A. 1981. "Linear Differentiation of the C-Band Pattern of the W Chromosome in Snakes and Birds." *Chromosoma* 83: 275–87.

Mengden, Gregory A., and Dean A. Stock. 1980. "Chromosomal Evolution in Serpentes; a Comparison of G and C Chromosome Banding Patterns of Some Colubrid and Boid Genera." *Chromosoma* 79 (1): 53–64.

Michelmore, R. W., I. Paran, and R. V. Kesseli. 1991. "Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregrant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations." *Proceedings of the National Academy of Sciences* 88: 9828–32.

Miller, Michael R, Joseph P Dunham, Angel Amores, William A Cresko, and Eric A Johnson. 2007. "Rapid and Cost-Effective Polymorphism Identification and Genotyping Using Restriction Site Associated DNA (RAD) Markers." *Genome Research* 17 (2): 240–48. https://doi.org/10.1101/gr.5681207.

Modi, Williams S, and David Crews. 2005. "Sex Chromosomes and Sex Determination in Reptiles." *Current Opinion in Genetics & Development* 15: 660–65.

Morin, PJ. 1999. "Beta-Catenin Signaling and Cancer." *BioEssays* 12: 1021–30.

Muller, H.J. 1914. “A Gene for the Fourth Chromosome of Drosophila.” *Journal of Experimental Zoology* 17: 325–36.

Muller U. 1996. “H-Y Antigens.” *Human Genetics* 97 (6): 701–4.

Murtagh, Veronica J., Paul D. Waters, and Jennifer A. Marshall Graves. 2010. *Compact but Complex - The Marsupial Y Chromosome*. Vol. 4. Marsupial Genetics and Genomics.

Na, Kiyong, Eun Kyung Kim, Wonjun Jang, and Hyun-Soo Kim. 2017. “CTNNB1 Mutations in Ovarian Microcystic Stromal Tumors: Identification of a Novel Deletion Mutation and the Use of Pyrosequencing to Identify Reported Point Mutation.” *Anticancer Research* 37 (6): 3249–58. https://doi.org/10.21873/anticanres.11688.

Nakashima, K, I Nobuhisa, M Deshimaru, T Ogawa, Y Shimohigashi, Y Fukumaki, M Hattori, Y Sakaki, S Hattori, and M Ohno. 1995. “Structures of Genes Encoding TATA Box-Binding Proteins from Trimeresurus Gramineus and T. Flavoviridis Snakes.” *Gene* 152 (2): 209–13.

Nanda, I., T. Haaf, M. Schartl, M. Schmid, and D.W. Burt. 2002. “Comparative Mapping of Z-Orthologous Genes in Vertebrates: Implications for the Evolution of Avian Sex Chromosomes.” *Cytogenetic and Genome Research* 99: 178–84.

Nanda, Indrajit, Zhihong Shan, Manfred Schartl, Dave W. Burt, Michael Koehler, Hans-Gerd Nothwang, Frank Grutzner, et al. 1999. “300 Million Years of Conserved Synteny between Chicken Z and Human Chromosome 9.” *Nature Genetics* 21: 258–59.

NCBI Resource Coordinators. 2017. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research* 45 (Database issue): D12–17. https://doi.org/10.1093/nar/gkw1071.

O’Brien, SJ, JF Eisenberg, M Miyamoto, SB Hedges, S Kumar, DE Wilson, M Menotti-Raymond, et al. 1999. “Genome Maps 10. Comparative Genomics. Mammalian Radiations. Wall Chart.” *Science* 286: 463–78.

Ohno, Susumu. 1966. *Sex Chromosomes and Sex-Linked Genes*. Monographs on Endocrinology. Springer-Verlag Berlin Heidelberg.

Olmo, Ettore, and Giorgio Gelo Signorino. 2010. “CHROMOREP: A Reptiles Chromosomes Database.” 2010. http://ginux.univpm.it/scienze/chromorep/.

O’Meally, Denis, Hardip R. Patel, Rami Stiglec, Stephen D. Sarre, Arthur Georges, Jennifer A. Marshall Graves, and Tariq Ezaz. 2010. “Non-Homologous Sex Chromosomes of Birds and Snakes Share Repetitive Sequences.” *Chromosome Research* 18: 787–800.

