

Medical University of South Carolina

MEDICA

MUSC Theses and Dissertations

1-1-2017

A Comprehensive Molecular Phylogeny of Holocephalans and Identification of Sex-Specific Genetic Markers in Chondrichthyan Fishes

Jenny Marie Kemper
Medical University of South Carolina

Follow this and additional works at: <https://medica-musc.researchcommons.org/theses>

Recommended Citation

Kemper, Jenny Marie, "A Comprehensive Molecular Phylogeny of Holocephalans and Identification of Sex-Specific Genetic Markers in Chondrichthyan Fishes" (2017). *MUSC Theses and Dissertations*. 937.
<https://medica-musc.researchcommons.org/theses/937>

This Dissertation is brought to you for free and open access by MEDICA. It has been accepted for inclusion in MUSC Theses and Dissertations by an authorized administrator of MEDICA. For more information, please contact medica@musc.edu.

A Comprehensive Molecular Phylogeny of Holocephalans and Identification of Sex-Specific Genetic Markers in Chondrichthyan Fishes

by

Jenny Marie Kemper

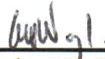
A dissertation submitted to the faculty of the Medical University of South Carolina in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Graduate Studies.

Department of Molecular and Cellular Biology and Pathobiology
Marine Biomedicine and Environmental Sciences

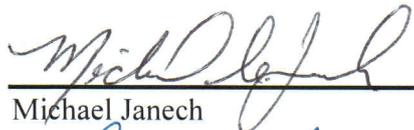
2017

Approved by:

Chairman, Advisory Committee



Gavin J.P. Naylor



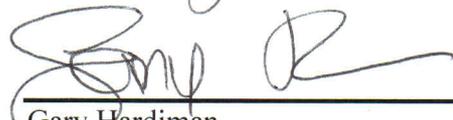
Michael Janech



Paula Traktman



Dayna Wolff



Gary Hardiman

ACKNOWLEDGEMENTS

This dissertation was supported in part by the National Science Foundation grant DEB-1132229 to Gavin Naylor, funding support from the Medical University of South Carolina College of Graduate Studies, and the Marine Biomedicine and Environmental Sciences program. Additional support came from the Southern Association of Marine Laboratories (SAML) student support program and the Henry and Anne Mollet Elasmobranch Research Award through the American Elasmobranch Society.

I would like to gratefully acknowledge my committee members, Gavin Naylor, Michael Janech, Daynna Wolff, Paula Traktman, and Gary Hardiman, for their support, guidance, encouragement, and expertise throughout my dissertation. I would also like to thank Satomi Kohno for his help with several aspects of this research.

I am grateful to my many past and present lab mates. In particular, I would like to thank Shannon Corrigan for her invaluable help and advice throughout this entire project, as well as thoughtful review of chapters and manuscripts. I'd additionally like to acknowledge Lei Yang who's work on mitochondrial genome capture was critical to this work, as well as always lending support and good discussion. Last, I would like to thank Elisabeth Rochel for all of her assistance.

Many people and institutions generously provided tissue and/or DNA samples that were vital to this research. These include: A. Graham, A. Bentley, A. Stewart, C. Cotton, D. Ebert, D. Didier, H. Calumpong, J. Caira, K. Jensen, J. Moore, J. Diaz de Astarloa, K. Sato, K. Graham, K. Bineesh, K. Walovich, L. Singh, M. Francis, P. Clerkin, P. Psomadakis, P. Smith, R. Hanel, R. Robertson, S. Iglesias, T. Pietsch, D. Roje, W. White. I am truly grateful to Dr. Edgardo Ernesto di Giacomo who collected and sent *Callorhynchus callorynchus* tissue samples from Argentina, that were critical for the sex determination research.

Many others have generously provided assistance for this project. I would like to thank Shane Boylan, Bryan Frazier, and Ashely Shaw for help in collection of shark tissue for cell culture and chromosome analysis. Cathy Walsh and Carl Luer provided advice and expertise on cell culture techniques in elasmobranchs. I greatly appreciate the help and time spent by the staff of MUSC cytogenetics department who generously allowed me to learn their techniques, and helped me design, carryout and troubleshoot an experiment in culturing shark tissue and preparing chromosome spreads. I'd like to acknowledge Travis Glenn and Troy Kieran at the University of Georgia, for their help and guidance in preparing the 3RAD libraries and subsequent sequencing. Additionally, I am grateful to Tony Gamble whose knowledge of RAD-seq and sex determination was vital to this project. He also provided much assistance throughout the library preparation process, data analysis, and validation assays. C. Underwood provided fossil calibration information for divergence time estimation. I would also like to recognize the many researchers and staff at the Hollings Marine Lab. In particular, Peter Moeller, Lou Burnett, and Peter Lee who were very helpful when setting up cell culture and incubation at Hollings Marine Lab. Also, I am very grateful to Lisa May and Cheryl Woodley who took me into their lab and generously providing me with supplies and equipment to carry out PCR experiments.

I would like to acknowledge the students, faculty, and staff of the Marine Biomedicine and Environmental Sciences program for all their help and support during my dissertation, especially, Lou Guillette, John Baatz, and Taneisha Simpson. Last, I am grateful to my family and friends for their enduring love, support, and encouragement not only throughout my dissertation, but in everything that I do.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
LIST OF ABBREVIATIONS.....	xi
ABSTRACT.....	xv
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: A comprehensive phylogeny of holocephalans (Chondrichthyes: Chimaeriformes) using multi-locus nuclear data	
INTRODUCTION.....	16
MATERIALS AND METHODS.....	25
RESULTS AND DISCUSSION.....	41
CONCLUSIONS.....	87
CHAPTER 3: Molecular phylogeny of holocephalans inferred from mitochondrial genomes	
INTRODUCTION.....	97
MATERIALS AND METHODS.....	98
RESULTS AND DISCUSSION.....	106
CONCLUSIONS.....	147
CHAPTER 4: Identification of sex-specific genomic markers in chondrichthyan fishes	
INTRODUCTION.....	156
MATERIALS AND METHODS.....	179
RESULTS AND DISCUSSION.....	196
CONCLUSIONS.....	244
CHAPTER 5: DISCUSSION.....	251
GLOSSARY.....	263

APPENDIX.....	265
REFERENCES	286

LIST OF TABLES

CHAPTER 2: Phylogeny of holocephalans using nuclear gene capture data

Table 2.1	Holocephalan samples for gene capture	29
Table 2.2	Elasmobranch species for gene capture	31
Table 2.3	Maximum likelihood analyses	36
Table 2.4	Bayesian inference analyses	36
Table 2.5	Fossil calibrations in divergence time estimation.....	40
Table 2.6	Divergence node ages	84

CHAPTER 3: Phylogeny of holocephalans using mitochondrial genome capture data

Table 3.1	Maximum likelihood analyses	101
Table 3.2	Bayesian inference analyses	101
Table 3.3	Fossil calibrations in divergence time estimation.....	104
Table 3.4	Mitochondrial genome capture statistics	107
Table 3.5	Characteristics of mitochondrial protein-coding genes	111
Table 3.6	Divergence node ages	132

CHAPTER 4: Identification of sex-specific genomic markers in chondrichthyans

Table 4.1	Known vertebrate sex-determining genes.....	158
Table 4.2	<i>Carcharhinus plumbeus</i> samples for 3RAD sequencing.....	185
Table 4.3	Chondrichthyan samples used for single digest RAD-seq.....	189
Table 4.4	Primers for candidate sex-specific RAD contigs in <i>Callorhynchus callorynchus</i>	194
Table 4.5	Shark tissue samples for chromosome analysis.....	197
Table 4.6	Shark blood samples for cell culture and chromosome analysis	198
Table 4.7	Shark blood samples with concanavalin A exposure experiment.....	200
Table 4.8	<i>Carcharhinus isodon</i> chromosome analysis experiment at MUSC.....	200

Table 4.9	<i>Carcharhinus plumbeus</i> chromosome analysis experiment at MSUC	201
Table 4.10	BrdU cell proliferation assay for <i>Carcharhinus isodon</i>	204
Table 4.11	BrdU cell proliferation assay for <i>Rhizoprionodon terraenovae</i>	206
Table 4.12	Summary of 3RAD data analysis using STACKS.....	213
Table 4.13	Summary of 3RAD putative sex-specific loci	214
Table 4.14	Primers for sex-specific RAD loci under 3RAD	215
Table 4.15	Summary of sequenced individual RAD libraries	228
Table 4.16	Summary of RAD-seq library analyses	229
Table 4.17	Putative sex-specific RAD markers in <i>Callorhynchus callorynchus</i>	230
Table 4.18	Putative sex-specific RAD markers in <i>Carcharodon carcharias</i>	231
Table 4.19	Summary of <i>Callorhynchus callorynchus</i> candidate male-specific contigs.....	235
Table 4.20	RAD marker 1889 contig 1 sequence	235
Table 4.21	RAD marker 1895 contig 1 sequence	235

LIST OF FIGURES

CHAPTER 1: Introduction

Figure 1.1	Phylogenetic relationships of major groups of vertebrates	3
Figure 1.2	Schematic of phylogenetic tree for character state reconstruction	8

CHAPTER 2: Phylogeny of holocephalans using nuclear gene capture data

Figure 2.1	Drawing of <i>Callorhinchus</i> sp.	18
Figure 2.2	Photograph of <i>Rhinochimaera atlantica</i>	19
Figure 2.3	Photograph of <i>Neoharriotta pinnata</i>	19
Figure 2.4	Photograph of <i>Harriotta raleighana</i>	19
Figure 2.5	Photograph of <i>Chimaera notafricana</i>	20
Figure 2.6	Photograph of <i>Hydrolagus africanus</i>	20
Figure 2.7	Schematic of taxon identification and nuclear exon data sets workflow	26
Figure 2.8	Length distribution of 1,264 captured nuclear exons	42
Figure 2.9	Maximum likelihood tree of full nucleotide data	44
Figure 2.10	Maximum likelihood tree of clock-filtered nucleotide data	46
Figure 2.11	Maximum likelihood tree of full amino acid data	52
Figure 2.12	Maximum likelihood and consensus of clock-filtered amino acid data	56
Figure 2.13	Maximum likelihood and consensus of 1077 exon nucleotide data.....	58
Figure 2.14	Maximum likelihood and consensus of 50 exon nucleotide data, no partition	60
Figure 2.15	Maximum likelihood and consensus of 50 exon nucleotide data, codon partition.....	62
Figure 2.16	Maximum likelihood and consensus of 50 exon nucleotide data, gene partition.	64
Figure 2.17	Maximum likelihood and consensus of 50 exon nucleotide data, gene+codon partition.....	66
Figure 2.18	Maximum likelihood and consensus of 50 exon amino acid data.....	68

Figure 2.19	Bayesian inference tree of clock-like filtered nucleotide data, codon partition ...	71
Figure 2.20	Bayesian inference tree of 50 exon nucleotide data, no partition.....	72
Figure 2.21	Bayesian inference tree of 50 exon nucleotide data, codon partitioning.....	73
Figure 2.22	Bayesian inference tree of 50 exon amino acid data	74
Figure 2.23	Divergence time tree of 50 exon nucleotide data	79
Figure 2.24	Strict consensus tree	89

CHAPTER 3: Phylogeny of holocephalans using mitochondrial genome capture data

Figure 3.1	Histogram of percent of mitochondrial capture reads on target	108
Figure 3.2	Histogram of average coverage of mitochondrial genome capture.....	109
Figure 3.3	Maximum likelihood tree of nucleotide data, codon partition	113
Figure 3.4	Maximum likelihood tree of nucleotide data, gene+codon partition	114
Figure 3.5	Maximum likelihood tree of amino acid data.....	115
Figure 3.6	Strict consensus tree of nucleotide and amino acid data	116
Figure 3.7	Bayesian inference tree of nucleotide data, codon partition.....	124
Figure 3.8	Bayesian inference tree of nucleotide data, gene+codon partition.....	125
Figure 3.9	Bayesian inference tree of amino acid data	126
Figure 3.10	Divergence time tree of nucleotide data.....	128
Figure 3.11	Strict consensus of nucleotide mitochondrial and nuclear data.....	135

CHAPTER 4: Identification of sex-specific molecular markers in chondrichthyans

Figure 4.1	Male RAD marker 48396 PCR in <i>Carcharhinus plumbeus</i>	216
Figure 4.2	Male RAD marker 17488 PCR in <i>Carcharhinus plumbeus</i>	216
Figure 4.3	Male RAD marker 9873 PCR in <i>Carcharhinus plumbeus</i>	217
Figure 4.4	Male RAD marker 12965 PCR in <i>Carcharhinus plumbeus</i>	217
Figure 4.5	Male RAD marker 515 PCR in <i>Carcharhinus plumbeus</i>	218
Figure 4.6	Female RAD marker 67066 PCR in <i>Carcharhinus plumbeus</i>	218

Figure 4.7	Female RAD marker 62654 PCR in <i>Carcharhinus plumbeus</i>	219
Figure 4.8	Female RAD marker 63104 PCR in <i>Carcharhinus plumbeus</i>	219
Figure 4.9	Female RAD marker 56104 PCR in <i>Carcharhinus plumbeus</i>	220
Figure 4.10	Female RAD marker 54677 PCR in <i>Carcharhinus plumbeus</i>	222
Figure 4.11	Female RAD marker 9998 PCR in <i>Carcharhinus plumbeus</i>	222
Figure 4.12	Female RAD marker 66178 PCR in <i>Carcharhinus plumbeus</i>	223
Figure 4.13	Female RAD marker 58030 PCR in <i>Carcharhinus plumbeus</i>	223
Figure 4.14	Female RAD marker 12075 PCR in <i>Carcharhinus plumbeus</i>	224
Figure 4.15	Positive control PCR amplification.....	236
Figure 4.16	Male-specific RAD marker 1889 primer validation PCR in <i>Callorhynchus callorynchus</i>	236
Figure 4.17	Male-specific RAD marker 1895 primer validation PCR in <i>Callorhynchus callorynchus</i>	237
Figure 4.18	Male-specific RAD marker primer and internal control PCR validation.....	238
Figure 4.19	RAD marker 1889 PCR validation in male and female <i>Callorhynchus callorynchus</i>	239
Figure 4.20	RAD marker 1895 PCR validation in male and female <i>Callorhynchus callorynchus</i>	240

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
amh	Anti-Müllerian hormone
amhr2	amh type II receptor
BAC	Bacterial artificial chromosome
Ba(OH) ₂	Barium hydroxide
BI	Bayesian inference
BIC	Bayesian information criterion
BLAST	Basic local alignment search tool
bp	Base pair
BrdU	Bromodeoxyuridine
cf	Confer
CIPRES	Cyberinfrastructure for Phylogenetic Research
CNE	Conserved non-coding element
cyp19a	Aromatase
cyp19a1a	Cytochrome P450/Aromatase
DMRT1	DM-related transcription factor 1
DNA	Deoxyribonucleic acid
ESD	Environmental sex determination
ESS	Effective sample size
F	Empirical base frequencies
FISH	Fluorescent in situ hybridization
foxl2	Forkhead box L2
Gb	Giga base pairs

GSD	Genetic sex determination
gsdf	Gonadal soma derived growth factor
GTR	General time-reversible
HCl	Hydrochloric acid
HF	High fidelity
H-strand	Heavy strand
HPD	Highest posterior density
I	Invariable sites
IL-2	Intereukin 2
ILS	Incomplete lineage sorting
Irf9	Interferon regulatory factor 9
kb	Kilo base pairs
KCl	Potassium chloride
LRT	Likelihood ratio test
LSR	Lazy sub-tree rearrangement
L-strand	Light strand
Ma	Million years ago
Mb	Mega base pairs
(MC) ³	Metropolis-coupled Markkov chain Monte Carlo
MCMC	Markov Chain Monte Carlo
miRNA	microRNA
ML	Maximum likelihood
MRCA	Most recent common ancestor
MUSC	Medical University of South Carolina
NaCl	Sodium chloride

NADH2	NADH dehydrogenase subunit 2
NCBI	National center for biotechnology information
ORF	Open reading frame
PBL	Peripheral blood leukocytes
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PHA	Phytohaemagglutinin
pHMM	profile hidden Markov model
PSRF	Potential Scale Reduction Factor
RAD	Restriction site-associated DNA
RAD-seq	Restriction site-associated DNA sequencing
RAG-1	Recombination-activating gene 1
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
secs	Seconds
sf-1	Steroidogenic factor 1
SNP	Single nucleotide polymorphism
sp.	Species
spp.	Species
SSC	Sodium saline citrate
TGF- β	Transforming growth factor beta
tmrca	Time to most recent common ancestor
tRNA	Transfer ribonucleic acid
TSD	Temperature-dependent sex determination

Γ

Gamma parameter

JENNY MARIE KEMPER. A Comprehensive Molecular Phylogeny of Holocephalans and Identification of Sex-Specific Genetic Markers in Chondrichthyan Fishes. (Under the direction of GAVIN JP NAYLOR).

ABSTRACT

Holocephalan fishes occupy an important phylogenetic position as members of the basal clade of jawed vertebrates, as well as the sister group to bony vertebrates. Thus, they can be used to provide an important reference for comparative studies aimed at understanding early gnathostome genome evolution. However, despite this importance, holocephalans have been understudied, and no comprehensive phylogeny for their inter-relationships has been estimated. A phylogeny will provide information about ancestor-descendant relationships, forming a framework that can be used to test future hypotheses by mapping characters, identifying patterns, and estimating ancestral states. A cross-species hybridization capture method was used to sequence over 1,000 nuclear exons and whole mitochondrial genomes from 55 chimaeroid lineages. Phylogenetic analysis used different methods, partitioning schemes, and character data to estimate and compare trees. Nuclear and mitochondrial data produced mainly congruent results across different data sets and analyses. All trees recovered the same major clades and species within. Major differences among topologies were found in shallower nodes, species-level relationships, and placement of a few clades, and were less resolved. Overall, a robust phylogenetic framework representing the majority of chimaeroid lineages and a consensus tree has been estimated. Divergence times of lineages were estimated, largely congruent between data sets. These results have significant taxonomic implications for the group. Sex determination is an important biological mechanism among organisms. Yet, within chondrichthyans, their sex-determining mechanism is not yet identified nor confirmed. Therefore, restriction site-associated

DNA sequencing was used to interrogate the genome of three species, a chimaera, shark, and ray, for sex-linked markers. Candidate sex-specific markers were identified in all three species, consistent with a genetic mechanism of sex determination. Two male-specific markers were identified and validated in *Callorhynchus callorhynchus*, a pattern that is consistent with a male heterogametic, XX/XY sex chromosome system. The shark and ray species were not validated, but their candidate sex-linked marker results also indicate male heterogamety. Estimated evolutionary relationships provides a framework for future research aimed at better understanding this group, and for comparative studies in vertebrate evolution. Specifically, future work can further explore the sex-determining mechanisms and genes among chimaeras and chondrichthyans.

CHAPTER 1

INTRODUCTION

Evolutionary Importance of Chondrichthyan Fishes

Extant gnathostomes (jawed vertebrates) are represented by two primary divisions, chondrichthyan fishes (sharks, rays, skates, and chimaeras) that are characterized by a mostly cartilaginous endoskeleton, and bony vertebrates (bony fishes and tetrapods), characterized by an ossified endoskeleton (Figure 1.1). Paleontological, morphological, and molecular studies have shown that chondrichthyans and bony vertebrates are sister groups, with chondrichthyans as the basal clade (Janvier, 1996; Cole & Currie, 2007; Venkatesh et al., 2014). Chondrichthyan fishes are phylogenetically considered to be the most basal living group of jawed vertebrates, as their lineage diverged early on in gnathostome history (Janvier, 1996; Kikugawa et al., 2004; Cole & Currie, 2007). Chondrichthyans shared a common ancestor with bony vertebrates approximately 462 to 421 million years ago (Ma) (Benton et al., 2009), and they first appeared in the fossil record in the Paleozoic era (Sansom et al., 1996; Coates & Sequeira, 2001). Chondrichthyans have widely been recognized as an important outgroup for studies on bony vertebrates based on their basal phylogenetic position to other jawed vertebrates, as they are assumed to be the extant group that approximates the ancestral gnathostome condition (Janvier, 1996; Neyt et al., 2000; Venkatesh et al., 2001; Chiu et al., 2002, 2004; Tanaka et al., 2002; Amores et al., 2004; Kikugawa et al., 2004; Robinson-Rechavi et al., 2004; Mulley et al., 2009). It should be noted that while chondrichthyans represent a very old lineage, it does not mean that they are more primitive than other vertebrates, and in fact, a recent study hypothesized that modern sharks may be more derived (Pradel et al., 2014). Thus, the term ‘basal’ does not imply that an extant clade is more or less ancestral than any other clade. It simply states that the clade is a descendant of a lineage that diverged early in history, as seen by divergence near the root of a phylogenetic tree. Nevertheless, their key phylogenetic position as basal jawed vertebrates and their sister group

relationship to bony vertebrates, makes them an important reference for understanding gnathostome genome evolution, identifying origins of complex organismal systems, genes, and their regulation, and detecting genomic, physiological and developmental transitions over the course of evolution. Additionally, they have the potential to provide insight into the ancestral gnathostome condition, and represent an important outgroup for comparative studies across bony vertebrate lineages.

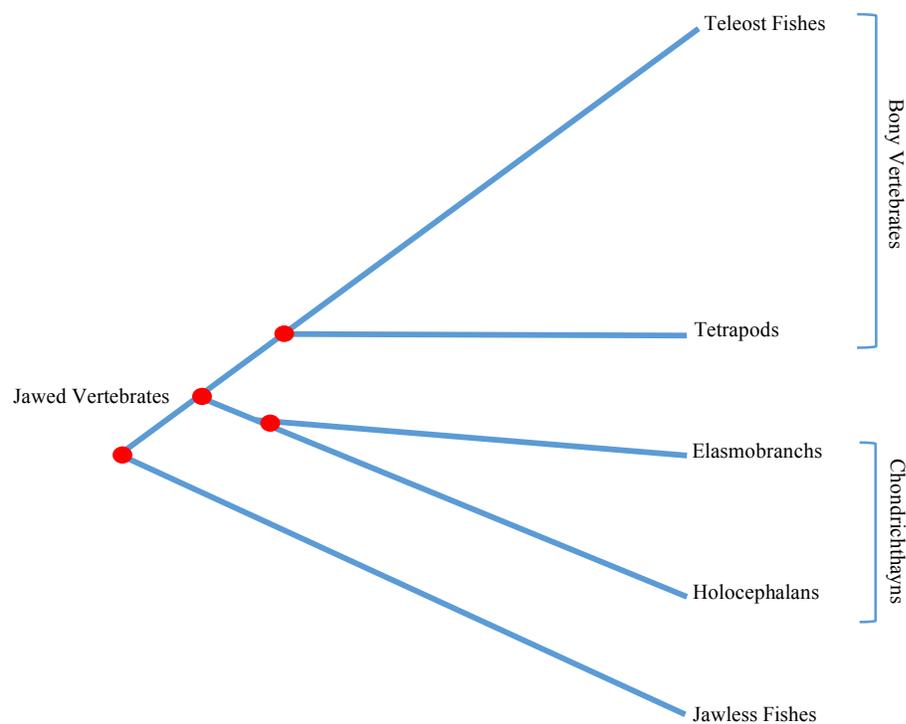


Figure 1.1 Schematic of phylogenetic relationships among the major divisions of vertebrates.

Within chondrichthyan fishes there are two monophyletic sister groups, the holocephalans (chimaeras) and the elasmobranchs (sharks, rays, skates). Collectively, they are appreciated as the basal sister group to bony vertebrates (Kikugawa et al., 2004; Cole & Currie, 2007; Venkatesh et al., 2014), and occupy a phylogenetic position as the most distantly related

extant group of jawed vertebrates to bony vertebrates (Figure 1.1). In order to expand on and incorporate a more distantly related genome into comparative studies on human vertebrate genome evolution, a model chondrichthyan species was identified. The elephantfish, *Callorhinchus milii*, was chosen as a good candidate for whole genome sequencing due to its small genome size relative to other chondrichthyans (Venkatesh et al., 2005).

The whole genome of a single male *C. milii* was sequenced largely from fosmid clones (35-40 kb inserts), resulting in 330,000 contigs and 240,000 singletons, a resulting genome size of approximately 0.91 Gb, with 1.4X coverage, representing about 75% of the genome (Venkatesh et al., 2007). Additionally, Venkatesh et al. (2014) used a shotgun sequencing approach and bacterial artificial chromosome (BAC) libraries to assemble and annotate the *C. milii* genome further. Their assembly resulted in a genome size of 0.937 Gb, with an average depth of 19.25X. They also used RNA sequencing data to supplement genome data. They retrieved a total of 21,208 scaffolds, with a N50 of 46.6 kb for contigs and 4.6 Mb for scaffolds (Venkatesh et al., 2014). The average GC content was 42.3%, and showed heterogeneity across scaffolds and at genome level, with ~ 46% of genome organized into 246 isochores across three families (Venkatesh et al., 2014). A total of 18,872 protein-coding genes were predicted based on genome and RNA sequence data (Venkatesh et al., 2014). They also characterized microRNAs (miRNAs), conserved non-coding elements (CNEs), rate of molecular evolution, intron evolution, synteny, and evolution of some protein domains and gene families.

Overall, the whole genome sequence of *C. milii* revealed that the genome organization of mammals is more similar to holocephalans than it is to teleost fishes, even though tetrapods and teleosts are more closely related (Venkatesh et al., 2007). The *C. milii* genome was found to share a higher level of conserved synteny with the human genome, relative to the human-teleost genome comparisons (Venkatesh et al., 2007). Also, there was found to be a higher level of

sequence similarity between the genomes of *C. milii* and human, with a greater number of CNEs between *C. milii* and human compared with human and teleost fish (Venkatesh et al., 2006, 2007). This could be an effect of an extra whole genome duplication event in teleost fishes, not present in other jawed vertebrates (Amores et al., 1998). A relative rates test on a protein alignment of 699 orthologous protein-coding genes from *C. milii* and 12 other vertebrates revealed that the *C. milii* genes were evolving much slower than other vertebrates, including the ‘living fossil’ coelacanth, and concluded that the genome of *C. milii* was the least derived of all sequenced vertebrate genomes to date (Venkatesh et al., 2014). These findings collectively suggest that this group of holocephalans represents a good reference for ancestral chondrichthyan and gnathostome genome evolution (Venkatesh et al., 2014).

Several researchers have understood the value in the *C. milii* genome since its assembly, and used the data to explore gnathostome evolution (Yu et al., 2008; Larsson et al., 2009; Ravi et al., 2009; Wang et al., 2009; Nah et al., 2014). In several cases, the genome was used to identify genes or gene clusters in *C. milii* and compare them with bony vertebrates. This led to a better understanding of the likely ancestral jawed vertebrate genome, as well as evolution in modern vertebrates, by identifying gains and losses of particular genes (Yu et al., 2008; Larsson et al., 2009; Ravi et al., 2009; Wang et al., 2009; Nah et al., 2014). The *C. milii* genome has retained more ancestral vertebrate genes than bony vertebrates (Venkatesh et al., 2007; Yu et al., 2008; Larsson et al., 2009; Ravi et al. 2009), making it likely that they exhibit a genome more similar to the ancestral vertebrate condition than other bony vertebrates. Several studies also identified CNEs, and found that the majority of CNEs in *C. milii* are conserved in mammals, which likely represent putative cis-regulatory elements (Ravi et al., 2009; Wang et al., 2009; Nah et al., 2014). Many of the CNEs shared between *C. milii* and mammals were found to be either lost or have diverged in teleost fishes (Wang et al., 2009; Nah et al., 2014). This indicates that teleost fish

genomes may be evolving at a significantly higher rate compared to *C. milii* and mammals (Wang et al., 2009).

Many molecular studies have shown and alluded that holocephalans have a basal position among chondrichthyans (e.g., Le et al., 1993; Arnason et al., 2001; Kikugawa et al., 2004; Chen et al., 2012). Additionally, the genome organization of *C. milii*, which is more similar to humans than humans to teleosts, and a slower rate of evolution, underscores the importance of holocephalans as a reference and/or an outgroup for vertebrate genome evolution and comparative genomics. However, despite their obvious importance, holocephalans have been poorly studied. No comprehensive framework for the inter-relationships among species has been resolved, which makes interpreting the ancestral state and genomic alterations within the group difficult.

Molecular Phylogenetics

There are many applications of phylogenetics, with the major goal to aid in the understanding of evolutionary origins, histories, diversifications, and directions of change in organismal features. Phylogenies provide a framework that allows us to infer information about patterns of evolution, the distribution of characters across groups of organisms, the structure and function of genomes, and timing of species diversification. Phylogenies also are fundamental in comparative genomic studies, as they provide important ancestor-descendant relationship information (Soltis and Soltis, 2003; Garamszegi & Gonzalez-Voyer, 2014).

By estimating the evolutionary relationships among the extant species of holocephalans, the phylogeny can then be used in future downstream analyses to explore patterns of change and innovation for traits of interest. Molecular phylogenetics is the study of evolutionary relationships among groups of organisms based on comparisons of homologous DNA sequences.

A phylogenetic tree provides a hypothesis about the evolutionary relationships of life based on similarities and differences in inherited genotypic characters (Figure 1.2). It is a hypothesis because we cannot go back in time and observe the evolution of species, and know with certainty their relationships. Our data also cannot provide certainty about the inter-relationships, and thus, a tree will only postulate a hypothesis about the relatedness of taxa. It is thought that understanding and utilizing phylogenies should be a fundamental aspect to many areas of biological research (Soltis & Soltis, 2003). First, phylogenetic trees provide the necessary data on the relationships among species, populations, genes, etc. But, more importantly, phylogenetic trees help us interpret patterns by providing the likely ordering of changes over evolutionary time. This ordering is useful because it shows where to look to find out how something functions. This is important because it allows one to make predictions since the tree can be used to infer when and where molecules, traits, processes, or behaviors evolved. Phylogenetic estimates can then be used to test evolutionary hypotheses in a comparative framework by mapping characters onto the topology and estimating the ancestral state (Figure 1.2). In addition, one may also gain a better understanding of evolutionary lability, and timing and rate of evolution of traits of interest. Since genetic changes have accumulated over time as speciation has occurred and species have diverged, a single extant species does not provide the best estimate of the ancestral state. Instead, by inferring the tree topology for all or most of the extant species, one can better estimate the character states that are ancestral versus putatively derived, which can then be used to explore trait evolution in a comparative context.

The primary goal of molecular phylogenetics is to reconstruct an accurate and well-supported phylogeny in order to understand the true evolutionary history of a particular group of organisms, and the structure and function of their genomes. Obtaining a reliable and accurate estimate of phylogeny is contingent upon several factors, including appropriate selection of

informative molecular markers, sufficient sequence data, adequate taxon sampling, and use of appropriate methods and evolutionary models (e.g., Swofford et al., 1996; Delsuc et al., 2005; Philippe et al., 2005a; Heath et al., 2008; Patwardhan et al., 2014).

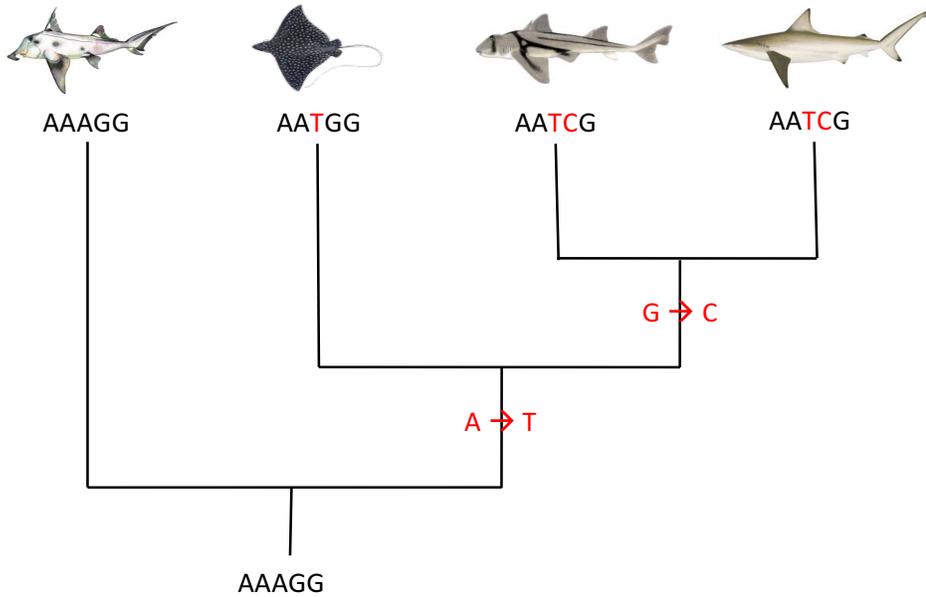


Figure 1.2 Schematic of a phylogenetic tree showing a hypothetical DNA sequence for 4 unique lineages, mapping of DNA sequence changes, and estimation of ancestral state.

Sparse taxon sampling can lead to a highly supported, yet incorrect, phylogenetic tree due to strong systematic biases (Heath et al., 2008). Thus, taxon sampling is an extremely important factor to consider when designing phylogenetic studies. Many studies on this topic have indicated that increasing taxon sampling increases the accuracy of phylogenetic estimation (Pollock et al., 2002; Pollock & Bruno, 2002; Zwickl & Hillis, 2002; Philippe et al., 2005a; Heath et al., 2008; Nabhan & Sarkar, 2011). This is because as additional taxa are added, there is an increase in the number of internodes, which can expose previously unknown substitutions, as well as spreading homoplasy across the tree (Pollock & Bruno, 2002; Heath et al., 2008; Nabhan &

Sarkar, 2011). This additional information, along with increased knowledge of site-specific rates, can improve accuracy by improving parameter estimation (Pollock & Bruno, 2002; Heath et al., 2008; Nabhan & Sarkar, 2011).

Additionally, one must also consider the goal of the phylogenetic study, and develop the appropriate molecular markers. Often, when one or a few genes are used, the result is insufficient character data to fully resolve the phylogenetic tree. Individual genes can often produce incongruent phylogenies from one another and from the species phylogeny due to systematic error (Fitch, 1970; Pamilo & Nei, 1988; Avise, 2004; Ballard & Whitlock, 2004; Patwardhan et al., 2014; Philippe et al., 2005a; Rubinoff & Holland, 2005; Galtier et al., 2009). It has been shown that using multiple, independent molecular markers may reduce error and produce a more accurate phylogenetic estimate with better resolution (Baptiste et al., 2002; Rokas et al., 2003; Driskell et al., 2004; Delsuc et al., 2005; Rokas & Carroll, 2005). Additionally, the type of marker must also be considered. The mitochondrial genome represents only a single marker, as it is a single, non-recombining locus, whereas the nuclear genome contains many potentially independent DNA markers, depending on the structure of genomic linkage groups. The substitution rate and quality of the signal within the markers of interest should be evaluated. Depending on the evolutionary depth of interest, ancient divergences may require more conserved genes, with slow rates of evolution that would have less homoplasy and enough information still present to resolve relationships. However, there would likely not be enough variation to resolve recent divergences, which may require less conserved genes with faster rates of evolution to provide enough variation to resolve the phylogeny. Thus, when aiming to estimate a phylogeny of a group of taxa where little information is known, it may be prudent to select an array of markers at different depths of evolutionary time, and compare the resulting topologies to gain better clarity as to the true species relationships.

Another important factor to consider are models of evolution used to describe DNA sequence evolution. The parameters of the model include the tree topology, branch lengths, and the parameters of the evolutionary model (e.g., transition/transversion ratio, base frequencies, among-site rate variation). When particular assumptions are violated (e.g., stationarity of base frequencies, equal substitution rate), this can result in error, and mislead the phylogenetic estimation (Swofford et al., 1996; Felsenstein, 2004). One way to minimize this error is to use appropriate models of evolution for the data at hand. The assumption of a homogeneous substitution rate across sites is often violated, as rate variation has been shown to occur within most genes and proteins (Yang, 1996). This rate variation at different sites is thought to be due to selective pressures associated with functional and/or structural constraints (Yang, 1996). Thus, the distribution of sites that are free to vary may be different within and between genes. Sites with no or little constraints are free to vary more, and will experience higher substitution rates (Yang, 1996). Whereas, constrained sites, which may see no or very little change, will have a much lower substitution rate. When among-site rate variation is present, but not incorporated into the model, this can impact the phylogenetic inference and parameter estimates (Yang, 1996; Sullivan & Swofford, 2001).

Nuclear and mitochondrial genes have independent gene histories, and in general have different rates of substitution. Using both types of markers may have value in that they are informative at different time depths (Avice, 2004). Also, the advantage in analyzing both types of data is the significance produced by shared clades. The probability of two random trees producing similar topologies and sharing clades is low. So, when there is congruence, this provides confidence that shared clades likely represent real clades on the species phylogeny (Cunningham, 1997; Rubinoff & Sperling, 2002; Rubinoff & Holland, 2005), since there is no gene history constraint, it must be shared species phylogeny constraint. On the other hand,

conflicts between the data sets can also provide unique insight into the evolutionary history of taxa (e.g., introgression, population structure, sex-biased gene flow) (Rubinoff & Holland, 2005).

Comparative genomic studies of humans and other bony vertebrates has significantly increased our understanding of the human genome and vertebrate genome evolution. A major goal is to reconstruct the history of vertebrate genomes in order to identify functional genes, regulatory regions, and determine function (Moreno et al., 2008; Yu et al., 2008). Comparative studies rely upon and require that genomes have a high degree of orthology and conservation between species (Alföldi & Lindblad-Toh, 2013; Meadows & Lindblad-Toh, 2017), and this can be used to identify conserved and divergent regions. Comparing the genome of other organisms to humans can provide insight into human physiology and disease, among many things (see Meadows & Lindblad-Toh, 2017). However, it is more powerful when data is available for multiple species within a group to better understand true changes and innovations (Meadows & Lindblad-Toh, 2017). Typically, teleost fishes have been used as a distantly related genome for comparisons with humans, and have been quite valuable for identifying novel genes (Aparicio et al., 2002; Jaillon et al., 2004). However, the teleost fish lineage has undergone an additional whole genome duplication event that may hinder human-teleost comparisons (Venkatesh et al., 2007). The many reasons stated above (i.e., basal gnathostome group, slow-evolving genome, similar genome organization to humans), make holocephalans a good lineage to use in comparative studies with humans. The problem is the general lack of information on the inter-relationships and evolutionary history within chimaeroid fishes.

Sex Determination in Chondrichthyans

Sex determination is a fundamental process in biology necessary for proper development and reproduction in sexually reproducing organisms. While several different sex-determining

mechanisms have been found in vertebrates studied thus far, little is known about the ancestral condition or how sex-determining mechanisms transition across vertebrate taxa. It results from an initial genetic or environmental signal that determines whether an embryo develops as male or female. There are two main types of primary sex determination: 1) genetic sex determination (GSD) where sex is determined either by the inheritance of heteromorphic or homomorphic sex chromosomes, and 2) environmental sex determination (ESD) where sex is determined by extrinsic factors such as temperature (TSD) or social interactions. Birds, mammals, and even crocodylians show extreme conservation in their sex determination mechanisms (Mank et al., 2006). However, other vertebrate lineages like teleost fishes, amphibians, turtles, and lizards exhibit a wide variety of sex-determining mechanisms within their respective groups (Mank et al., 2006; Bachtrog et al., 2014). To understand the process of sex determination, it is necessary to know both the genetic basis underpinning the mechanisms, as well as how the mechanisms have transitioned over time.

Despite their important basal position in the vertebrate phylogenetic tree, the mechanism of sex determination among chondrichthyans, and holocephalans in particular, is not yet known. Relatively few cytogenetic studies have been conducted on these fishes. There have been investigations into the presence of putative sex chromosomes in 21 species, 17 of which have apparent heteromorphic chromosomes (Donahue, 1974; Kikuno & Ojima, 1987; Asahida et al., 1993; Asahida & Ida, 1995; Maddock & Schwartz, 1996; da Cruz et al., 2011; Aichino et al., 2013; Valentim et al., 2013). Yet, in several of these cases the karyotypes were not published or the published karyotype was of poor quality limiting the ability to confirm the findings. This previous work suggests the potential of a genetic mechanism of sex determination in chondrichthyan fishes. However, no studies have attempted to identify genes or markers that may be responsible for sex determination in these fishes, nor confirm GSD.

Research Objectives

This dissertation aims to fill an important gap in our understanding of the evolutionary relationships among holocephalans fishes, as previous studies have lacked the resolution and taxon sampling to infer relationships below the family level. The resulting phylogenetic tree will provide a framework to estimate the ancestral conditions for traits or genes of interest within holocephalans fishes. Since their basal phylogenetic position makes them a useful reference for vertebrate evolution, understanding their patterns of evolutionary diversification is thus fundamental to our ability to estimate the ancestral state of jawed vertebrates. This can then be used comparatively across vertebrates to better understand how genomes have evolved and organismal processes have changed throughout vertebrate evolution. The goal is to obtain an accurate estimate of the evolutionary history among holocephalans lineages using genomic comparisons, which will in turn be used to provide information about patterns and timing of evolutionary transitions within this group. The objectives are to provide a comprehensive sampling of taxa and use two independent molecular data sets, one derived from the nuclear genome, and the other from the mitochondrial genome, to generate hypotheses of the evolutionary relationships among holocephalans.

This study also investigates one of the most important developmental mechanisms in biology, sex determination, within chondrichthyan fishes. Surprisingly, the mechanism of sex determination in chondrichthyans is not known. Previous research examining chromosome number and morphology in some shark and ray species revealed putative sex chromosomes, consistent with a genetic sex-determining mechanism. The goal is to screen a subset of the genome of male and female specimens of a representative chondrichthyan species to identify candidate sex-linked molecular markers. The hypothesis is that there is at least one sex-linked

molecular marker present in the male sex of the representative species, none in the female sex, indicating a genetic mechanism of sex determination. This will provide previously unknown and important information about the sex-determining mechanism and potential genes involved for this group. There is value in understanding the history of a developmental mechanism such as sex determination, in order to fully understand how it functions, how it has changed and transitioned throughout vertebrate evolution. Thus, future studies can use the findings in a comparative framework to better understand sex determination across vertebrates.