O’Neill, Michael, Michele Biinder, Craig Smith, Jane Andrews, Kirsty Reed, Matthijs Smith, Craig Millar, David Lambert, and Andrew Sinclair. 2000. “ASW: A Gene with Conserved Avian W-Linkage and Female Specific Expression in Chick Embryonic Gonad.” *Development Genes and Evolution* 210: 243–49.

Orsulic, S, and M Peifer. 1996. “An in Vivo Structure-Function Study of Armadillo, the Beta-Catenin Homologue, Reveals Both Separate and Overlapping Regions of the Protein Required for Cell Adhesion and for Wingless Signaling.” *The Journal of Cell Biology* 134 (5): 1283. https://doi.org/10.1083/jcb.134.5.1283.

Ouimet, Mireille, Elizabeth J Hennessy, Coen van Solingen, Graeme J Koelwyn, Maryem A Hussein, Bhama Ramkhelawon, Katey J Rayner, et al. 2016. “MiRNA Targeting of Oxysterol Binding Protein-like 6 (OSBPL6) Regulates Cholesterol Trafficking and Efflux.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 36 (5): 942–51. https://doi.org/10.1161/ATVBAHA.116.307282.

Pace, Helen C, and Charles Brenner. 2003. “Feminizing Chicks: A Model for Avian Sex Determination Based on Titration of Hint Enzyme Activity and the Predicted Structure of an Asw-Hint Heterodimer.” *Genome Biology* 4: R18.

Page, David C., Rebecca Mosher, Elizabeth M. Simpson, Elizabeth M.C. Fisher, Graeme Mardon, Jonathan Pollack, Barbara McGillivray, Albert de la Chapelle, and Laura G. Brown. 1987. “The Sex-Determining Region of the Human Y Chromosome Encodes a Finger Protein.” *Cell* 51: 1091–1104.

"Paired End Library Preparation Method Manual, 20 Kb and 8 Kb Span GS Junior Titanium Series." 2012. 454 Life Science Corporation.
Palmer, M. S., A. H. Sinclair, P. Berta, N. A. Ellis, P. N. Goodfellow, N. E. Abbas, and M. Fellous. 1989. "Genetic Evidence That ZFY Is Not the Testis-Determining Factor." *Nature* 342: 937–39.
Pask, A, and J.A. Marshall Graves. 1999. "Sex Chromosome and Sex-Determining Genes: Insights from Marsupials and Monotremes." *Cellular and Molecular Life Sciences* 55: 864–75.
Pearson, D, R Shine, and R How. 2002. "Sex-Specific Niche Partitioning and Sexual Size Dimorphism in Australian Pythons (Morelia Spilota Imbricata)." *Biological Journal of the Linnean Society* 77: 113–25.
Pearson, David, Richard Shine, and Andrew Williams. 2002. "Geographic Variation in Sexual Size Dimorphism within a Single Snake Species (Morelia Spilota, Pythonidae)." *Oecologia* 131 (3): 418–26.
Peter Uetz, Jirí Hošek, and Jakob Hallerman. 2012. "The Reptile Database." 2012. http://www.reptile-database.org/.
Peyret, N. 2000. "Prediction of Nucleic Acid Hybridization: Parameters and Algorithms." Department of Chemistry, Detroit: Wayne State University.
Polley, Andreas, Martin W. Ganal, and Elisabeth Seigner. 1997. "Identification of Sex in Hop (Humulus Lupulus) Using Molecular Markers." *Genome* 40 (3): 357–61.
Poulakakis, N, G Goulielmos, A Antoniou, E Zouros, and M Mylonas. 2005. "Isolation and Characterization of Polymorphic Microsatellite Markers in the Wall Lizard Podarcis Erhardii (Squamata: Lacertidae)." *Molecular Ecology* 3 (5): 549–51.
Quinn, Alexander E., Arthur Georges, Stephen D. Sarre, Fiorenzo Guarino, Tariq Ezaz, and Jennifer A. Marshall Graves. 2007. "Temperature Sex Reversal Implies Sex Gene Dosage in a Reptile." *Science* 316: 411.
Radder, Rajkumar S, Alexander E. Quinn, Arthur Georges, Stephen D Sarre, and Richard Shine. 2007. "Genetic Evidence for Co-Occurrence of Chromosomal and Thermal Sex-Determining Systems in a Lizard." *Biology Letters* 4: 176–78.
Raymond, Christopher S., Jae R. Kettlewell, Betsy Hirsch, Vivian J. Bardwell, and David Zarkower. 1999. "Expression of Dmrt1 in the Genital Ridge of Mouse and Chicken Embryos Suggests a Role in Vertebrate Sexual Development." *Developmental Biology* 215: 208–20.
Raymond, Christopher S., Mark W. Murphy, M. Gerard O'Sullivan, Vivian J. Bardwell, and David Zarkower. 2000. "Dmrt1, a Gene Related to Worm and Fly Sexual Regulators, Is Required for Mammalian Testis Differentiation." *Genes and Development* 14: 2587–95.
Raymond, Christopher S., Emily D. Parker, Jae R. Kettlewell, Laura G. Brown, David C. Page, Kamila Kusz, Jadwiga Jaruzelska, et al. 1999. "A Region of Human Chromosome 9p Required for Testis Development Contains Two Genes Related to Known Sexual Regulators." *Human Molecular Genetics* 8: 989–96.
Rens, Willem, Frank Grutzner, Patricia C. M. O'Brien, Helen Fairclough, Jennifer A. M. Graves, and Malcolm A. Ferguson-Smith. 2004. "Resolution and Evolution of the Duck-Billed Platypus Karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 Male Sex Chromosome Constitution." *Proceedings of the National Academy of Sciences* 101: 16257–61.
Rens, Willem, Patricia CM O'Brien, Frank Grützner, Oliver Clarke, Daria Graphodatskaya, Enkhjargal Tsend-Ayush, Vladimir A Trifonov, et al. 2007. "The Multiple Sex Chromosomes of Platypus and Echidna Are Not Completely Identical and Several Share Homology with the Avian Z." *Genome Biology* 8: 8:R243.1-8:R243.21.