In summary, the overall goals of this dissertation research are to estimate an accurate phylogeny of extant chimaeroid fishes using nuclear and mitochondrial DNA markers, and then, use three species of chondrichthyans to explore the potential mode of sex determination by probing the genome for sex-linked markers. A cross-species hybridization gene capture approach will be used to target pre-specified nuclear and mitochondrial DNA genes. Chapter 2 will present the phylogeny and diversification of holocephalans using the nuclear data. Chapter 3 will estimate the phylogeny and divergence of holocephalans using mitochondrial protein-coding genes, and compare results to Chapter 2. Chapter 4 will present the results of investigating the sex-determining mechanism in chondrichthyan fishes through chromosome analysis and restriction site-associated DNA sequencing of males and females, along with validation of candidate sex-linked molecular markers. Chapter 5 will be a general discussion of the results, implications of these findings, and future directions of the work.

CHAPTER 2

A Comprehensive Phylogeny of Holocephalans (Chondrichthyes: Chimaeriformes) Using Multi-Locus Nuclear Data

Introduction

The living jawed vertebrates (Gnathostomes) are divided into two lineages, the cartilaginous fishes (Class Chondrichthyes) and the bony vertebrates (Osteichthyes), which are further divided into Actinopterygii (e.g. teleost fishes) and Sarcopterygii (e.g. lungfishes, coelacanths, tetrapods). Chondrichthyan fishes are a monophyletic group that includes sharks, skates, rays, and chimaeras (Schaeffer & Williams, 1977; Maisey, 1984, 1986; Le et al., 1993; Didier, 1995; Arnason et al., 2001; Douady et al., 2003; Kikugawa et al., 2004; Winchell et al., 2004; Grogan et al., 2012; Ward et al., 2005; Inoue et al., 2010; Licht et al., 2012). They are considered to be the most basal living group of jawed vertebrates, diverging from Osteichthyes approximately 462 to 421 million years ago (Ma) in the Paleozoic era (Benton et al., 2009). Extant chondrichthyans are divided into two lineages, the Elasmobranchii (e.g., sharks, rays) and the Holocephali (chimaeras, ghost sharks), which are thought to have diverged at least 400 Ma based on the earliest Chondrichthyes fossils from the Paleozoic (Sansom et al., 1996; Coates & Sequeira, 2001) and recent molecular divergence dating (Inoue et al., 2010; Licht et al., 2012).

Chondrichthyan fishes occupy an important phylogenetic position as the sister group to bony vertebrates, as well as the basal living clade of jawed vertebrates, as their lineage diverged early on in gnathostome history (Janvier, 1996; Kikugawa et al., 2004; Cole & Currie, 2007; Venkatesh et al., 2014). Thus, they represent an important reference for comparative studies aimed at understanding early jawed vertebrate evolution. This is especially true of holocephalans, which have been thought to be the more basal group of chondrichthyans and therefore potentially more representative of the basal gnathostome condition (de Beer & Moy-Thomas, 1935).

Due to their position in the vertebrate tree, and in an effort to better understand human and vertebrate genome evolution, Venkatesh et al. (2005) set out to identify a chondrichthyan model genome. The elephantfish, *C. milii*, was selected due to its small genome size (910 Mb),

and its genome sequenced with 1.4X coverage (Venkatesh et al., 2007). Interestingly, this revealed that the genome organization of chimaeras is more similar to tetrapods than teleost fishes, although tetrapods are more closely related to teleost fishes. Further, Venkatesh et al., (2014) concluded that the *C. milii* genome is the least derived and slowest evolving of known vertebrates, including the coelacanth, making holocephalans a good model for both ancestral chondrichthyans and vertebrate genome evolution.

The living holocephalans are a relatively small yet once diverse (Helfman et al., 2009) group of marine fishes found worldwide in all oceans, except for Antarctic waters (Didier et al., 2012). They generally inhabit the deep-sea, with a few species occurring in or migrating to shallower seas (e.g. *Callorhinchus* sp.). They are typically found near the bottom, along the shelf of continental landmasses, around islands, seamounts, and ridges (Didier et al., 2012). Adult size is highly variable, but most species reach a total length of between 50 cm and 100 cm, with some species reaching lengths greater than one meter. Overall, very little information exists on this group of fishes with regards to biology, ecology, and behavior. This is due to the deep-water habitat of the majority of species. However, in recent years, with more deep-sea exploration, more specimens have been collected and studied.

Holocephalans belong to the order Chimaeriformes with three families, each distinguished by a unique snout morphology (Didier et al., 2012). Callorhinchidae comprises one genus and three species, and is characterized by a plow-shaped snout. The three species, *C. milii* (New Zealand and Australia), *C. capensis* (southern Africa), and *C. callorynchus* (southern South America) are morphologically indistinguishable and have the same color pattern but are separated as unique species based on geographic distribution, all in the southern hemisphere (Figure 2.1). This genus is morphologically distinct in other aspects besides snout morphology, including a torpedo-shaped body, heterocercal tail, and a large anal fin (Didier et al., 2012).

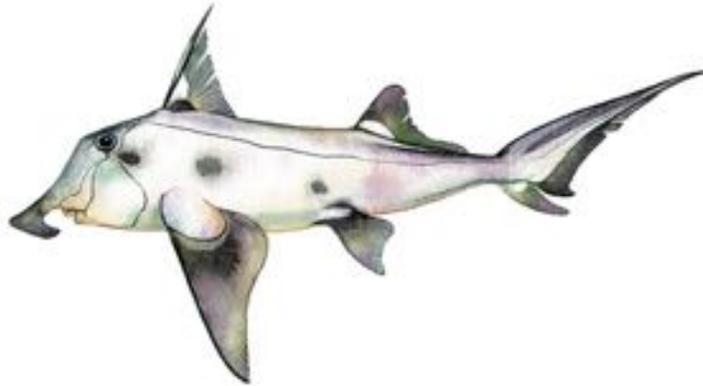


Figure 2.1 Drawing of *Callorhinchus* sp. Photo retrieved from <https://sharksrays.org>.

The family Rhinochimaeridae comprises three genera and eight described species, and is characterized by a long spear-like snout. The species contained within the Rhinochimaeridae range from medium to quite large in size and are found in Atlantic, Indian, and Pacific Ocean basins. This family is characterized by a more compressed, yet elongate body and narrow tail. The three genera have previously been separated into two distinct groups, the Rhinochimaerinae with *Rhinochimaeridae* (Figure 2.2), and Harriottinae with *Harriotta* and *Neoharriotta* (Didier, 1995); however, this is not supported by any synapomorphies. *Harriotta* and *Neoharriotta* are separated morphologically by the presence of a distinct anal fin in *Neoharriotta* (Figure 2.4) and absence in *Harriotta* (Figure 2.3). Recent molecular analyses have shown that these two groupings are likely inaccurate, where *Neoharriotta* and *Harriotta* species were not more closely related to one another (Inoue et al., 2010; Licht et al., 2012).

The family Chimaeridae contains two genera and 39 recognized species, characterized by a short, conical snout and an elongate body that tapers to a whip-like tail, often with a long filament. Species vary in size from small to very large, maturing at over 1 m in length, occurring in all ocean basins, except for far northern and southern polar waters (Didier et al., 2012). The two genera, *Chimaera* with 16 species and *Hydrolagus* with 23 species, are morphologically very



Figure 2.2 Photograph of a *Rhinochimaera atlantica* specimen. Photo credit: Rob Leslie.



Figure 2.3 Photograph of *Neoharriotta pinnata* specimen. Photo credit: KK Bineesh.



Figure 2.4 Photograph of a *Harriotta raleighana* specimen.

similar. They differ in the presence of a small anal fin separated from ventral caudal fin in *Chimaera* (Figure 2.5), and absence in *Hydrolagus* (Figure 2.6). However, this character has been shown to be both present and absent in specimens of a single species (Didier et al., 2008), questioning its usefulness as a diagnostic feature between genera. Additionally, many species are morphologically similar, and distributions may overlap, making species identification difficult.



Figure 2.5 Photograph of a *Chimaera notafriicana* specimen. Photo credit: Kristin Walovich.



Figure 2.6 Photograph of a *Hydrolagus africanus* specimen. Photo credit: Kristin Walovich.

Accurate species identification is necessary for understanding the diversity of this group of fishes, to study their biology and ecology, and assess population dynamics for fishery and conservation efforts. Unfortunately, many chimaeroid species are difficult to identify due to similarities in morphology, coloration, and distribution. This has led to taxonomic uncertainty and questions regarding the validity of some species level designations. For example, the three species of *Callorhinchus* are indistinguishable morphologically, yet are considered separate species based on locality. The only morphological character separating the genera *Chimaera* and *Hydrolagus* is the presence of a small anal fin in *Chimaera* that is absent in *Hydrolagus*. However, this trait was found to be plastic in at least one species, raising questions about its utility as a diagnostic character (Didier et al., 2008). These issues are compounded by the fact that there is little available data on intra-specific variation as most species descriptions are based on few individuals of usually large adults.

By estimating the evolutionary relationships among the extant species of chimaeroids, the phylogeny can then be used in future downstream analyses to explore patterns of change and innovation for traits of interest. One way to estimate a phylogeny is through molecular phylogenetics, the study of evolutionary relationships among groups of organisms based on comparisons of homologous DNA sequences. Phylogenetic trees provide a framework to interpret patterns by ordering inferred changes over time. A tree can then be used to test evolutionary hypotheses by mapping characters (e.g., DNA sequences, traits) onto the topology for extant taxa and using this information to estimate internode character states and the ancestral character state. In addition to determining ancestral states, one may also gain a better understanding of evolutionary lability, and timing and rate of evolution of traits of interest.

Most phylogenetic studies on chondrichthyans have focused on understanding the evolutionary relationships among sharks and rays. Typically, chimaeroids are only used as an

outgroup or few species are included in such studies (Le et al., 1993; Arnason et al., 2001; Douady et al., 2003; Kikugawa et al., 2004; Winchell et al., 2004; Naylor et al., 2005; Ward et al., 2005; Mallatt & Winchell, 2007; Heinicke et al., 2009; Vélez-Zuazo & Agnarsson, 2011; Naylor et al., 2012a). Most previous studies on chimaeroid systematics have focused on morphological characters of fossil taxa (e.g. Patterson, 1965; Maisey, 1986; Lund & Grogan, 1997), leaving the inter-relationships within the extant chimaeroids poorly studied and unresolved. Didier (1995) first examined the inter-relationships among extant Chimaeriformes by undertaking an extensive morphological study and interpreting their phylogenetic relationships based on comparative anatomy. Her results produced a higher-level phylogeny of chimaeroids, and suggested that Callorhynchidae is basal to a sister relationship between Rhinochimaeridae and Chimaeridae. However, the relationships at the generic and species level were not resolved and still remain unclear.

The few studies that have attempted to address the inter-relationships among chimaeras using molecular data only used a few species, or did not include all the genera (Ward et al., 2008; Inoue et al., 2010; de la Cruz-Aguero et al., 2012; Licht et al., 2012), resulting in unresolved topologies. These studies also only used mitochondrial markers. One study used a nuclear gene (*RAG-1*) along with mitochondrial markers to estimate divergence times of the major groups of chondrichthyans, including Chimaeriformes (Heinicke et al., 2009). Overall, these studies provided evidence that the three families of holocephalans are monophyletic. Callorhynchidae was recovered as the most basal clade, and Rhinochimaeridae and Chimaeridae formed a monophyletic group to the exclusion of Callorhynchidae, as shown in morphological studies (e.g., Didier, 1995; Stahl, 1999)

Molecular studies have found conflicting results in the relationship between the genera *Chimaera* and *Hydrolagus*. Some tree topologies based on mitochondrial markers suggest that

Chimaera and *Hydrolagus* may be paraphyletic genera (Ward et al., 2008; Licht et al., 2012; de la Cruz-Aguero et al., 2012). On the other hand, Inoue et al. (2010) recovered the two genera as monophyletic, but only used three Chimaeridae species. Previous work also found that within Rhinochimaeridae *R. pacifica* was closest to *H. raleighana*, to the exclusion of *N. pinnata* (Licht et al., 2012), while morphological work placed *Harriotta* and *Neoharriotta* into a group to the exclusion of *Rhinochimaera* (Didier, 1995). While these studies have provided insights into the family level relationships, there is a need for denser taxon sampling and the use of multiple genetic markers to help resolve within family and genera relationships.

The modern holocephalan lineage has an origin dating back at least to the Paleozoic era (Patterson, 1965; Didier, 1995; Grogan et al., 2012). They are believed to have evolved from the great diversity of holocephalans present in the Carboniferous (Grogan et al., 2012). By the end of the Permian, it is thought that most of the holocephalans lineages had become extinct (Stahl, 1999; Grogan et al., 2012). Due to the nature of their cartilaginous skeletons, fossil chimaeroids are mainly known from tooth plates, dorsal fin spines, and very rarely, skeletons, making it difficult to interpret their fossil record and evolutionary history (Ward & Duffin, 1989; Stahl & Chatterjee, 2003). Some lineages must have survived the end Permian mass extinction, as new holocephalans appeared by the late Triassic (Stahl, 1999). A combination of mitochondrial markers and one nuclear marker have been used to estimate divergence times of chimaeroids. The split between holocephalans and elasmobranchs has been estimated at a range of 494-410 (Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012), with fossil evidence suggesting a split ~ 410 Ma (Coates & Sequeira, 2001). The three families were estimated to have diverged in the Mesozoic. Callorhynchidae has a divergence time range of between 320-125 Ma, with a much older point estimate of 220 Ma by Heinicke et al. (2009), compared to 167 Ma and 187 Ma by

Inoue et al. (2010) and Licht et al. (2012), respectively. Rhinochimaeridae and Chimaeridae shared a most recent common ancestor between 182-51 Ma (Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012).

Chimaeroid fishes occupy a basal phylogenetic position among chondrichthyans, and collectively, chondrichthyans are the basal, sister group to bony vertebrates. Also, chimaeras appear to have a genome architecture that may be more similar to humans, than humans are to teleosts, as well as the least derived genome to date (Venkatesh et al., 2014). Collectively, this illustrates the usefulness of these fishes as references for both chondrichthyan and vertebrate genome evolution studies. Their phylogeny can be used to estimate internode and ancestral state conditions, tracing the evolution of particular genes or traits. The ancestral state reconstruction can help inform how these traits have evolved across vertebrates. For example, one could identify whether a particular gene or trait is conserved, and if so, speculate that the gene likely has a significant function across vertebrates, that can be used to inform future studies on this particular gene of interest. On the other hand, if the gene appears to have transformed over vertebrate lineages, this too provides insight into the gene at study; one can identify what changes have occurred among lineages, and use this information to study function, regulation, and implications for the study species. However, we currently know very little about their evolutionary relationships. Studies have offered little insight into the systematics of this group, relationships among the various genera and species are unresolved, and taxon sampling has been inadequate for a comprehensive reconstruction of their phylogeny. Adequate taxon sampling is an important factor in phylogenetic reconstruction. Insufficient taxon sampling can mislead the phylogenetic estimation, resulting in an inaccurate estimate of evolutionary relationships (see Heath et al., 2008). The addition of taxa helps to estimate model parameters more accurately, thus, increasing the accuracy of the phylogenetic inference (Pollock & Bruno, 2002; Heath et al.,

2008; Nabhan & Sarkar, 2011). Previous molecular studies on chimaeroids mainly used mitochondrial genes to estimate their phylogenies. The mitochondrial genome is a single molecular marker, and thus represents only that particular evolutionary history. A single molecular marker may not necessarily reflect the true species history (gene tree-species tree incongruence), and the resulting phylogeny may be misleading (Avice, 2004; Ballard & Whitlock, 2004; Patwardhan et al., 2004; Philippe et al., 2005a; Rubinoff & Holland, 2005; Galtier et al., 2009). It is considered by many that estimating evolutionary relationships based on one or a few markers is inadequate for accurate gene tree reconstruction (Cummings et al., 1995; White et al., 2017). Multi-locus data (i.e., independent nuclear markers) has been shown to increase phylogenetic accuracy (Rokas & Carroll, 2005). Thus, the objective of this chapter was to use multi-locus nuclear genomic information and dense taxon sampling to obtain an accurate estimate of the evolutionary history among chimaeroid lineages and estimate times of divergence, providing critically important information about the pattern and timing of evolution for this group.

Materials and Methods

Taxon Sampling

A schematic of taxon identification through data set acquisition is presented in Figure 2.7. A total of 402 chimaeroid samples were collected, sampled, and sent to the lab by various individuals on research cruises, commercial fishing vessels, and from fish markets in the Atlantic, Indian, and Pacific Oceans. White muscle tissue was taken from each specimen in the field and stored in 95% ethanol at 4°C until processing in the laboratory. Total genomic DNA was extracted using the EZNA® Tissue DNA Kit (Omega Bio-Tek, Norcross, GA) and stored at -20°C. The mitochondrial NADH dehydrogenase subunit 2 (*NADH2*) gene was initially chosen to

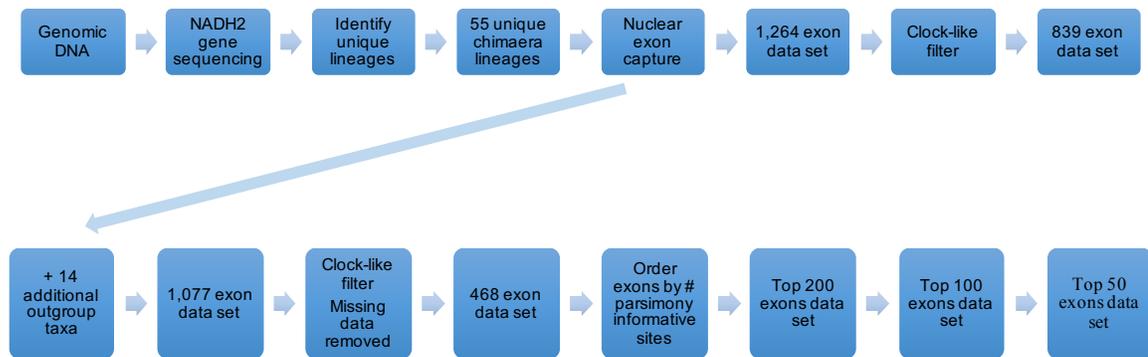


Figure 2.7 Schematic of workflow from genomic DNA extraction to taxon identification and the different nuclear exon data sets.

identify unique lineages and assess inter- and intra-specific variability, as it is fast-evolving, protein-coding, and considered a useful marker for species differentiation (Naylor et al., 2012b). Universal primers designed to target the complete coding sequence for *NADH2* following Naylor et al. (2005) or primers designed specifically for the *NADH2* region of chimaeroid fishes (Kemper et al., 2014) were used along with TaKaRa Ex Taq® (Clontech Laboratories, Inc). Polymerase chain reaction (PCR) was carried out in a 25 µl volume reaction (1x TaKaRa buffer, 2 to 3.5 mM MgCl₂, 200 µM of dNTPs, 0.32 µM forward and reverse primer, 0.625 units of TaKaRa Ex Taq, PCR grade water, and undiluted DNA template). The reaction was denatured at 94°C for 2 min, subjected to 30 cycles at 94°C for 30 sec, 50-58°C for 30 sec, and 72°C for 1 min, followed by 72°C for 5 min. A 3 µl sample of the PCR product was subjected to gel electrophoresis (1% agarose) to assess the effectiveness of the PCR amplification. Successfully amplified samples were sent for DNA sequencing (Retrogen, San Diego, CA). The software program Geneious® v.6.1.7 (Biomatters Ltd., Auckland, NZ) was used to view nucleotide and translated sequences, assess quality, and make nucleotide base calls. Translated sequences were aligned using the Geneious® alignment algorithm, which uses a progressive pairwise method to

first align the most similar sequences, and then successively less similar sequences based on a neighbor-joining guide tree (cost matrix = 65% similarity; gap open penalty = 12; gap extension penalty = 3; alignment type = global alignment with free end gaps), and back translated to nucleotide sequences (402 taxa, 1044 bp). Aligned sequences were subjected to a neighbor-joining analysis under the general-time reversible (GTR) substitution rate model using PAUP* v.4.0a147 (Swofford, 2002). Since the goal was to identify potentially unique lineages and not estimate the relationships among species at this point, a simpler method was chosen. The GTR model was used as it allows for different rates of substitution among nucleotides. Samples included specimens nominally identified as 41 of the 50 recognized species based on morphology, as well as unidentified chimaeroid specimens. The *NADH2* gene was then used to confirm taxon identification and identify potentially unrecognized diversity, using a reference database within the laboratory and vouchered specimens through museums and photographic images. A voucher is a representative of a particular organism that has been collected, expertly identified, and in many cases preserved and deposited in a permanent collection (i.e., museum) for authentication and future research. In total, 55 potentially unique chimaeroid lineages were chosen for further data collection and analysis, including 41 recognized species and 15 additional taxa (Table 2.1). Either five or nineteen representative elasmobranch species (Table 2.2) were chosen as outgroups to root the phylogenetic tree, depending on data set and analysis. Elasmobranchs were chosen to root the tree because they are the closest relatives of chimaeras, samples were available in the lab, and the RNA baits, described below, used in gene capture had been designed and successful at collecting the same exon and mitochondrial genome data as for the chimaera samples.

Nuclear Gene Capture Method

A modified method of cross-species hybridization gene capture was used to generate data for 1,264 slow-evolving, nuclear, putatively single-copy exons, shared across six jawed vertebrate genomes (Li et al., 2013). Library preparation for Illumina sequencing of multiplexed gene capture was modified from the method of Meyer and Kircher (2010), and follows Li et al. (2013), using the “with-bead” method (Fisher et al., 2011). Briefly, extracted genomic DNA (0.5 – 3 µg) was sheared to approximately 500 bp using a Covaris M220 focused-ultrasonicator (Covaris Inc, Woburn, MA) to form a library for each sample. Overhangs on DNA fragment ends were removed or filled in to leave blunt ends necessary for adapter ligation. Illumina-specific adapters were then ligated to each end. DNA fragments were PCR amplified to enrich the libraries before hybridization gene capture. Custom 5'-biotinylated RNA baits were designed for each of the 1,264 targeted exons based on the *C. milii* genome using the MYBaits target enrichment system (MYcroarray, Ann Arbor, MI). Targeted DNA gene capture hybridization followed the method described by Li et al. (2013), where amplified target DNA libraries were hybridized with the *C. milii* nuclear RNA baits. Two rounds of DNA hybridization capture were performed, as Li et al. (2013) found that two rounds of gene capture, using product from the first round of capture as template for a second round, increased the number of captured targets. The captured libraries were PCR amplified to incorporate unique indexes. Approximately fifteen indexed samples were pooled in equimolar ratios. Pooled libraries were prepared for sequencing on an Illumina MiSeq benchtop sequencer (Illumina Inc, San Diego, CA) by diluting to ~ 8–13.3 pM for 2 x 300 bp sequencing with the MiSeq Reagent kit v3 (600 cycles).

Table 2.1 Holocephalan lineages sampled for nuclear and mitochondrial gene capture. * = Other specimen(s) in the same clade based on *NADH2* gene analysis are vouchered; ** = photograph available of another specimen(s) in the same clade.

Species	GN #	Sample Locality	Vouchered	Photograph
<i>Callorhynchus callorynchus</i>	15484	South Atlantic Ocean, Argentina	No	No
<i>Callorhynchus capensis</i>	7200	Indian Ocean, South Africa	No	No **
<i>Callorhynchus milii</i>	6995	Tasman Sea, New South Wales	CSIRO H 3777-02	Yes
<i>Chimaera argiloba</i>	11101	Indian Ocean, Western Australia	CSIRO H 7140-06	Yes
<i>Chimaera carophila</i>	12993	South Pacific Ocean, New Zealand	NMNZ P.40174	Yes
<i>Chimaera cf phantasma</i>	4387	North Pacific Ocean, Philippines	No	Yes
<i>Chimaera cubana</i>	12746	Caribbean Sea, Nicaragua	No	Yes
<i>Chimaera fulva</i>	10943	Tasman Sea, Tasmania	CSIRO H 7052-04	No
<i>Chimaera lignaria</i>	6992	South Pacific Ocean, New Zealand	NMNZ P.37873	Yes
<i>Chimaera macrospina</i>	10955	Indian Ocean, Western Australia	CSIRO H 6417-02	Yes
<i>Chimaera monstrosa</i>	12209	Mediterranean Sea, Corsica	No	No
<i>Chimaera notafriicana</i>	14838	Indian Ocean, South Africa	No *	Yes
<i>Chimaera obscura</i>	10957	Tasman Sea, New South Wales	CSIRO H 1383-02	Yes
<i>Chimaera opalescens</i>	13522	North Atlantic Ocean, Porcupine Bank	MNHN 2007-1557	Original description
<i>Chimaera panthera</i>	6987	South Pacific Ocean, New Zealand	NMNZ P.037258	Yes
<i>Chimaera panthera</i>	10951	Tasman Sea	CSIRO H 5911-01	Yes
<i>Chimaera phantasma</i>	9968	North Pacific Ocean, Taiwan	No	Yes
<i>Chimaera phantasma</i>	10131	North Pacific Ocean, Taiwan	No	Yes
<i>Chimaera</i> sp. 3	15977	South Indian Ocean	MNHN 2005-1749	Yes
<i>Chimaera willwatchi</i>	10953	Southwestern Indian Ocean	CSIRO H 5356-01	Yes
<i>Chimaera didierae</i>	11724	Indian Ocean, Mauritius	CAS 242334	Yes
<i>Chimaera buccanigella</i>	11492	Indian Ocean, Mauritius	CAS 242335	Yes
<i>Hydrolagus affinis</i>	3774	North Atlantic Ocean, Portugal	No *	No **
<i>Hydrolagus africanus</i>	10493	South Africa	No *	No **
<i>Hydrolagus alberti</i>	12713	Caribbean Sea, Honduras	No	Yes
<i>Hydrolagus barbouri</i>	13453	North Pacific Ocean, Japan	NSMT P76382	Yes
<i>Hydrolagus bemisi</i>	6982	South Pacific Ocean, New Zealand	NMNZ P.37869	Yes
<i>Hydrolagus bemisi</i>	6988	South Pacific Ocean, New Zealand	NMNZ P.37868	Yes

<i>Hydrolagus cf lemures</i>	11228	Pacific Ocean, Indonesia		CSIRO - C20	Yes
<i>Hydrolagus cf trolli</i>	14842	Indian Ocean, South Africa		No	Yes
<i>Hydrolagus colliei</i>	6715	North Pacific Ocean, Washington, USA		UW47649	Yes
<i>Hydrolagus homonycteris</i>	13674	Tasman Sea, Southern Australia		CSIRO H 7534-01	Yes
<i>Hydrolagus lemures</i>	10970	Indian Ocean, Western Australia		CSIRO H 6579-13	No
<i>Hydrolagus melanophasma</i>	12565	North Pacific Ocean, California, USA		No	Yes
<i>Hydrolagus mirabilis</i>	12731	Caribbean Sea, Costa Rica		No *	Yes
<i>Hydrolagus mitsukurii</i>	9973	North Pacific Ocean, Taiwan		No	Yes
<i>Hydrolagus novaezealandiae</i>	2704	South Pacific Ocean, New Zealand		No *	No **
<i>Hydrolagus ogilbyi</i>	10982	Tasman Sea, Victoria		CSIRO H 5322-02	Yes
<i>Hydrolagus pallidus</i>	13689	North Atlantic Ocean, Portugal		No *	No **
<i>Hydrolagus purpurescens</i>	15894	South Indian Ocean		MNHN 2004-0820	Yes
<i>Hydrolagus erithacus</i>	10465	South Atlantic Ocean, South Africa		SAIAB 200579	Yes
<i>Hydrolagus</i> sp. B	12561	Indian Ocean, Reunion Island		No	Yes
<i>Hydrolagus</i> sp. C	11755	Indian Ocean, Mauritius		No	Yes
<i>Hydrolagus trolli</i>	6983	South Pacific Ocean, New Zealand		NMNZ P.37313	Yes
<i>Hydrolagus trolli</i>	10975	Tasman Sea, Victoria		CSIRO H 5304-22	No **
<i>Hydrolagus</i> sp. D	15425	Indian Ocean, Kerala, India		No	No
<i>Harriotta haeckeli</i>	12991	South Pacific Ocean, New Zealand		NMNZ P.046910	Yes
<i>Harriotta raleighana</i>	16084	North Atlantic Ocean, Scotland		No	No
<i>Harriotta raleighana</i>	6976	South Pacific Ocean, New Zealand		No	No
<i>Neoharriotta carri</i>	12725	Caribbean Sea, Panama		No	Yes
<i>Neoharriotta pinnata</i>	14631	Arabian Sea, Kochi, India		No	No **
<i>Rhinochimaera africana</i>	10586	North Pacific Ocean, Japan		No *	No **
<i>Rhinochimaera atlantica</i>	1083	North Atlantic Ocean, Connecticut, USA		YPM 11254	Yes
<i>Rhinochimaera atlantica</i>	14847	Indian Ocean, South Africa		No	Yes
<i>Rhinochimaera pacifica</i>	10990	Tasman Sea, Tasmania		CSIRO H 6967-01	Yes

Table 2.2 Elasmobranch species chosen for outgroup in nuclear and mitochondrial capture. Species with a • next to the species name indicated outgroup species used in all phylogenetic methods.

Order	Species	GN #	Sample Locality
Heterodontiformes	<i>Heterodontus mexicanus</i>	5224	Northeast Pacific Ocean, Gulf of California
	<i>Heterodontus portusjacksoni</i> •	2316	New South Wales, Australia
Orectolobiformes	<i>Orectolobus hutchinsi</i> •	4847	Indian Ocean, Western Australia
Lamniformes	<i>Alopias pelagicus</i>	13878	Northwest Pacific Ocean, Taiwan
	<i>Mitsukurina owstoni</i>	3164	No locality data available
Carcharhiniformes	<i>Hemipristis elongata</i>	12267	United Arab Emirates
	<i>Mustelus griseus</i> •	10111	Taiwan
Squaliformes	<i>Deania calcea</i>	1710	Japan
	<i>Etmopterus spinax</i>	5162	Portugal
Pristiophoriformes	<i>Pliotrema warreni</i>	7338	South Africa
	<i>Pristiophorus japonicas</i>	1049	Japan
Rhinopristiformes	<i>Rhynchobatus immaculatus</i>	10067	Taiwan
	<i>Rhynchobatus laevis</i>	12249	United Arab Emirates
Myliobatiformes	<i>Mobula eregoodootenkee</i> •	15461	South Africa
	<i>Mobula mobular</i>	15654	Spain
Torpediniformes	<i>Narcine entermedor</i>	5446	Northeast Pacific Ocean, Gulf of California
	<i>Narcine tasmaniensis</i>	2566	New South Wales, Australia
Rajiformes	<i>Leucoraja erinacea</i>	4694	Northwest Atlantic Ocean
	<i>Pavoraja niida</i> •	2574	New South Wales, Australia

Read assembly and alignment

Sequence reads were de-multiplexed to sort reads into their respective samples by indices using Illumina MiSeq Reporter. Adapter sequences and low quality reads were trimmed using Cutadapt within the wrapper script Trim Galore! v.0.3.1 (Krueger, 2012). Trimmed sequence reads were assembled *de novo* into contigs for each sample in ABySS v.1.3.6 (Simpson et al., 2009; k-mer values = 51 to 251, increments of 10). Assembled contigs were filtered, extended and merged using Trans-ABYSS v1.4.4 (Robertson et al., 2010). Assembled contigs were validated using HaMStR v.13.2.4 (Ebersberg et al., 2009). HaMStR begins by introducing a taxa set, the six model vertebrate taxa used to identify targeted exons, and core-ortholog groups, the DNA sequences of each targeted exons for each taxa. The sequences of the model taxa were then aligned and compiled into a database of profile hidden Markov models (pHMMs) for each of the 1,264 orthologous sequence groups. Assembled contigs were translated into protein sequences and searched for matches to the individual pHMMs. Each contig that had a match to one of the pHMMs (e-value < $1.0e^{-05}$) was conditionally assigned to that ortholog group. Each conditional match was then compared to *C. milii* using BLASTP (Altschul et al., 1997) following a reciprocal best BLAST hit criterion. Since *C. milii* should be the closest taxon, if the best BLASTP hit was the sequence from *C. milii* that contributed to the pHMM (e-value < $1.0e^{-05}$), the contig was retained as an ortholog. The retained contigs were back-translated to nucleotide sequences, concatenated and aligned using custom Perl scripts, which used the MAFFT alignment algorithm (Katoh & Standley, 2013).

Nuclear Data Sets

In total, seven different nuclear data sets ranging from 1,264 to 50 exons were used to some extent for phylogenetic analysis. Originally, the full nuclear data set (1,264 exons) and the

clock-like filtered full data set (839 exons) were used to reconstruct the phylogeny. If available computing resources were unlimited, these two data sets would have been explored in more detail, including divergence time estimation. However, currently available computing infrastructure was unable to complete a full suite of phylogenetic analyses and divergence time estimation on these two data sets due to the amount of data. Also, additional outgroup taxa were included for the divergence time analysis, which resulted in a new nuclear data set of 1,077 exons. This data set was further reduced by filtering till a small enough data set was reached that still resulted in the same overall phylogeny as the larger data sets, but could be used in divergence time analysis in a timely manner.

The full nuclear data set consisted of 1,264 exons with 55 chimaera species and five outgroup elasmobranchs. The full nuclear data set was then subjected to a clock-like filter in Paup*. This was applied in order to remove potentially paralogous sequences that may have resulted from duplication events within chimaeroids. In the event of a duplication, the ortholog and paralog will begin to diverge over time. If you have a speciation event, and compare a gene paralog in one species to the ortholog in the other species, it may appear to have an accelerated rate of evolution, indicated by a long branch. This is due to the fact that the ortholog and paralog diverged further back in time, before the speciation event, and would likely have accumulated more changes between genes. Whereas if you compared the orthologous gene in the two species, their divergence occurred more recently, and thus, less change has occurred. Using a clock-like filter should help to remove exons that appear to have accelerated rates of evolution in some lineages. Comparison of clock and non-clock models for each exon along an input maximum likelihood tree was carried out using likelihood ratio tests (LRT). If there was a significant difference (P -value < 0.05) in the log-likelihood values as evaluated by the LRT, indicating rejection of a clock model, those exons were removed from further analyses. In order to proceed

with divergence time analysis, an additional fourteen elasmobranchs were chosen as they have reliable fossil information along their lineages. With this addition, a new alignment of the nuclear exon data was conducted, which consisted of 1,077 exons. This reduced data set from the original 1,264 exons was due to missing data from the intersection of targeted exons between chimaeras and the additional sharks and rays. The 1,077 exon data set was too computationally intensive for divergence time analysis on available computing resources, so a smaller set of exons was identified that was consistent overall with the previous results, yet small enough to use in the analysis. The 1,077 exon set was first filtered for missing data using the program FASconCAT-G (Kück & Longo, 2014) along with scanning the alignment by eye, and then subjected to the clock-like filter in Paup*. The result was 468 exons that were concatenated using FASconCAT-G. Three additional data sets were explored by ordering the number of parsimony informative sites (PIS) of the 468 exon set and choosing the top 200, 100 and 50 exons. FASconCAT-G was used to determine the number of PIS per exon, and concatenate the data sets.

Model Choice and Partitioning

The best-fitting partitioning scheme and model of molecular evolution for nucleotide and amino acid alignments for each data set was selected in PartitionFinder v.1.1.0 (Lanfear et al., 2012) using the Bayesian information criterion (BIC). This was carried out for both maximum likelihood and Bayesian inference methods. This method was used to identify an appropriate model of evolution for the alignment to be used in subsequent analyses as well as partitioning of the data set by splitting an alignment into subsets to account for any heterogeneity among sites. A greedy algorithm was employed, and branch lengths were “linked”. Partitioning schemes that were evaluated included by codon position (nucleotide only), by gene (nucleotide and amino

acid), and by gene + codon (nucleotide only). PartitionFinder also was used to determine the best-fitting model of evolution for un-partitioned data sets.

Maximum Likelihood Analyses

Maximum likelihood (ML) analyses were conducted in RAxML v.8.0.26 (Stamatakis, 2014), and using RAxML within the Cyberinfrastructure for Phylogenetic Research (CIPRES) gateway (Miller et al., 2010). All nucleotide alignments used the general time reversible (GTR) substitution model (six substitution rate parameters, allows for unequal base frequencies) + gamma parameter (Γ), which models substitution rate heterogeneity over sites + invariable sites (I), which is the fraction of sites assumed to be invariable (Table 2.3). Amino acid alignments used either the JTT (Jones et al., 1992) or LG (Le & Gascuel, 2008) substitution model and a combination of Γ , I, and F (empirical base frequencies; Table 2.3). RAxML uses a rapid hill-climbing algorithm, and builds a starting tree using randomized stepwise addition order Parsimony. Trees are optimized using lazy sub-tree rearrangement (LSR), model parameters estimated for each inference, and log-likelihood scores obtained. Nonparametric bootstrap support values for nodes were obtained using 1,000 replicates in RAxML.

Bayesian Inference

Bayesian inference (BI) was implemented in the program MrBayes 3.2.6 (Ronquist et al., 2012) in CIPRES to estimate the posterior probability distribution of evolutionary model parameters, tree topology, and branch lengths. A list of all BI analyses conducted can be found in Table 2.4. Metropolis-coupled Markov chain Monte Carlo (MC)³ algorithm was used with two

Table 2.3 Maximum likelihood analyses conducted on the nuclear gene captured exons. All analyses were conducted in the program RAxML. Data set indicates the exon set and character type used. For model details, see methods.

Data Set	Model	Partition	Inferences	Bootstrap Replicates
1264 exons - Nucleotide	GTR+I+G	None	200	1000
1264 exons - Nucleotide	GTR+I+G	Codon	1000	1000
1264 exons - Amino Acid	JTT+I+G+F	None	200	1000
Clock-filter - Nucleotide	GTR+I+G	None	1000	1000
Clock-filter - Nucleotide	GTR+I+G	Codon	1000	1000
Clock-filter - Amino Acid	JTT+I+G+F	None	1000	1000
1077 exons - Nucleotide	GTR+I+G	Codon	200	1000
468 exons - Nucleotide	GTR+I+G	None	200	-
200 exons - Nucleotide	GTR+I+G	None	200	-
100 exons - Nucleotide	GTR+I+G	None	200	-
50 exons - Nucleotide	GTR+I+G	None	1000	1000
50 exons - Nucleotide	GTR+I+G	Codon	1000	1000
50 exons - Nucleotide	GTR+I+G	Gene	200	1000
50 exons - Nucleotide	GTR+I+G	Gene + Codon	200	1000
50 exons - Amino Acid	JTT+I+G	None	1000	1000
50 exons - Amino Acid	JTT+G; LG+G	Gene	200	1000

Table 2.4 Bayesian inference analyses conducted on the nuclear gene captured exons. All analyses were conducted in the program MrBayes. Data set indicates the exon set and character type used. For model details, see methods.

Data Set	Model	Partition	Runs	Generations/run	Sample Frequency
Clock-filtered - Nucleotide	GTR+I+G; SYM+G	Codon	2	10,000,000	1000
50 exons - Nucleotide	GTR+I+G	None	2	20,000,000	2000
50 exons - Nucleotide	GTR+I+G; SYM+G	Codon	2	20,000,000	2000
50 exons - Amino Acid	JTT+I+G	None	2	9,730,000	1000

parallel runs, each with one cold and seven heated chains. Chains ran for approximately 9 million to 20 million generations, depending on data set, with a burn-in of 25%, and sampled every 1,000th to 2,000th generation. The starting tree was random with arbitrary values for branch lengths and model parameters. Prior distributions for all model parameters were set as the default in MrBayes, which indicated no prior knowledge of parameters. Topology prior was set as uniform distribution; branch lengths set as unconstrained, gamma-dirichlet (1, 0.1, 1, 1); four stationary nucleotide frequencies set as dirichlet (1, 1, 1, 1); six substitution rates set as dirichlet (1, 1, 1, 1, 1, 1); shape parameter set as exponential (1); proportion of invariable sites set at uniform (0, 1). Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to visualize the (MC)³ output, the parameter values sampled from the chain, check mixing and convergence, and evaluate the burn-in. Samples were summarized using histograms, trace plots, means, and credible intervals to assess mixing and convergence in Tracer. All estimated parameters from each run showed good mixing and convergence onto their respective posterior distributions as evident by ESS values > 200, Potential Scale Reduction Factor (PSRF) approaching one, and visualization of trace plots and distributions. A 50% majority rule consensus tree was used to summarize trees in MrBayes. The tree, along with associated posterior probability values and branch lengths was visualized in FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Divergence Time Estimation

Divergence times of the sampled lineages were estimated using a Bayesian relaxed molecular clock method under an uncorrelated lognormal distribution for lineage-specific rate heterogeneity in the program BEAST v2.4.3 (Bouckaert et al., 2014). An additional five vertebrate species, *Gallus gallus*, *Anolis carolinensis*, *Homo sapiens*, *Xenopus tropicalis*, and *Danio rerio*, were chosen to be included in the divergence time tree estimation. These five

vertebrates were chosen as they had high quality genomes available at the time and were used to identify and design the nuclear exons and baits (Li et al., 2013), and thus, the nuclear exon data was readily available. An initial pilot run of the analysis resulted in a very old divergence time estimate for the split between holocephalans and elasmobranchs. Additional vertebrate species were added to the alignment, as reliable fossil information is present for some of the splits within these species, which may be useful due to the lack of abundant fossil calibrations within holocephalans. A new nuclear exon alignment was made in the same manner as stated above, with the addition of the five vertebrate species. The same set of 50 exons was chosen for this analysis. The program AMAS (Borowiec, 2016) was used to split the alignment into individual exons. Then FASconCAT-G was used to concatenate the 50 exons. A ML analysis was conducted in RAxML under the GTR + Γ + I model, with a codon partitioning scheme, with 200 individual runs. The 50 exon data set was partitioned by codon for the BEAST analysis. BEAUti v2.4.3 (Bouckaert et al., 2014) was used to generate input files for BEAST.

In order to estimate divergence ages of nodes, external information is needed to calibrate the phylogenetic tree, and one way is to use fossil information (Bromhan & Penny, 2003). Reliable fossil dating information from the literature is first identified. Fossils are typically assigned to a particular group of extinct or extant organisms based on morphological similarities. If we know that a fossil belongs to a particular lineage, and may be an ancestor, we can use the information to place a constraint on that lineage in order to scale the tree to actual time. Since we know that the fossil specimen was alive at a certain date in time, we can use this as a minimum age for that particular node. Fifteen fossil calibrations were assigned to appropriate nodes based on the confidence of fossil record (Benton et al., 2009; Underwood, Pers. Comm.). Fossil calibration ages and prior settings can be found in Table 2.5. Calibration fossils were used to estimate dates for the most recent common ancestor (MRCA) of the respective lineages.