Riggio, Gina. 2017. “How to Troubleshoot Your DNA Libraries For 454 Next Generation Sequencing- An Essential Guide.” Science Squared. 2017. https://bitesizebio.com/13578/why-did-my-454-run-fail-a-troubleshooting-guide/.

Rooij, Jasmijn D.E. de, Marry M. van den Heuvel-Eibrink, Wouter J.W. Kollen, Edwin Sonneveld, Gertjan J.L. Kaspers, H. Berna Beverloo, Maarten Fornerod, Rob Pieters, and C. Michel Zwaan. 2016. “Recurrent Translocation t(10;17)(P15;Q21) in Minimally Differentiated Acute Myeloid Leukemia Results in ZMYND11/MBTD1 Fusion.” *Genes, Chromosomes and Cancer* 55 (3): 237–41. https://doi.org/10.1002/gcc.22326.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, D. Muzny, M. Platzer, et al. 2005. “The DNA Sequence of the Human X Chromosome.” *Nature* 434: 325–37.

Rossi, Pellegrino, Susanna Dolci, Cristina Albanesi, Paola Grimaldi, and Raffaele Geremia. 1993. “Direct Evidence That the Mouse Sex-Determining Gene Sry Is Expressed in the Somatic Cells of Male Foetal Gonads and in the Germ Cell Line in the Adult Testis.” *Molecular Reproduction and Developmement* 34: 369–73.

Sakamoto, K., K. Shimomura, Y. Komeda, H. Kamada, and S. Satoh. 1995. “A Male-Associated DNA Sequence in a Dioecious Plant, Cannabis Sativa L.” *Plant & Cell Physiology* 36 (8): 1549–54.

Santa, Lucia, Jr. 1998. “A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbour Thermodynamics.” *Proceedings of the National Academy of Sciences* 95: 1460–65.

Sarre, Stephen D., Tariq Ezaz, and Arthur Georges. 2011. “Transitions between Sex-Determining Systems in Reptiles and Amphibians.” *Annual Review of Genomics and Human Genetics* 12: 391–406.