The best scoring ML tree under a codon partitioning scheme for the nucleotide 50 exon data set was used to make a starting tree. The R package ‘ape’ v.3.4 (Paradis et al., 2004), using the ‘chronos’ function, was used to create an ultrametric tree from the ML tree. Figtree was used to visualize the branch lengths of the nodes used in fossil calibrations. The program Mesquite v3.2 (Maddison & Maddison, 2017) was used to re-scale the branch lengths on the ultrametric tree so calibration fossil ages fit appropriately. Each partition was analyzed under a separate GTR+ Γ +I model, while clock and tree models remained linked, respectively. Substitution rate, shape, proportion of invariant sites, GTR substitution model parameters, and frequencies were set to "estimate", and the gamma category count set to 4. The relaxed lognormal clock model was selected, and clock rate set to “estimate”. The birth-death model was used for the tree prior to model branching rates on the tree. Tree model priors were set as default: birthRate2.t:tree Uniform distribution, initial value 1.0, [0, 1000] lower and upper bound; reltiveDeathRate2.t:tree uniform distribution, initial value 0.5, $[-\infty, \infty]$. Default evolutionary model parameter priors were used: gamma shape parameter with an exponential distribution, initial value 1, $[-\infty, \infty]$; proportion invariant sites had a uniform distribution, initial value 0.05, [0, 1.0]; substitution rates had a gamma distribution, initial value 1.0, [0, ∞]. Hyperpriors on the clock model were also set as default: uclMean.c:Clock uniform distribution, initial value 1.0, $[-\infty, \infty]$; uclStdev.c:Clock gamma distribution, initial value 0.1, [0 ∞]. An exponential prior distribution was set for time to most recent common ancestor (tmrca) for each of the fossil calibration nodes and monophyly was enforced. A soft minimum age constraint (exponential offset) was used for each prior, and a soft maximum age constraint was used by setting the exponential mean value so that 95% of the distribution lies between the minimum and maximum age constraint.

Table 2.5 Fossil calibrations used in divergence time estimation. Fossil node number refers to the corresponding node on the divergence time tree where the fossil calibration was placed. Minimum and maximum soft age constraints are given in millions of years (Ma). The exponential prior distribution used in the analysis for each fossil is given by the exponential offset and mean.

Fossil	Lineage	Minimum age (Ma)	Maximum age (Ma)	Exponential Offset	Exponential Mean	Min. age reference
2	Chimaeriformes	280	462.5 ^a	280	60.7	Coates et al., 2017
3	Callorhynchidae	145	280 ^b	145	45	Underwood, Pers. Comm.
6	Rhinochimaeridae	152.5	280 ^b	152.5	42.5	Underwood, Pers. Comm.
6	Chimaeridae	112	280 ^b	112	56	Nesov & Averainov, 1996
61	Lamniformes	135	197 ^c	135	20.7	Rees, 2005
61	Carcharhiniformes	41	197 ^c	41	52	Underwood et al., 2011
59	Heterodontiformes	156	197 ^c	156	13.6	Underwood, 2002
70	Mobulidae	56	180 ^d	56	41.5	Cappetta, 2012
70	<i>Rhynchobatus</i>	52	180 ^d	52	42.7	Cappetta, 2012
68	<i>Narcine</i>	92	180 ^d	92	29.4	Claeson et al., 2013
69	Rajiformes	72	180 ^d	72	36	Siverson & Cappetta, 2001
76	Crown tetrapods	330.4	350.1 ^e	330.4	6.5	Benton et al., 2009
77	Amniotes	312.3	330.4 ^e	312.3	6	Benton et al., 2009

^a Maximum age based on stem gnathostome fossil evidence in Benton et al. 2009.

^b Maximum age based on fossil dating MRCA of Chimaeriformes (Coates et al., 2017).

^c Maximum age based on fossil dating MRCA of Galea (Duffin & Ward, 1983).

^d Maximum age based on Batoidea fossil skeleton (Duffy, Pers. Comm.).

^e Maximum age based on fossil evidence given in Benton et al. 2009.

Four independent Markov Chain Monte Carlo (MCMC) BEAST runs were conducted, each with 400 million generations, sampled every 1,000th generation. Each individual run was checked in Tracer to assess proper mixing and convergence on to the posterior distribution. An effective sample size (ESS) value > 200 is assumed to represent adequate sampling from the posterior distribution for each parameter. Tree and log files from the four independent runs were combined in LogCombiner v2.4.3 (<https://github.com/CompEvol/beast2/releases>), resampling states at a lower frequency of 5000, with a 25% burn-in. The combined log file was visualized in Tracer to assess model parameter values, node-height estimates, summary statistics, and trace files of parameters. TreeAnnotator v.2.4.3 (<https://github.com/CompEvol/beast2/releases>) was used to summarize the posterior probability density of the combined tree file as a maximum clade credibility tree. FigTree was used to visualize the mean and 95% highest posterior density (HPD) limits of node heights (divergence time estimates) and the posterior probabilities of the nodes.

Results and Discussion

Nuclear Gene Capture Statistics

The modified capture protocol for nuclear exons generated 1,264 targeted exons for the 55 chimaeroid lineages and five elasmobranch species. The 1,264 exons were retrieved from a total of 915 different nuclear genes. The data set was 95.05% complete, with 4.95% missing data. The length of individual concatenated sequences ranged from 277,779 bp (*H. cf lemures*) to 353,088 bp (*C. milii*), with a mean of 340,223 bp and a median of 350,112 bp. The total alignment length with gaps was 357,924 bp, with 119,308 amino acids. Exon length ranged from 117 to 4,086 bp, with a mean of 283 bp and median of 222 bp (Figure 2.8).

The clock-like filtered data set of 839 exons had a total alignment length of 206,112 bp. The 1,077 exons, 468 exons, 200 exons, and 100 exons data set alignments contained 290,871 bp, 102,138 bp, 47,691 bp, and 25,218 bp, respectively. The total alignment length for the 50 exon data set was 13,548 bp with 74 taxa. The 50 exons came from 48 different nuclear genes. Exon lengths ranged from 189 to 504 bp, with a mean of 271 bp and a median of 254 bp. Parsimony informative sites for each exon ranged from 94 to 205, with a mean of 109 and a median of 102. The 50 exon alignment contained 5,470 parsimony informative characters.

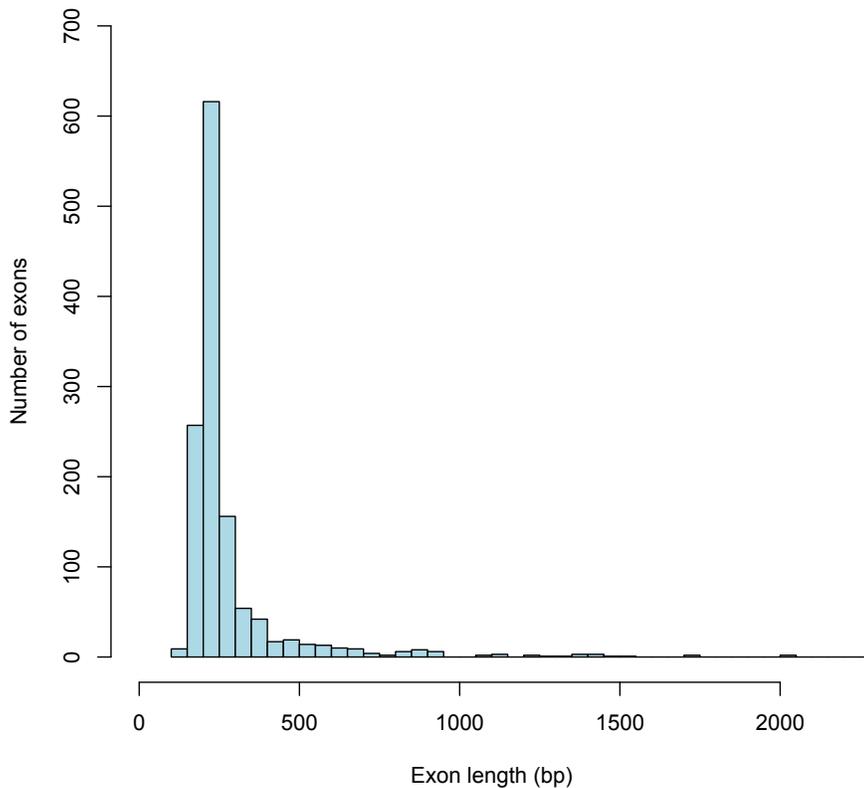


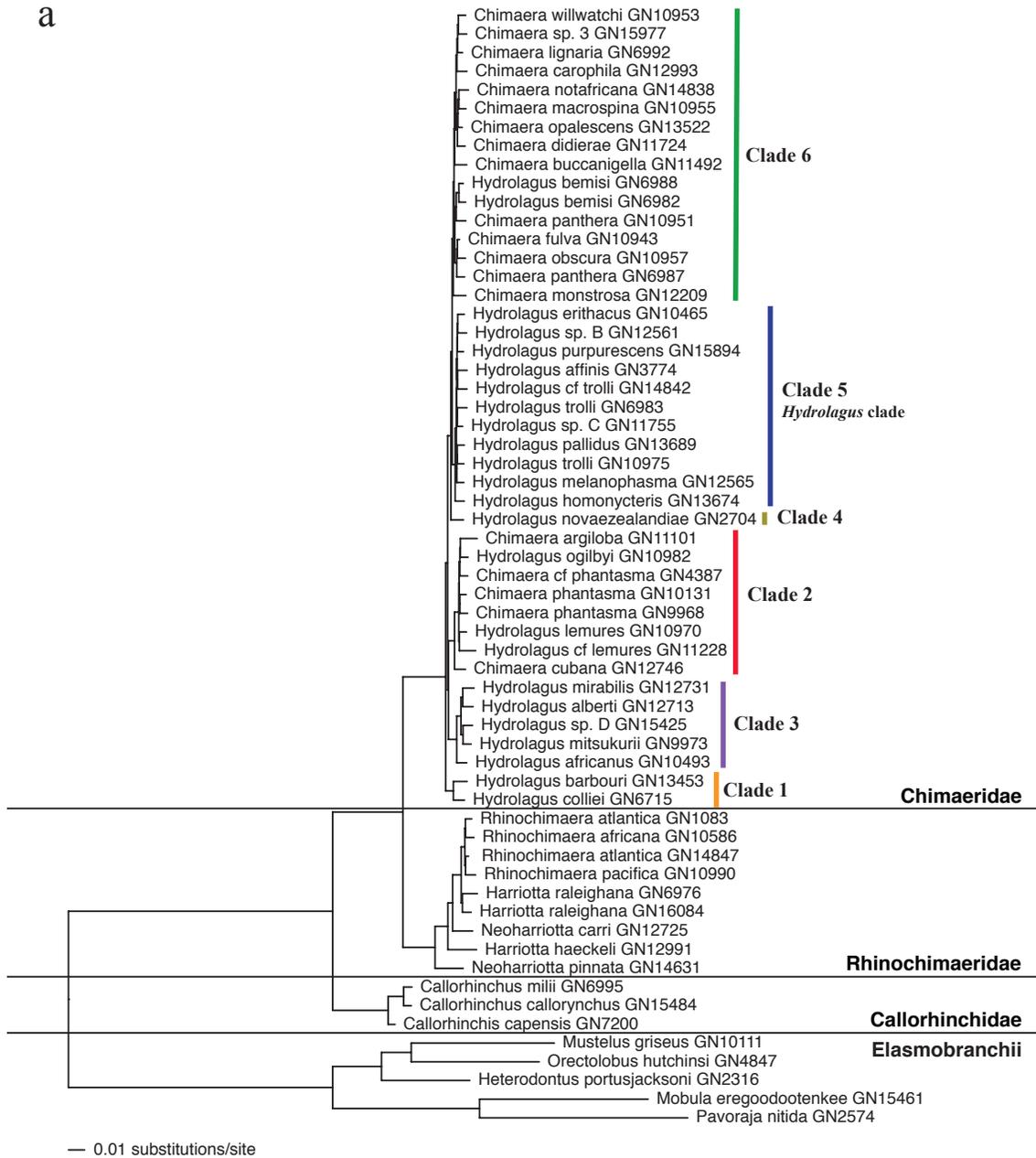
Figure 2.8 Exon length in base pairs (bp) distribution of the 1,264 captured nuclear exons.

Maximum Likelihood Analyses

Resulting maximum likelihood tree topologies across data sets and partitioning schemes were congruent in that they recovered the same eight major clades. However, shallower internal nodes, species-level relationships at the terminal nodes, as well as the placement of two of the major Chimaeridae clades differed among some of the analyses. Each of the three families were recovered as monophyletic with 100% bootstrap support (Figures 2.9-2.18), consistent with the results of previous work (Didier, 1995; Inoue et al., 2010; Licht et al., 2012). The nucleotide full nuclear exon data set (Figure 2.9) and clock-like filtered data set (Figure 2.10) were the most highly supported phylogenies according to the analyses based on different data sets, and their resulting phylogenetic relationships will be discussed below in detail.

The family Callorhinchidae was recovered as the basal clade, highly diverged from Rhinochimaeridae and Chimaeridae. Callorhinchidae has been considered to be the basal group of extant holocephalans based on both morphological and molecular studies (Didier, 1995; Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012). The family Rhinochimaeridae, which contains three recognized genera, was recovered as a monophyletic group with 100% support, sister to Chimaeridae, to the exclusion of Callorhinchidae. Previous studies have concluded that Chimaeridae and Rhinochimaeridae form a monophyletic group (Didier, 1995; Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012). Previous works were characterized by insufficient taxon sampling to elucidate the relationships among *Harriotta* and *Neoharriotta* species. This is the first study to include a majority of the diversity within Rhinochimaeridae. The genus *Rhinochimaera* was recovered as monophyletic, however, *Harriotta* and *Neoharriotta* were paraphyletic. The family Chimaeridae was recovered as a monophyletic group with 100% support in all analyses, with six distinct clades (Figures 2.9-2.18). This is also the first study to

a



b

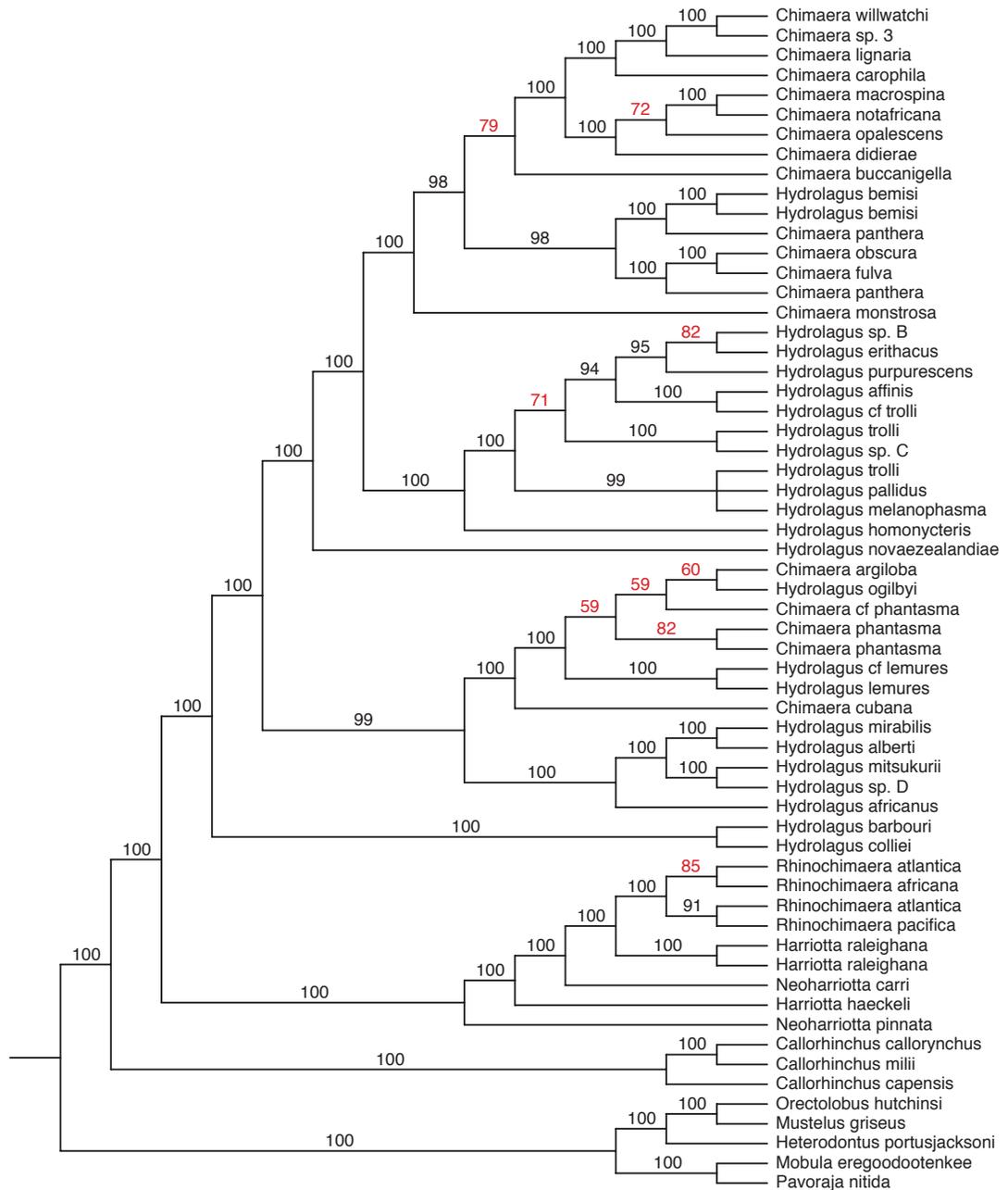
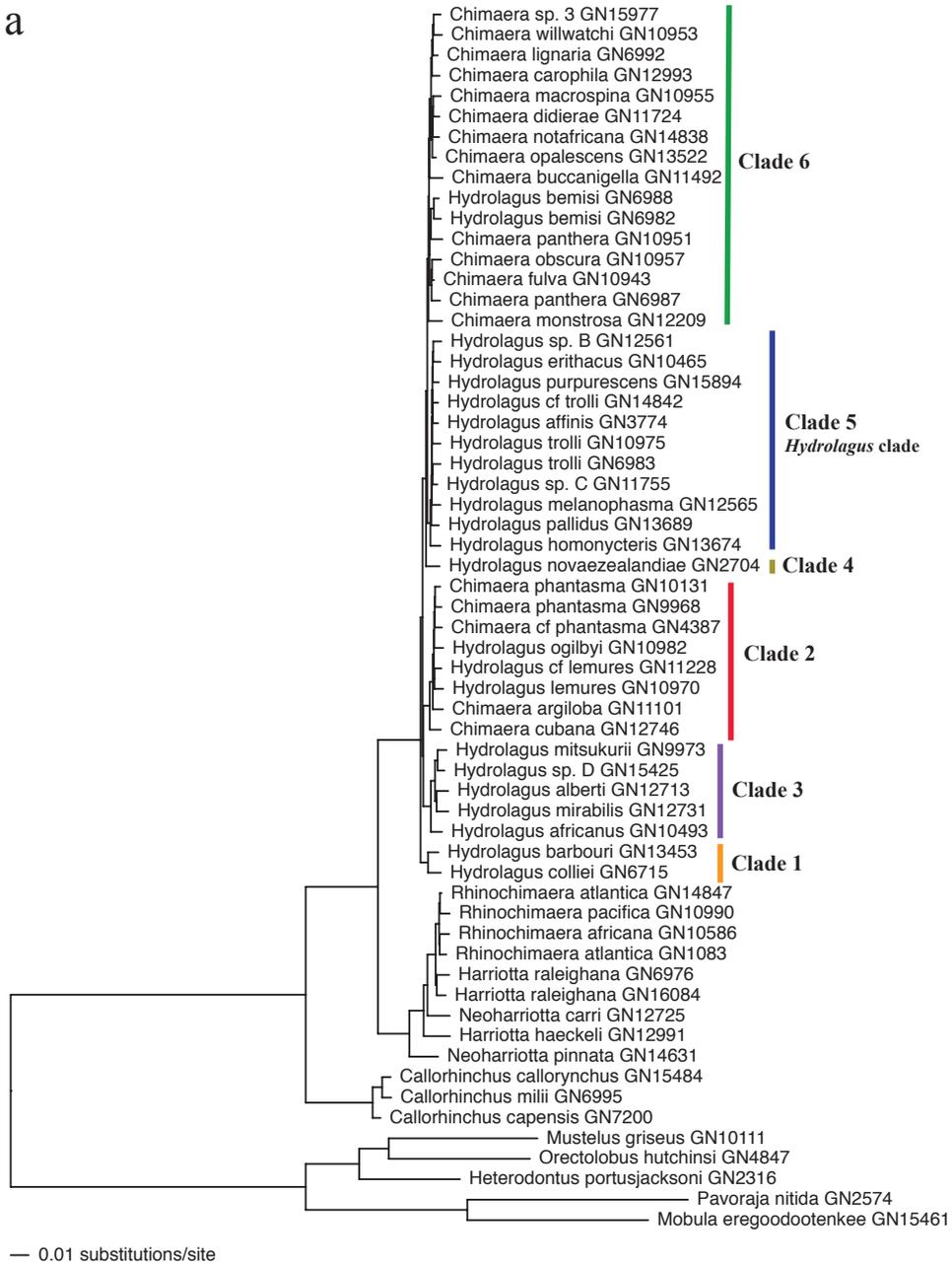


Figure 2.9 The a) maximum likelihood tree topology of the full nuclear nucleotide data set with 1264 exons partitioned by codon under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



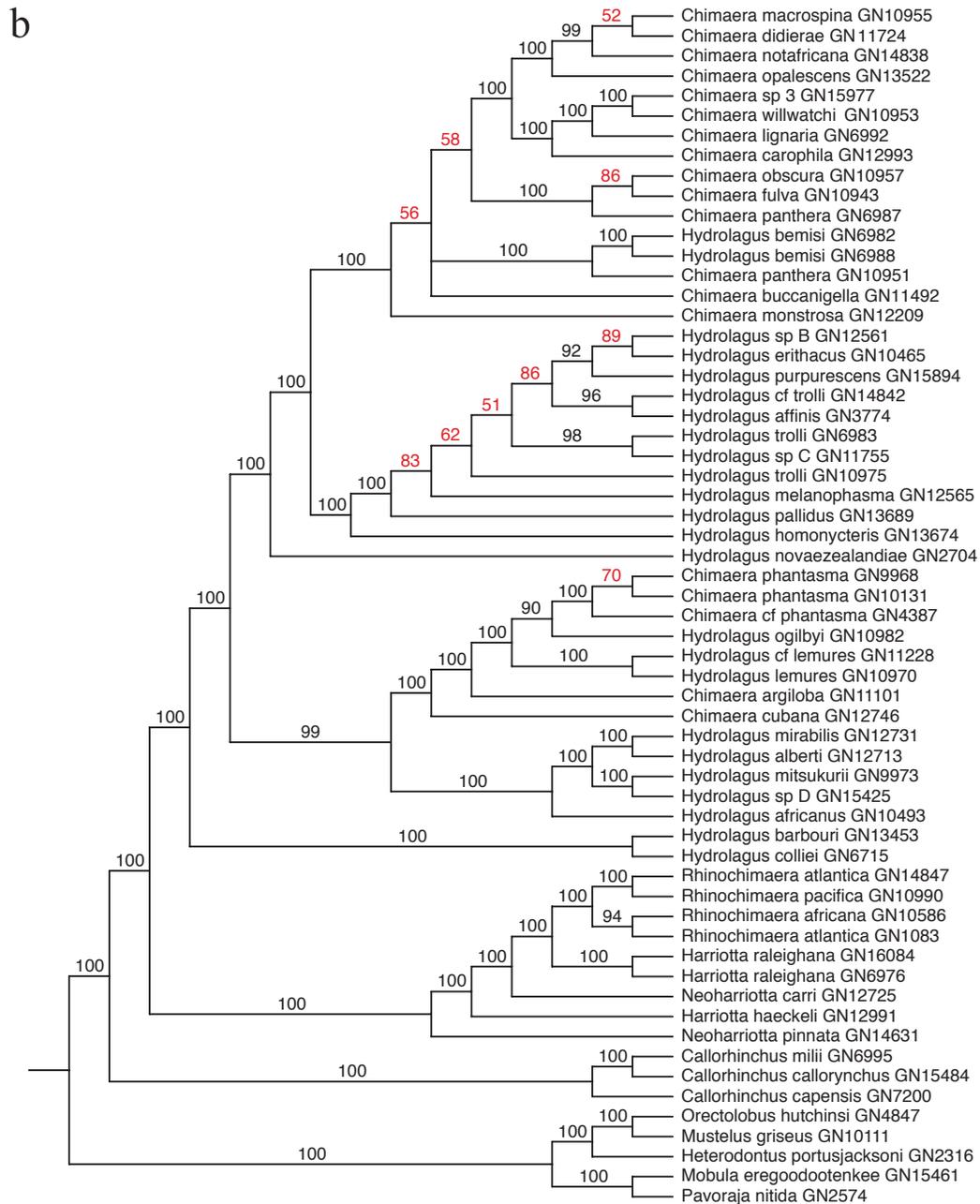


Figure 2.10 The a) maximum likelihood tree topology of the clock-filtered nuclear nucleotide data set under the GTR+ Γ +I model with codon partitioning and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

include dense taxon sampling of Chimaeridae species. While some structure can be seen, overall, the genera *Chimaera* and *Hydrolagus* were not monophyletic. It has been speculated that the anal fin character separating these two genera may not be a reliable means of separation. Previous molecular work has recovered the genera to be both monophyletic (Inoue et al., 2010) and paraphyletic (Ward et al., 2008; de la Cruz-Aguero et al., 2012; Licht et al., 2012), but lacked adequate taxon sampling.

The relationships between the three *Callorhynchus* species were consistent across all topologies, with *C. milii* sister to *C. callorhynchus*, to the exclusion of *C. capensis* (Figures 2.9-2.18). This relationship was consistently resolved with high support in all analyses except the 50 exon amino acid data set (Figure 2.18). These three recognized species are very closely related as indicated by the very short terminal branch lengths. This pattern was also found among the three species in previous molecular studies (Inoue et al., 2010; Licht et al., 2012). This is a very interesting finding given that these three species inhabit coastal waters of three different continents, *C. milii* from New Zealand and Australia, *C. capensis* from South Africa, and *C. callorhynchus* from South America. Previous works did not agree on the relationships among these three species (Inoue et al., 2010; Licht et al., 2012). Morphologically, these three species are nearly indistinguishable, currently separated based on geographic location. All species are characterized by a silvery color with dark spots or blotches along trunk. Color pattern has been suggested as one means of distinguishing species, but can be highly variable, and thus, not a reliable character for species separation. The estimated short terminal branch lengths and little genetic variation, along with no consistent morphological differentiation, one could hypothesize that these three species may actually represent one species with population structure. On the

other hand, the data may not contain enough phylogenetic signal to resolve the relationships, especially if the speciation events are close together (Saitou & Nei, 1986; Philippe et al., 1994).

Within the family Rhinochimaeridae, species-level relationships within the clade differed based on data sets. *Neoharriotta pinnata* was the basal lineage in all topologies, with high support, divergent from the remaining lineages. *Harriotta haeckeli* and *Neoharriotta carri* lineages also showed high support as unique lineages, divergent from others. *Harriotta raleighana* was represented by two distinct lineages, one from New Zealand, the other from off Scotland, sister taxa to one another. The implications of this result are difficult to interpret, as this could represent population variation or potentially unique species. The two *H. raleighana* lineages fall out as sister to the genus *Rhinochimaera*, not more closely related to the other member of *Harriotta*. Licht et al. (2012) also found that *H. raleighana* and *R. pacifica* clustered together to the exclusion of *N. pinnata*. The species-level relationships within *Rhinochimaera* differed among analyses, and generally showed lower bootstrap support. *Rhinochimaera atlantica* (GN1083; Atlantic Ocean off USA) clustered with *R. africana* (Japan), and *R. atlantica* (GN14847; South Africa) clustered with *R. pacifica* (Australia) in all analyses, except for full nuclear amino acid (Figure 2.11) and 50 exon nucleotide gene+codon partition (Figure 2.17). In no topologies did the two lineages identified as *R. atlantica* fall out as most closely related to one another. A morphological investigation of these lineages is needed to further explore this finding. Very short branches among the four *Rhinochimaera* lineages were seen, indicating little genetic variation among these lineages. This makes interpreting the relationships and species delimitation within this genus difficult. On one extreme, there may not be enough information in the data to fully resolve the relationships among members of the genus, or on the other end population variation may have been recovered for one wide-ranging species of *Rhinochimaera*.

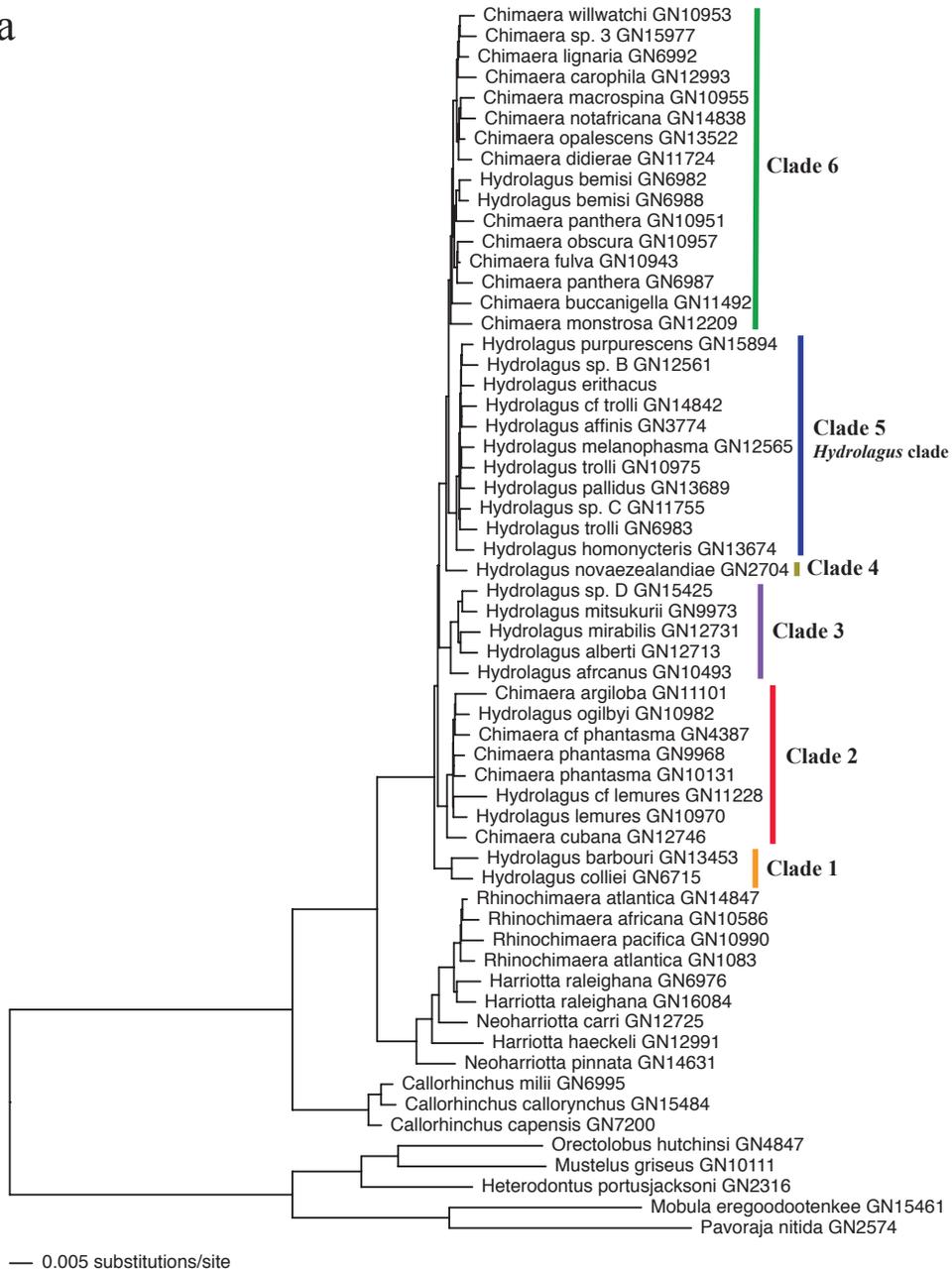
Six distinct clades within Chimaeridae were recovered across all data sets. The most basal Chimaeridae clade, which had 100% bootstrap support, comprised of two species, *H. colliei* and *H. barbouri*, consistent across all data sets (clade 1). Clade 2 (*C. cubana*, *H. cf lemures*, *H. lemures*, *C. phantasma* (GN9968), *C. phantasma* (GN10131), *C. cf phantasma*, *H. ogilbyi*, *C. argiloba*) contained both *Chimaera* and *Hydrolagus* species, indicating that these two genera are not monophyletic. Clade 3 (*H. africanus*, *H. mitsukurii*, *H. sp. D*, *H. alberti*, *H. mirabilis*) consisted of only *Hydrolagus* species. Both of these clades showed high support, and were sister clades to one another in the full (Figure 2.9), clock-filtered (Figure 2.10), and 1,077 exons (Figure 2.13) nucleotide data sets. The relationships within clade 2 differed between the full and clock-filtered nucleotide topologies. Overall, *C. cubana* was always the basal species, and *H. lemures*, and *H. cf. lemures* were sister taxa. Clade 3 species relationships were identical between the two data sets. Short internal and terminal branch lengths were estimated for relationships within clades 2 and 3. Again, this complicates interpretation of species boundaries within these clades. The next major Chimaeridae lineage showed 100% support in all resulting topologies, and contained one species, *H. novaezealandiae* (clade 4). It is placed as sister to the remaining two major groups, clade 5 (*H. homonycteris*, *H. melanophasma*, *H. trolli*-GN10975, *H. pallidus*, *H. sp. C*, *H. trolli*-GN6983, *H. cf trolli*, *H. affinis*, *H. purpurescens*, *H. sp. B*, *H. erithacus*) and clade 6 (*C. monstrosa*, *C. panthera*-GN6987, *C. obscura*, *C. fulva*, *C. panthera*-GN10951, *H. bemisi*-GN6982, *H. bemisi*-GN6988, *C. buccanigella*, *C. didierae*, *C. opalescens*, *C. macrospina*, *C. notafriicana*, *C. carophila*, *C. lignaria*, *C. sp 3*, *C. willwatchi*; Figures 2.9-2.17). Only in the 50 exon amino acid topology was clade 4 placed as sister taxon to clade 5, but this relationship had poor support (Figure 2.18). Clade 5 was recovered as sister to the clade 6 in all resulting topologies (Figures 2.9-2.17), with high bootstrap support. Clade 5 consisted of 11 identified *Hydrolagus* lineages. Some of the relationships within this group generally showed lower

support, and different species-level relationships resulted from analyses of the various data sets. Again, the very short branch lengths estimated for these lineages indicated little genetic variation within this group of holocephalans, leaving the interpretation of species-level relationships unclear. Another potential explanation is that the lineages within this clade may represent one species, which will be discussed in more detail in Chapter 3.

Within the full nuclear exon data set, the nucleotide analyses with no partition and a codon partitioning scheme recovered identical topologies (Figure 2.9; only codon partition tree shown). There was 100% bootstrap support for all major clades, and high support for many of the shallower nodes. The amino acid analysis resulted in the same major clades, which contained the same set of species, but differed in placement of Chimaeridae clade 2 and 3, and species-level relationships within the clades (Figure 2.11). While the major clades were well supported in these analyses, there was generally less support for shallower nodes in the amino acid analysis compared to nucleotide.

Identical tree topologies were also recovered for nucleotide analyses of the clock-like filtered data set under no partition and codon partitioning (Figure 2.10; only codon partition tree shown). The amino acid analysis differed from the nucleotide topology in some of the species-level relationships and the placement of clades 2 and 3 (Figure 2.12). However, members of the clades remained consistent between the analyses. The support for nodes in the amino acid analysis were high for the majority of the major clades, however, more polytomies were present in the consensus tree, representing less support for many of the species-level relationships. The full and clock-filtered data sets produced similar results with respect to both nucleotide and amino acid topologies, showed generally high bootstrap support, and were only incongruent with some of the relationships towards the tips of the trees.

a



b

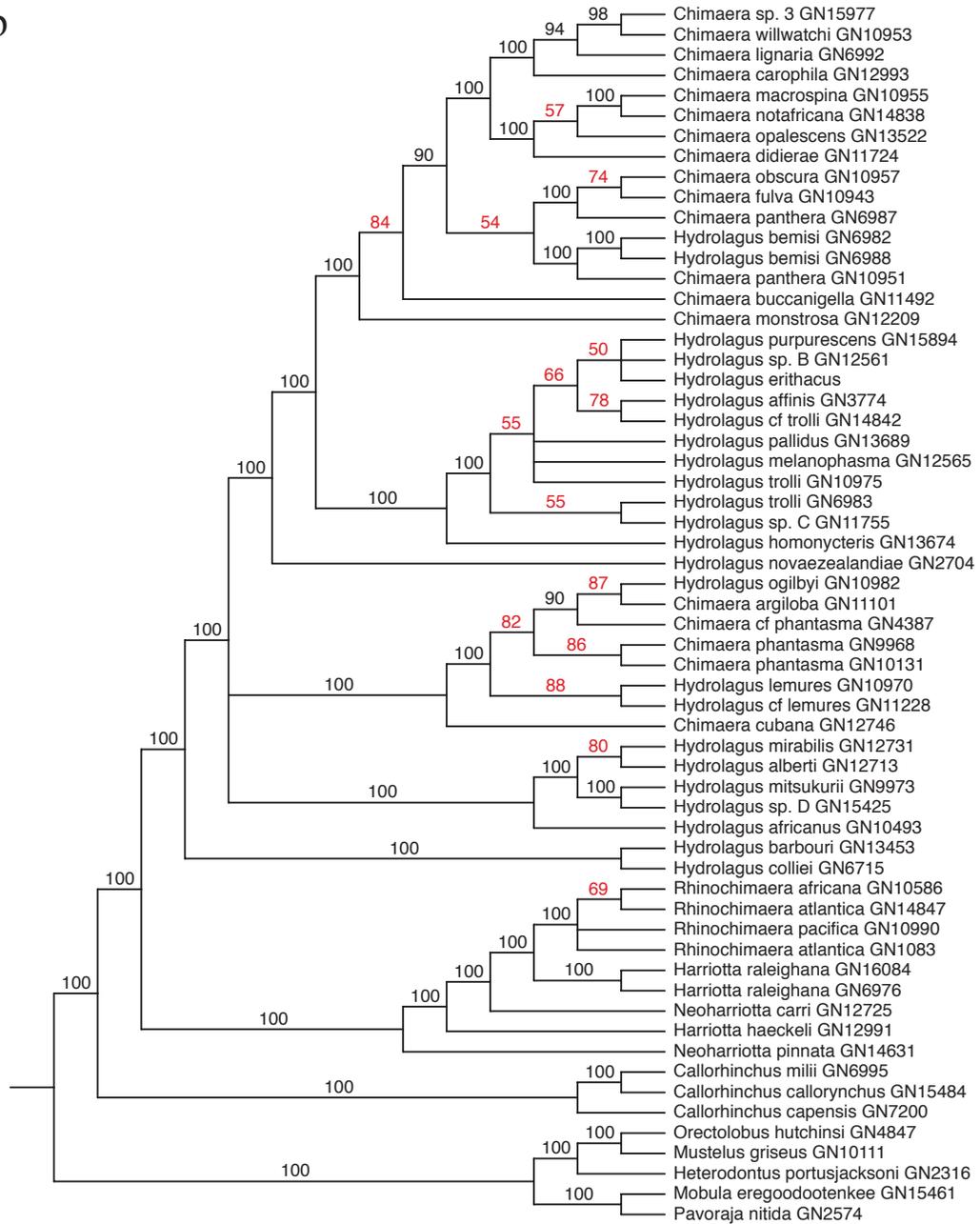


Figure 2.11 The a) maximum likelihood tree topology of the full nuclear amino acid data set with 1264 exons under the JTT+ Γ +I+F model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