Schiebel, K, M Winklemann, A Mertz, DC Page, D Weil, C Petit, and GA Rappold. 1997. “Abnormal XY Interchange between a Novel Isolated Protein Kinase Gene, PRKY, and Its Homologue, PRKX, Accounts for One Third of All (Y+)XX Males and (Y-)XY Females.” *Human Molecular Genetics* 6 (11): 1985–89.

Semerikov, V, U Lagercrantz, V Tsarouhas, A Rönnberg-Wästljung, C Alström-Rapaport, and M Lascoux. 2003. “Genetic Mapping of Sex-Linked Markers in Salix Viminalis L.” *Heredity* 91 (August): 293.

Seung-Min, Bae, Lim Whasun, Jeong Wooyoung, Lee Jin-Young, Kim Jinyoung, Bazer Fuller, and Song Gwonhwa. 2013. “Sex-Specific Expression of CTNNB1 in the Gonadal Morphogenesis of the Chicken.” *Reproductive Biology and Endocrinology* 11 (89): 1–10.

Shetty, S., P. Kirby, D. Zarkower, and J.A.M. Graves. 2002. “DMRT1 in a Ratite Bird: Evidence for a Role in Sex Determination and Discovery of a Putative Regulatory Element.” *Cytogenetic and Genome Research* 99: 245–51.

Shetty, Swathi, Darren K. Griffin, and Jennifer A. Marshall Graves. 1999. “Comparative Painting Reveals Strong Chromosome Homology over 80 Million Years of Bird Evolution.” *Chromosome Research* 7: 289–95.

Shine, Richard. 1999. “Why Is Sex Determined by Nest Temperature in Many Reptiles?” *Trends in Ecology and Evolution* 14: 186–89.

Shoemaker, Christina M., and David Crews. 2009. “Analyzing the Coordinated Gene Network Underlying Temperature-Dependent Sex Determination in Reptiles.” *Seminars in Cell & Development Biology* 20: 293–303.

Silber, Sherman J. 2011. “Human Male Infertility, the Y Chromosome, and Dinosaur Extinction.” *Middle East Fertility Society* 16: 114–20.

Sinclair, Andrew H., Philippe Berta, Mark S. Palmer, J. Ross Hawkins, Beatrice L. Griffiths, Matthijs J. Smith, Jamie W. Foster, Anna-Maria Frischauf, Robin Lovell-Badge, and Peter N. Goodfellow. 1990. “A Gene from the Human Sex-Determining Region

Encodes a Protein with Homology to a Conserved DNA-Binding Motif." *Nature* 346: 240–44.

Sinclair, Andrew H., Jamie W. Foster, James A. Spencer, David C. Page, Mark Palmer, Peter N. Goodfellow, and Jennifer A. Marshall Graves. 1988. "Sequences Homologous to ZFY, a Candidate Human Sex-Determining Gene, Are Autosomal in Marsupials." *Nature* 336: 780–83.

Singh, L., and S.P. Ray-Chaudhuri. 1975. "Localization of C-Band in the W Sex Chromosome of the Common Indian Krait Bungarus Caeruleus Schneider." *Nucleus* 18: 163–66.

Sites, JW Jr, JW Bickham, and MW Haiduk. 1979. "Derived X Chromosome in the Turtle Genus Staurotypus." *Science* 206: 1410–12.

Skaletsky, Helen, Tomoko Kuroda-Kawaguchi, Patrick J. Minx, Holland S. Cordum, LaDeana Hillier, Laura G. Brown, Sjoerd Repping, et al. 2003. "The Male-Specific Region of the Human Y Chromosome Is a Mosaic of Discrete Sequence Classes." *Nature* 423: 825–37.

Smit, AFA, R Hubley, and P Green. 2013. *RepeatMasker Open-4.0*. http://www.repeatmasker.org.

Smith, C.A., K.N. Roeszler, Q.J. Hudson, and A.H. Sinclair. 2007. "Avian Sex Determination: What, When and Where?" *Cytogenetic and Genome Research* 117: 165–73.

Smith, Craig A., Melissa Katz, and Andrew H. Sinclair. 2003. "DMRT1 Is Upregulated in the Gonads during Female-to-Male Sex Reversal in ZW Chicken Embryos." *Biology of Reproduction* 68: 560–70.

Smith, Craig A., Peter J. McClive, Patrick S. Western, Kirsty J. Reed, and Andrew H. Sinclair. 1999. "Conservation of a Sex-Determining Gene." *Nature* 402: 601–2.

Son, Dong Ju, Yu Yeon Jung, Young Sik Seo, Heonyong Park, Dong Hun Lee, Sanghyeon Kim, Yoon-Seok Roh, Sang Bae Han, Do Young Yoon, and Jin Tae Hong. 2017. "Interleukin-32alpha Inhibits Endothelial Inflammation, Vascular Smooth Muscle Cell Activation, and Atherosclerosis by Upregulating Timp3 and Reck through Suppressing MicroRNA-205 Biogenesis." *Theranostics* 7 (8): 2186–2203. https://doi.org/10.7150/thno.18407.

Soullier, Stephan, Catherine Hanni, Francois Catzeflis, Philippe Berta, and Vincent Laudet. 1998. "Male Sex Determination in the Spiny Rat Tokudaia Osimensis (Rodentia: Muridae) Is Not Sry Dependent." *Mammalian Genome* 9: 590–92.

St. John, Judy, and Tom Quinn. 2008. *Recent CRI Non-LTR Retrotransposon Activity in Coscoroba Reveals an Insertion Site Preference*. Vol. 9. https://doi.org/10.1186/1471-2164-9-567.

Sutou, S., Y. Mitsui, and K. Tsuchiya. 2001. "Sex Determination without the Y Chromosome in Two Japanese Rodents Tokudaia Osimensis Osimensis and Tokudaia Osimensis Spp." *Mammalian Genome* 12: 17–21.

Sven Wuertz, Sandrine Gaillard, Federica Barbisan, Serge Carle, Leonardo Congiu, Anna Forlani, Josiane Aubert, et al. 2006. "Extensive Screening of Sturgeon Genomes by Random Screening Techniques Revealed No Sex-Specific Markers." *Aquaculture* 258 (1–4): 685–88.

Takagi, N, M Itoh, and M Sasaki. 1972. "Chromosome Studies in Four Species of Ratitae (Aves)." *Chromosoma* 36: 281–91.

Takagi, N., and M. Sasaki. 1974. "A Phylogenetic Study of Bird Karyotypes." *Chromosoma* 46: 91–120.

Taylor, Renea A., Michael Fraser, Julie Livingstone, Shadrielle Melijah G. Espiritu, Heather Thorne, Vincent Huang, Winnie Lo, et al. 2017. "Germline BRCA2 Mutations Drive Prostate Cancers with Distinct Evolutionary Trajectories." *Nature Communications* 8 (January): 13671.

Tsuruoka, M, and I Karube. 2003. “Rapid Hybridization at High Salt Concentration and Detection of Bacterial DNA Using Fluorescence Polarization.” *Combinatorial Chemistry & High Throughput Screening* 6 (3): 225–34.

Urasaki, N., M. Tokumoto, K. Tarora, Y. Ban, T. Kayano, H. Tanaka, H. Oku, I. Chinen, and R. Terauchi. 2002. “A Male and Hermaphrodite Specific RAPD Marker for Papaya (Carica Papaya L.).” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 104 (2–3): 281–85. https://doi.org/10.1007/s001220100693.

Valenzuela, Nicole, and Valentine A. Lance. 2004. *Temperature-Dependent Sex Determination in Vertebrates*. Smithsonian Books.

Vallender, Eric J., and Bruce T. Lahn. 2006. “Multiple Independent Origins of Sex Chromosomes in Amniotes.” *Proceedings of the National Academy of Sciences* 103: 18031–32.

Veyrunes, Frédéric, Paul D. Waters, Pat Miethke, Willem Rens, Daniel McMillan, Amber E. Alsop, Frank Grützner, et al. 2008. “Bird-like Sex Chromosomes of Platypus Imply Recent Origin of Mammal Sex Chromosomes.” *Genome Research* 18: 965–73.