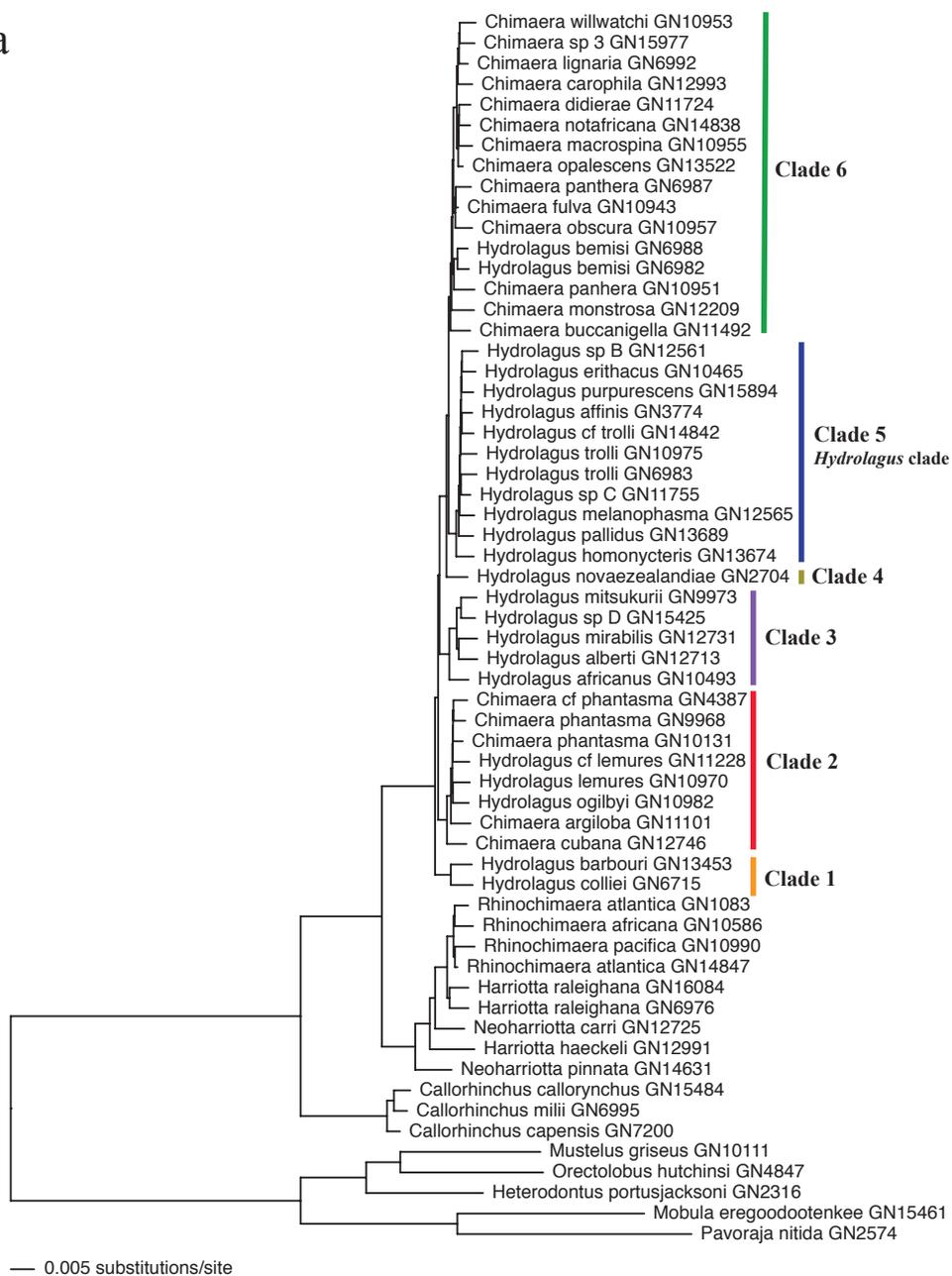
The 1,077 exon data set produced a similar topology to the previous nucleotide data sets (Figure 2.13). Differences were noted in the placement of clades 2 and 3, as well as relationships at the tips of the trees. The nucleotide alignment for the 50 exon data set under the different partitioning schemes resulted in mainly congruent tree topologies (Figures 2.14-2.17). Overall, for the 50 exon data set, the major clades were well supported, however, a few of the deeper node splits among major clades (i.e. split between clades 2 and 4-6, and between clades 5 and 6) and many of the species-level relationships were not well supported. In the 50 exon nucleotide topologies (Figures 2.14-2.17), clade 2 was recovered as a sister group to the group containing clades 4, 5 and 6, which was inconsistent with all other data sets and their corresponding analyses. The eight clades and their respective species remained the same among 50 exon analyses; it was the species-level relationships within one of the Chimaeridae clades (clade 5), and the relationships among the *Rhinochimaera* that differed. The amino acid analyses were identical between no partitioning scheme and a gene partitioning scheme for the 50 exon data set (Figure 2.18; only no partition tree topology shown). While some of the relationships were highly supported, the majority of relationships showed poor bootstrap support. The amino acid topology (Figure 2.18) showed the same major relationships as the full and clock-like filtered amino acid results (Figures 2.11-2.12). When comparing the nucleotide tree topologies to the amino acid topologies for all data sets, the major clades and the species contained within them remained the same. The relationships among the *Rhinochimaera*, the placement of two of the major clades within Chimaeridae (clades 2 and 3), and species-level relationships within the clades of Chimaeridae differed between the two character types. Partitioning did not have any major effects on the analysis results for any of the data sets.

In general, reliability of resulting tree topologies was greatest in those data sets that contained more characters (i.e. full and clock-like filtered nucleotide data sets). These data set

alignments were significantly longer than the 50 exon data set, and thus, likely contained many more phylogenetically informative sites that could be used for the phylogenetic inference. The 50 exon data set showed much less support for relationships, but is likely attributable to a decreased alignment length, and reduced information in the data. Amino acid analyses of the various data sets showed even less support than nucleotide analyses. The amino acid alignments contained fewer characters compared to their nucleotide counterpart, and thus, fewer informative sites. Since the same amino acid can be formed by multiple codon triplets, it is expected that in an alignment, there would be more variable nucleotide characters than amino acids. Thus, there would be less change in amino acid sequences. In this case, it appears that there is less phylogenetic signal in amino acid data sets, resulting in tree topologies with low support and more polytomies, compared to nucleotide data sets.

All nuclear exon data sets with ML analyses recovered the same eight major clades and species within these clades, with high support. In many cases, the relationships within the clades showed lower support depending on data set, and there was general topological incongruence at this level. Thus, only those relationships that were consistently recovered across data sets and inferences can be confidently resolved. Further support by utilizing a different method (e.g. Bayesian Inference, see below) and another independent data set (e.g. mitochondrial markers, see Chapter 3), will be needed to further support the estimated relationships among sampled lineages.

a



b

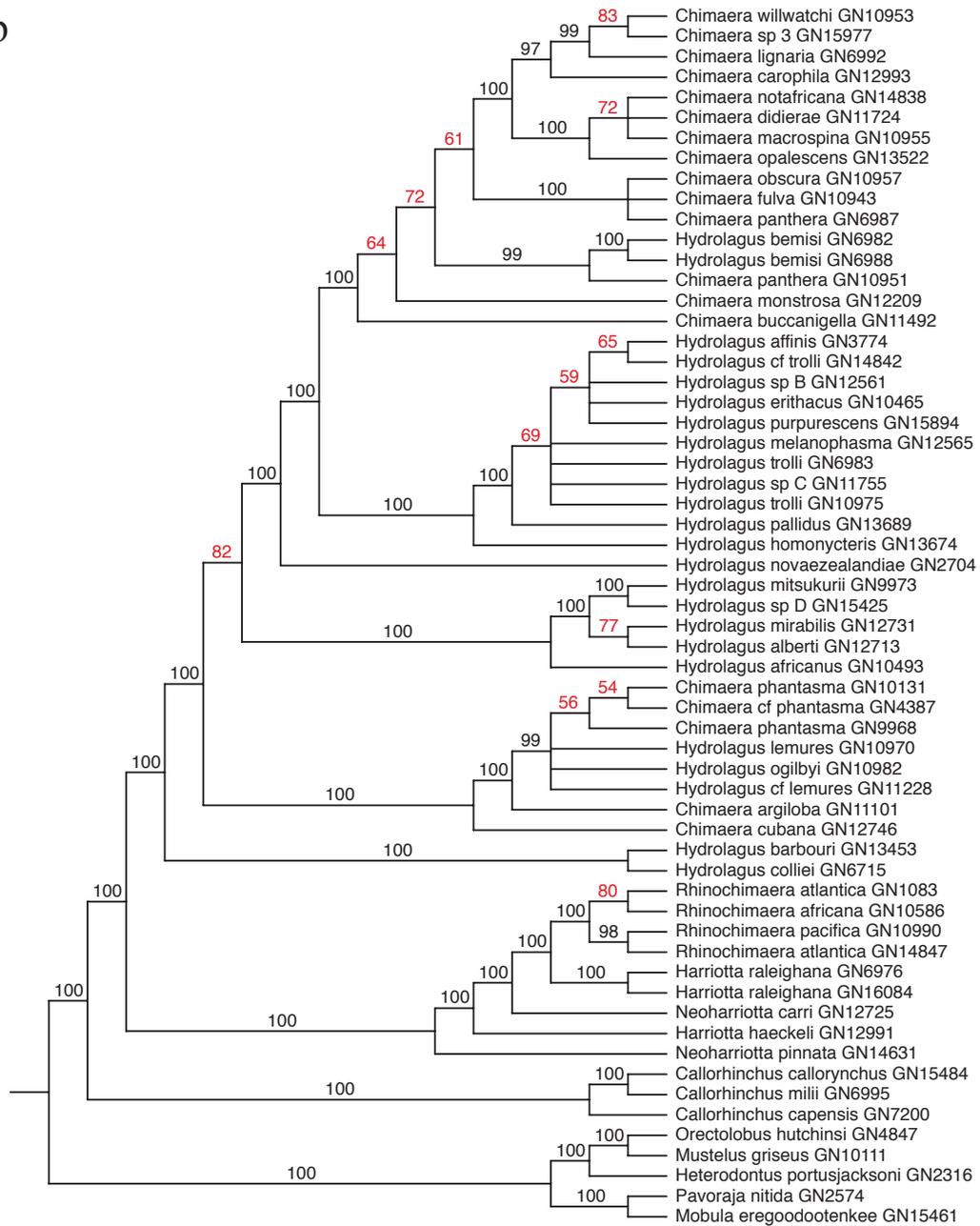
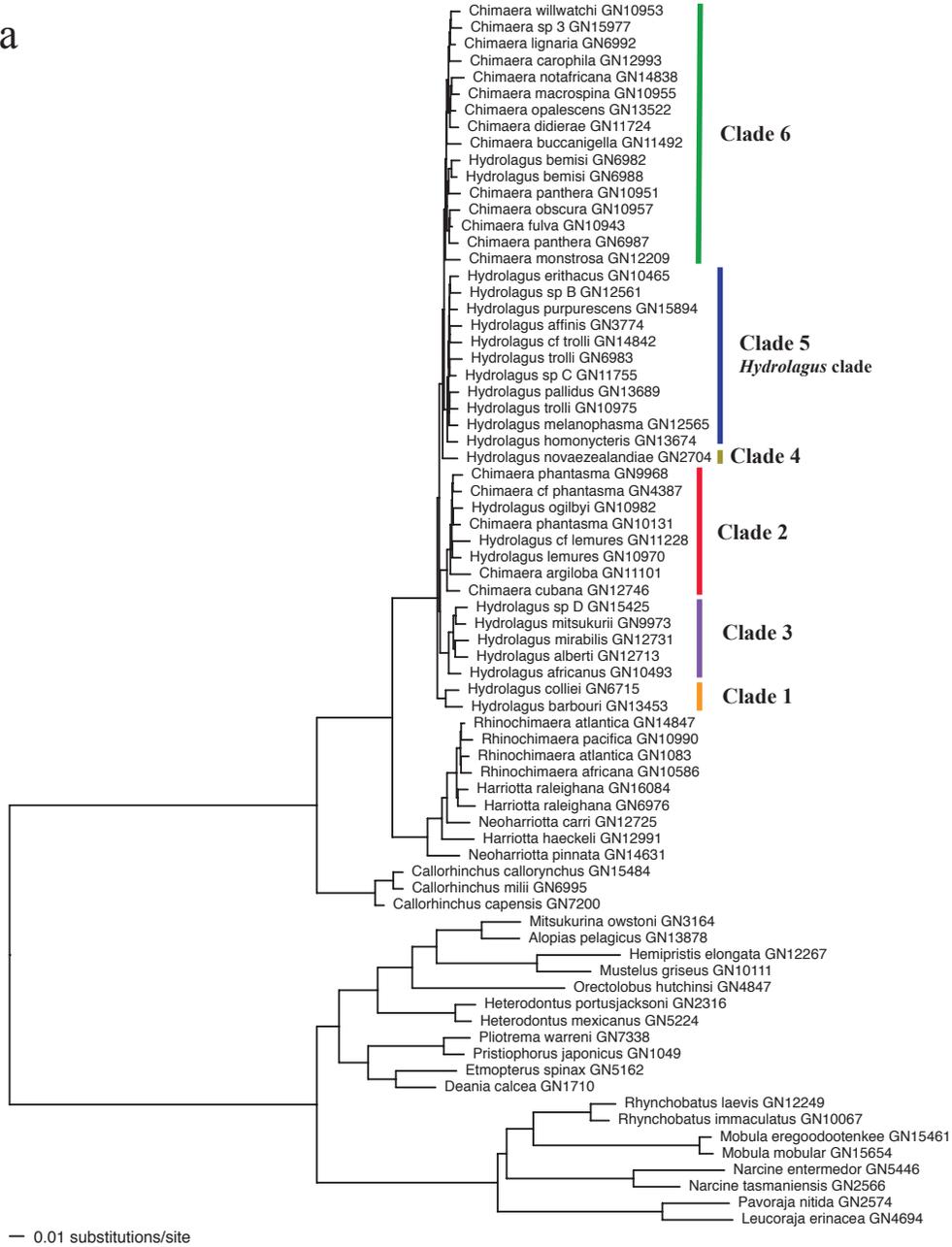


Figure 2.12 The a) maximum likelihood tree topology of the clock-filtered nuclear amino acid data set under the JTT+ Γ +I+F model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



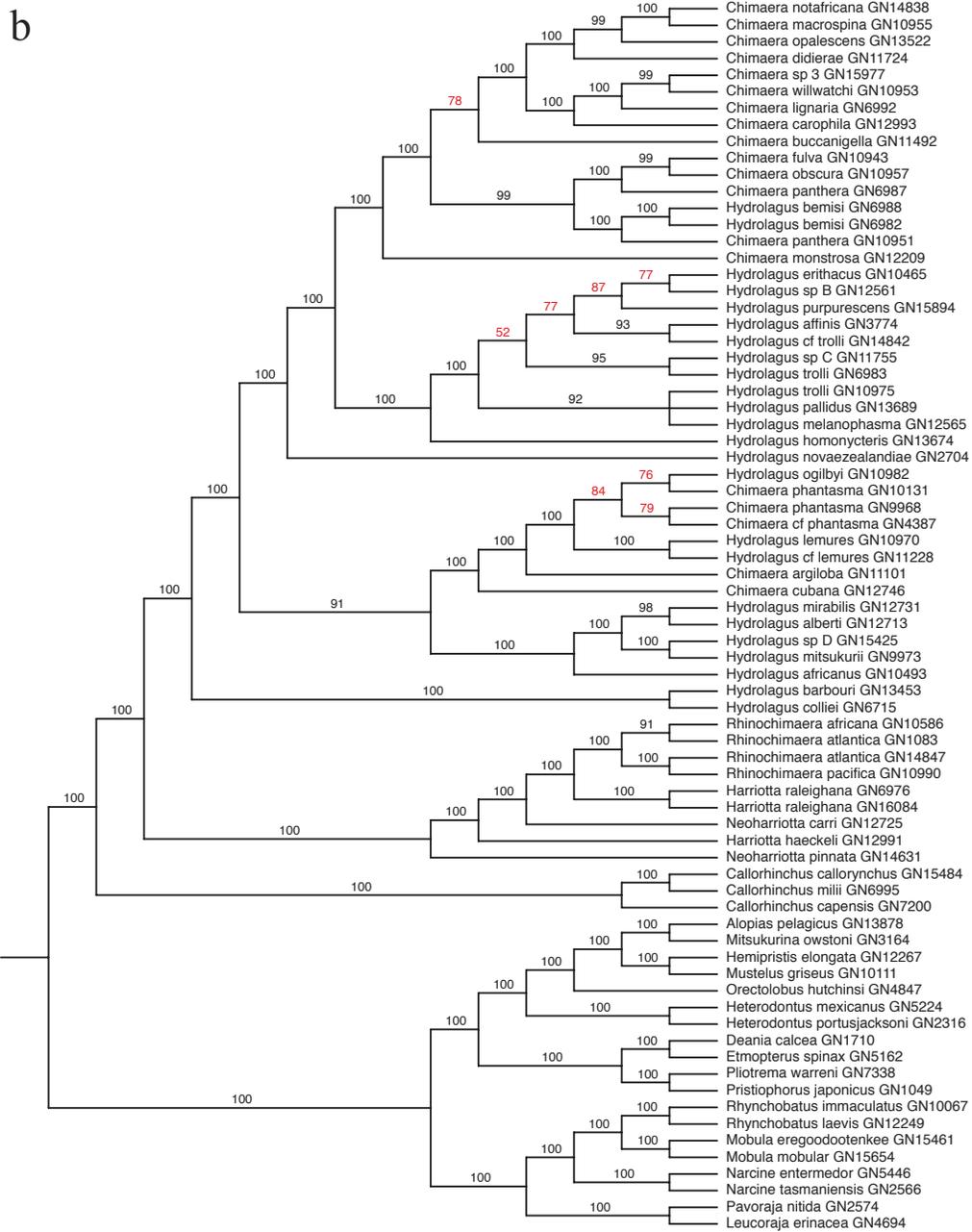
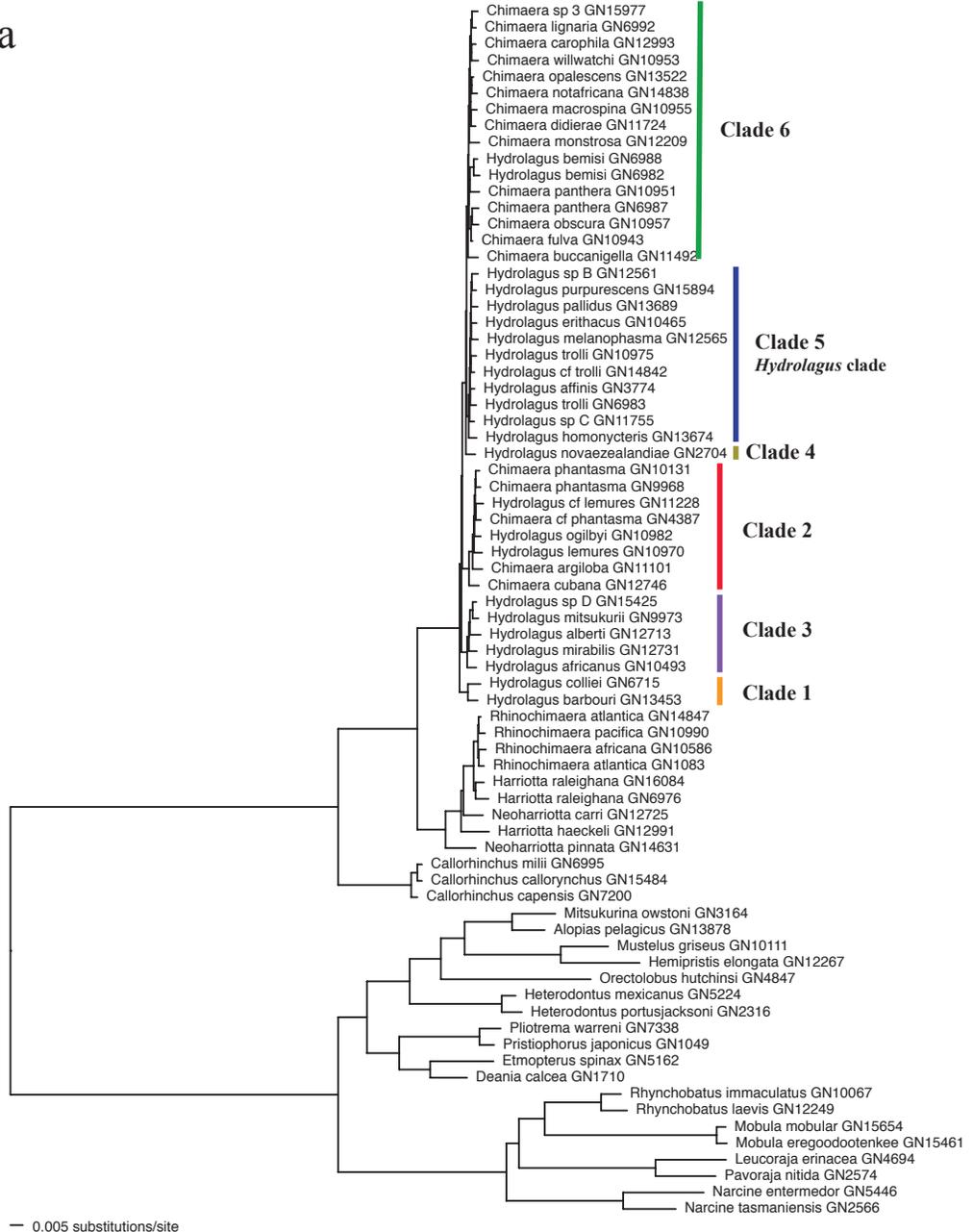


Figure 2.13 The a) maximum likelihood tree topology of the 1077 exon nuclear nucleotide data set under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