Vicoso, Beatriz, J. J. Emerson, Yulia Zektser, Shivani Mahajan, and Doris Bachtrog. 2013. “Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation.” *PLOS Biology* 11 (8): e1001643. https://doi.org/10.1371/journal.pbio.1001643.

Vidal, Valerie P, Marie-Christine Chaboissier, Dirk G. de Rooij, and Andreas Schedl. 2001. “Sox9 Induces Testis Development in XX Transgenic Mice.” *Nature Genetics* 28: 216–17.

Voineagu, Irina, Vidhya Narayanan, Kirill S. Lobachev, and Sergei M. Mirkin. 2008. “Replication Stalling at Unstable Inverted Repeats: Interplay between DNA Hairpins and Fork Stabilizing Proteins.” *Proceedings of the National Academy of Sciences* 105 (29): 9936–41.

Vonk, Freek J., Nicholas R. Casewell, Christiaan V. Henkel, Alysha M. Heimberg, Hans J. Jansen, Ryan J. R. McCleary, Harald M. E. Kerkkamp, et al. 2013. “The King Cobra Genome Reveals Dynamic Gene Evolution and Adaptation in the Snake Venom System.” *Proceedings of the National Academy of Sciences* 110 (51): 20651–56. https://doi.org/10.1073/pnas.1314702110.

Wallis, M. C., P. D. Waters, M. L. Delbridge, P. J. Kirby, A. J. Pask, F. Grutzner, W. Rens, M. A. Ferguson-Smith, and J. A. M. Graves. 2007. “Sex Determination in Platypus and Echidna: Autosomal Location of SOX3 Confirms the Absence of SRY from Monotremes.” *Chromosome Research* 15: 949–59.

Wallis, M. C., P. D. Waters, and J. A. M. Graves. 2008. “Sex Determination in Mammals - Before and after the Evolution of SRY.” *Cellular and Molecular Life Sciences* 65: 3182–95.

Warren, Wesley C., LaDeana W. Hillier, Jennifer A. Marshall Graves, Ewan Birney, Chris P. Ponting, Frank Grutzner, Katherine Belov, et al. 2008. “Genome Analysis of the Platypus Reveals Unique Signatures of Evolution.” *Nature* 453: 175–83.

Waters, Paul D., Margaret L. Delbridge, Janine E. Deakin, Nisrine El-Mogharbel, Patrick J. Kirby, Denise R. Carvalho-Silva, and Jennifer A. Marshall Graves. 2005. “Autosomal Location of Genes from the Conserved Mammalian X in the Platypus (Ornithorhynchus Anatinus): Implications for Mammalian Sex Chromosome Evolution.” *Chromosome Research* 13: 401–10.

Watson, Jaclyn M., James A. Spencer, Arthir D. Riggs, and Jennifer A. Marshall Graves. 1990. “The X Chromosome of Monotremes Shares a Highly Conserved Region with the Eutherian and Marsupial X Chromosomes despite the Absence of X Chromosome Inactivation.” *Proceedings of the National Academy of Sciences* 87: 7125–29.

Welsh, John, and Michael McClelland. 1990. “Fingerprinting Genomes Using PCR with Arbitrary Primers.” *Nucleic Acids Research* 18: 7213–18.

Welshons, W. J., and Liane Brauch Russell. 1959. “The Y-Chromosome as the Bearer of Male Determining Factors in the Mouse.” *Proceedings of the National Academy of Sciences* 45: 560–66.

Williams, John G.K., Anne R. Kubelik, Kenneth J. Livak, J.Antoni Rafalski, and Scott V. Tingey. 1990. “DNA Polymorphisms Amplified by Arbitrary Primers Are Useful as Genetic Markers.” *Nucleic Acids Research* 18: 6531–35.

Xiong, Y, and TH Eickbush. 1990. “Origin and Evolution of Retroelements Based upon Their Reverse Transcriptase Sequences.” *European Molecular Biology Organization* 9 (10): 3353–62.

Yin, Wei, Zong-ji Wang, Qi-ye Li, Jin-ming Lian, Yang Zhou, Bing-zheng Lu, Li-jun Jin, et al. 2016. “Evolutionary Trajectories of Snake Genes and Genomes Revealed by Comparative Analyses of Five-Pacer Viper.” *Nature Communications* 7 (13107).

Yuan, Chang. 2002. “Representational Difference Analysis.” *Current Protocols in Molecular Biology*, no. 25B.7.1: 1–12.

Zuker, M. 2003. “Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction.” *Nucleic Acid Research* 31 (13): 3406–3015.

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

Appendix A: Reagent Recipes

SM buffer

5.8 g of NaCl

2.0 g of $MgSO_4.7H_2O$

50 ml of 1M Tris.HCl pH 7.5

5 ml of 2% gelatine,

Made to a final volume of 1 L with hpH_2O and autoclaved.

LB broth

10 g NaCl

10 g tryptone

5 g yeast extract

Made to a final volume of 1 L with hpH_2O and a final pH of 7.0 and autoclaved. The broth is stored in its sealed autoclaved contained at 4° for up to 4 weeks.

LBA

10 g NaCl

10 g tryptone

5 g yeast extract

20 g agar

Made to a final volume of 1 L with hpH_2O and a final pH of 7.0 and autoclaved at 121°C for 30 minutes.

LBA plates containing Amp

100 ml of LBA

1 ml of 100 mg/ml ampicillin

Allowed the LBA to cool to approximately 50°C and add the ampicillin solution. Mix the LBA gently, limiting the production of bubbles. Pour 20 ml of LBA onto five 90 mm diameter petri dishes and allow the agar plates to set. The plates are dried, upside down, for 30 minutes at 37°C before being stored in cling-film at 4°C. LBA plates are stored no longer than 2 weeks.

LBA plates containing Amp, X-Gal and IPTG

100 ml of LBA

1 ml of 100 mg/ml ampicillin

200 µl of 100 mg/ml X-Gal solution in DMF

1 ml of 100 mM IPTG solution in hpH_2O

Allowed the LBA to cool to approximately 50°C and add the ampicillin, X-Gal and IPTG solution. Mix the LBA gently, limiting the production of bubbles. Pour 20 ml of LBA onto five 90 mm diameter petri dishes and allow the agar plates to set. The plates are dried, upside down, for 30 minutes at 37°C before being stored in cling-film at 4°C. LBA plates are stored no longer than 2 weeks.

TE buffer

1.21 g Tris-HCl

0.37 g $Na2EDTA.2H_2O$

Made to a final volume of 1 L with hpH_2O and a final pH of 8.0 and autoclaved at 121°C for 30 minutes.

50× TAE buffer

242 g Tris

57.1 mL glacial acetic acid

37.2 g $Na2EDTA.2H_2O$

Made to a final volume of 1 L with hpH_2O.

1× TAE buffer

Mix 20 ml of 50× TAE buffer with 980 ml of hpH_2O to a final volume of 1 L.

Cell lysis buffer

1.21 g Tris-HCl

3.72 g $Na2EDTA.2H_2O$

20 g SDS

Made to a final volume of 1 L with hpH_2O and a final pH of 8.0 and autoclaved at 121°C for 30 minutes.

SOC medium

5 g yeast extract

20 g tryptone

0.58 g NaCl

0.18 g KCl

0.95 g $MgCl_2$

1.20 g $MgSO_4$

3.60 g Glucose

Made to a final volume of 1 L with hpH_2O and a final pH of 7.0 and autoclaved at 121°C for 30 minutes. The medium is stored in its sealed autoclaved contained at 4° for up to 2 weeks.

Lysis buffer

0.58 g NaCl

1.21 g Tris-HCl

0.37 g Na2EDTA.2H_2O

5 g SDS

Made to a final volume of 1 L with hpH_2O and a final pH of 8.0 and autoclaved.

6× Agarose Electrophoresis Gel Loading Buffer

0.3 ml glycerol

25 mg bromophenol blue

10 mg SDS

Made to a final volume of 10 ml with hpH_2O.

Appendix B: DNA Sequences Data

See Zip file; Appendix B - Sequence Data