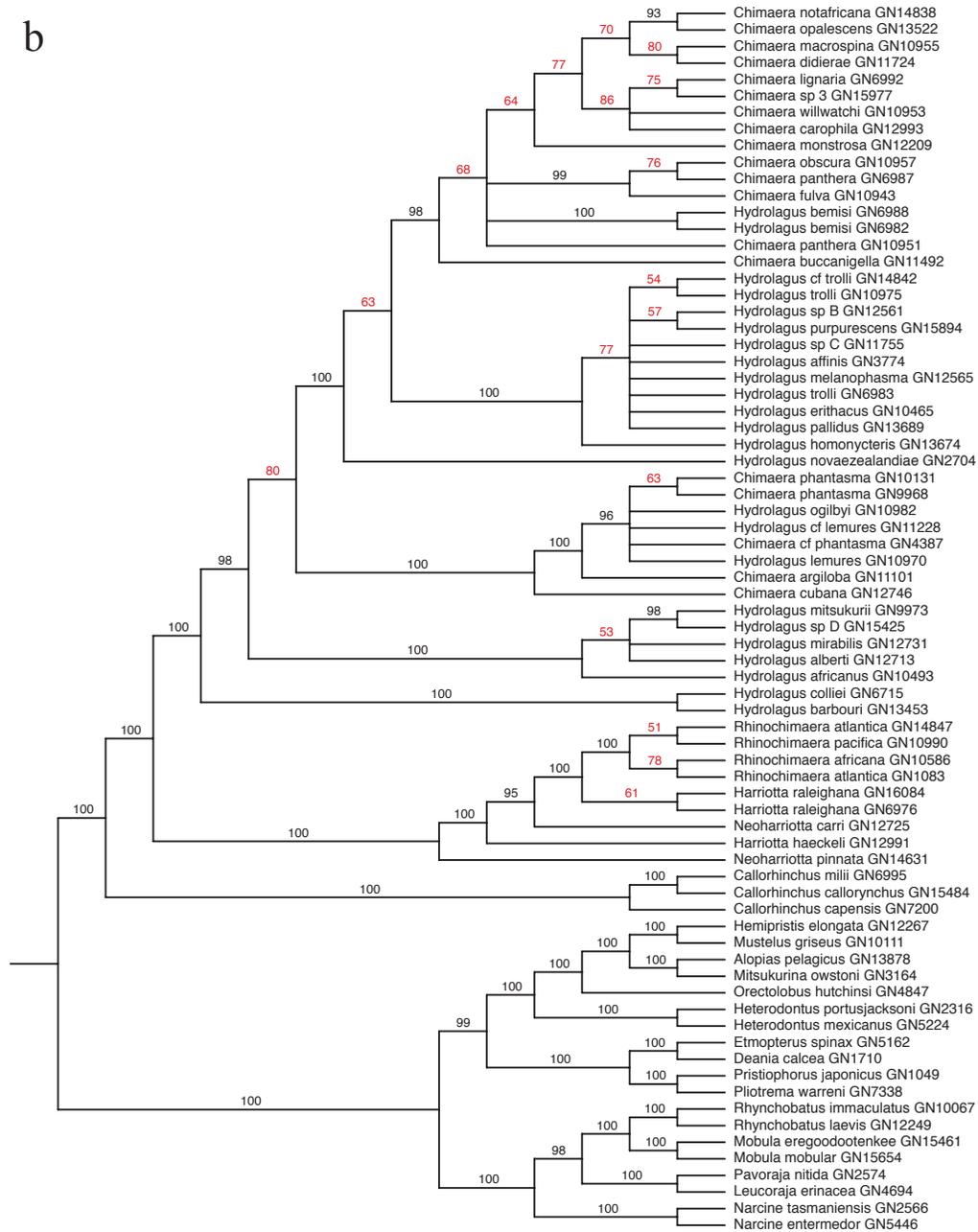
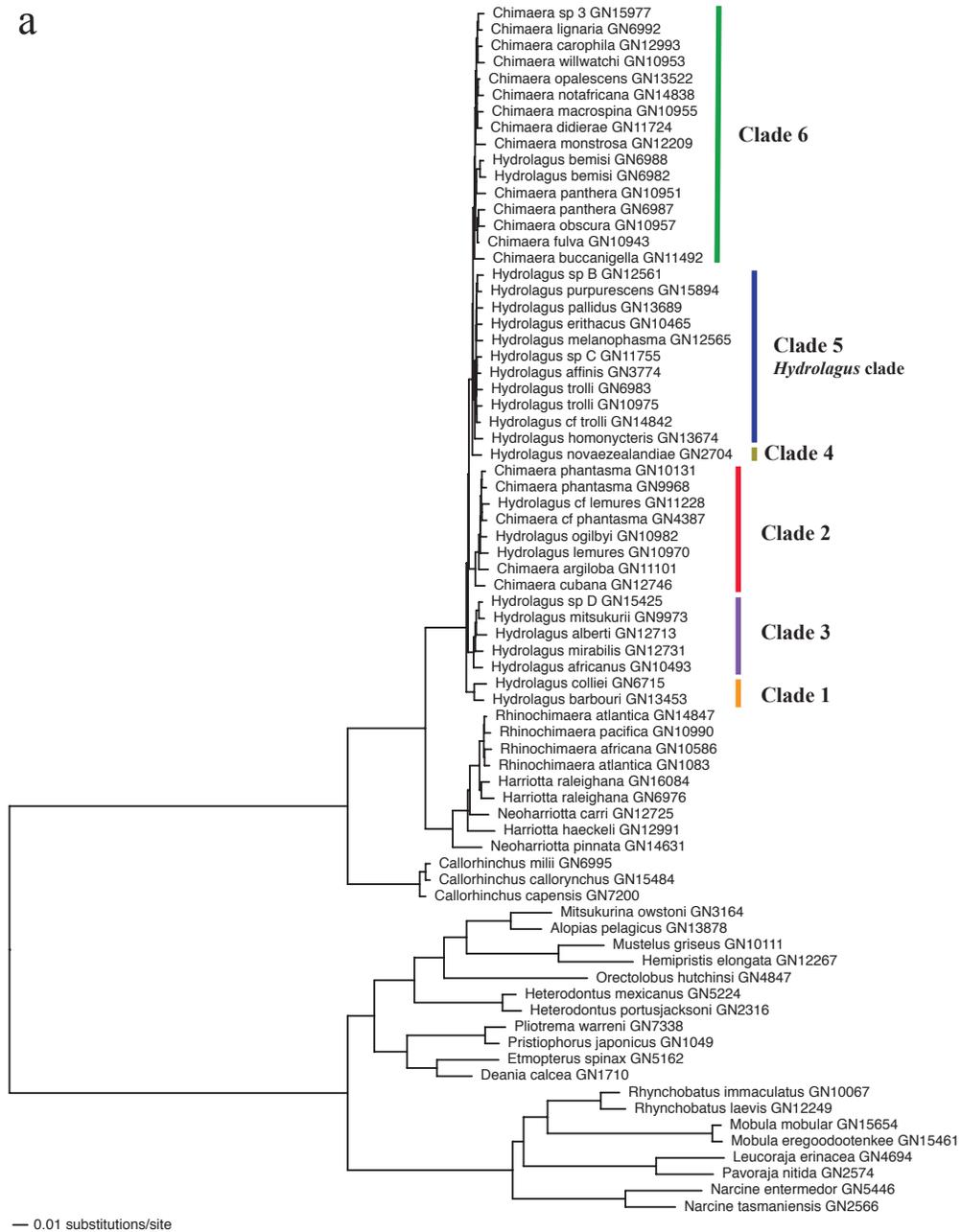


Figure 2.14 The a) maximum likelihood tree topology of the 50 exon nuclear nucleotide data set with no partitioning scheme under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



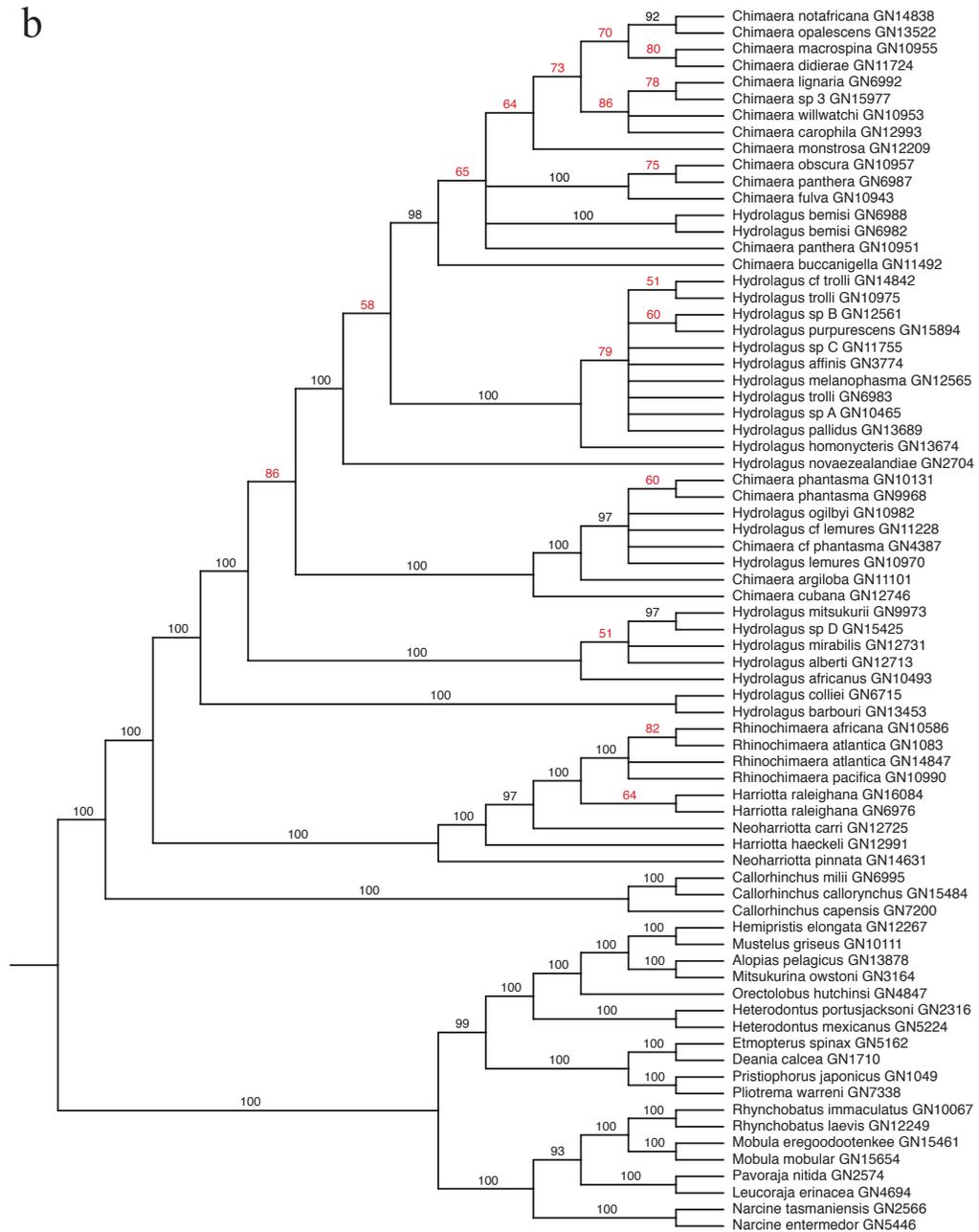
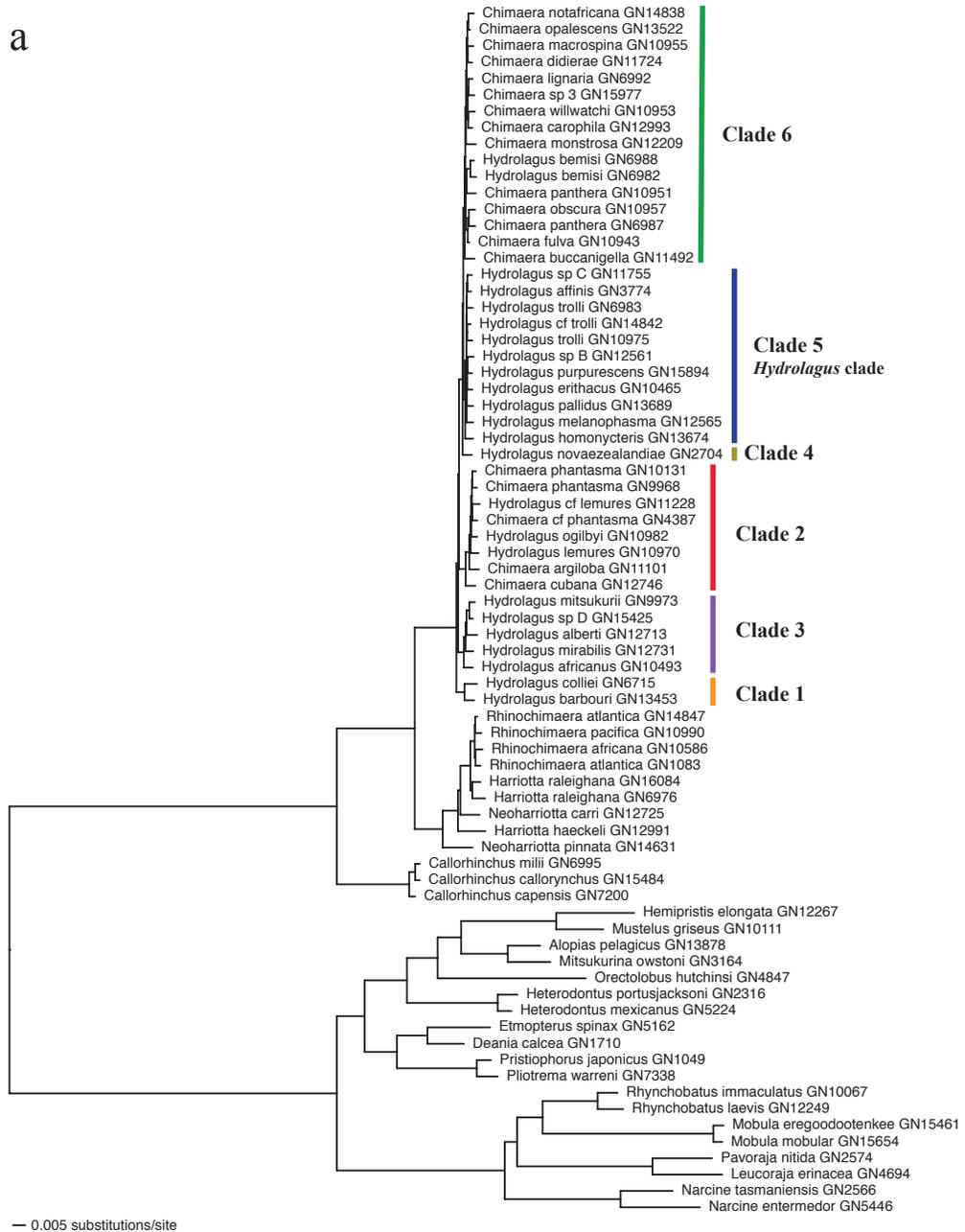


Figure 2.15 The a) maximum likelihood tree topology of the 50 exon nuclear nucleotide data set partitioned by codon under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



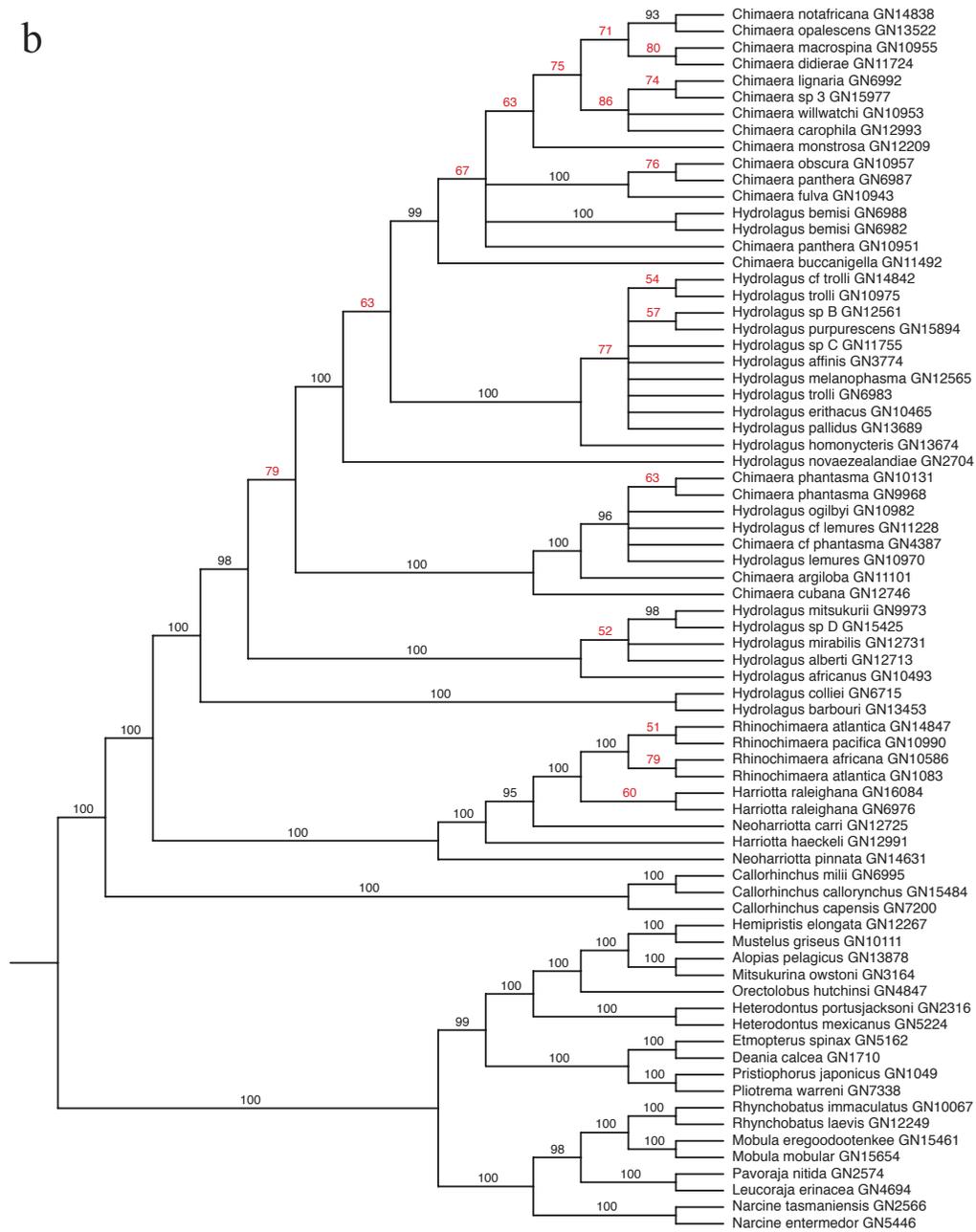
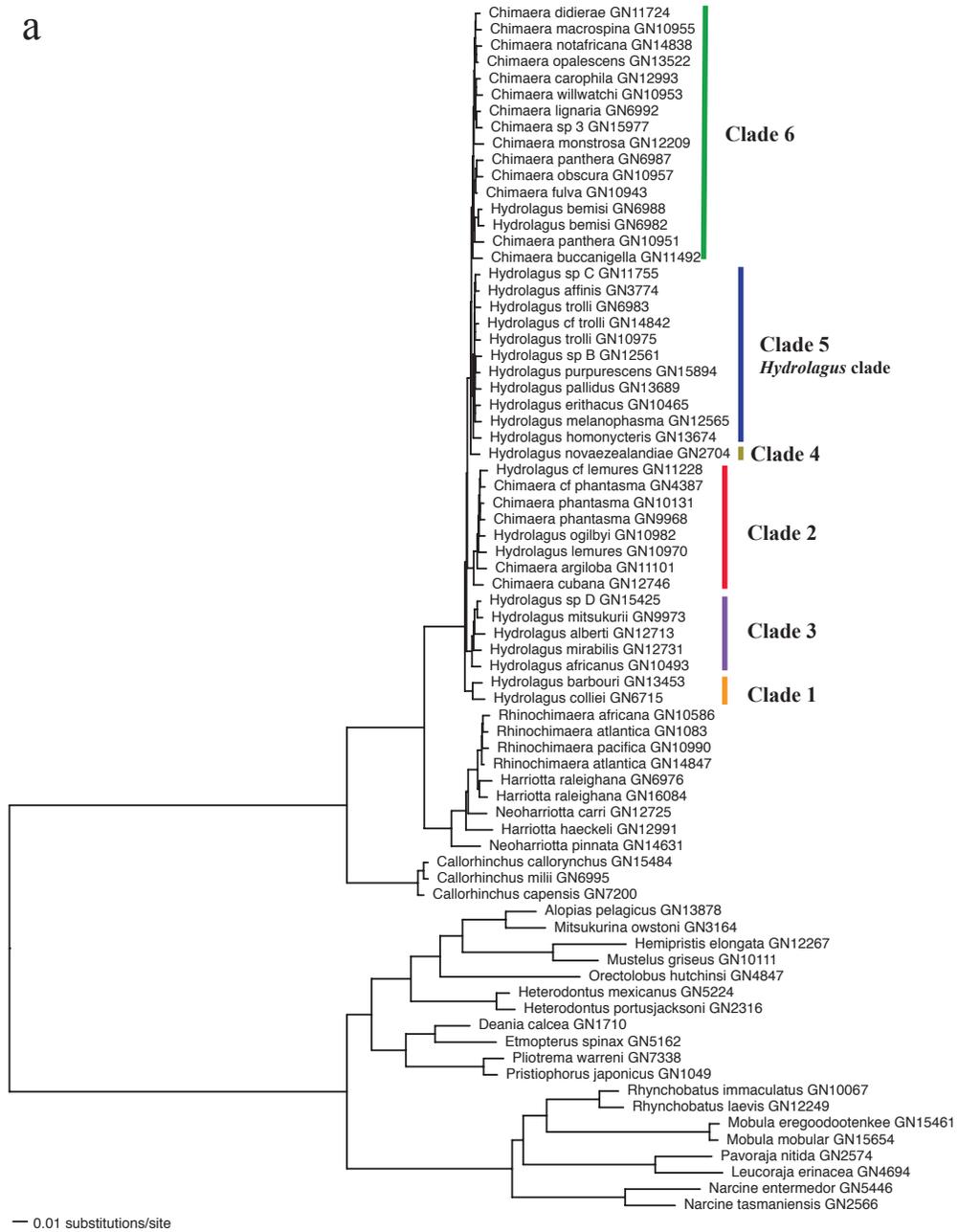


Figure 2.16 The a) maximum likelihood tree topology of the 50 exon nuclear nucleotide data set partitioned by gene under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



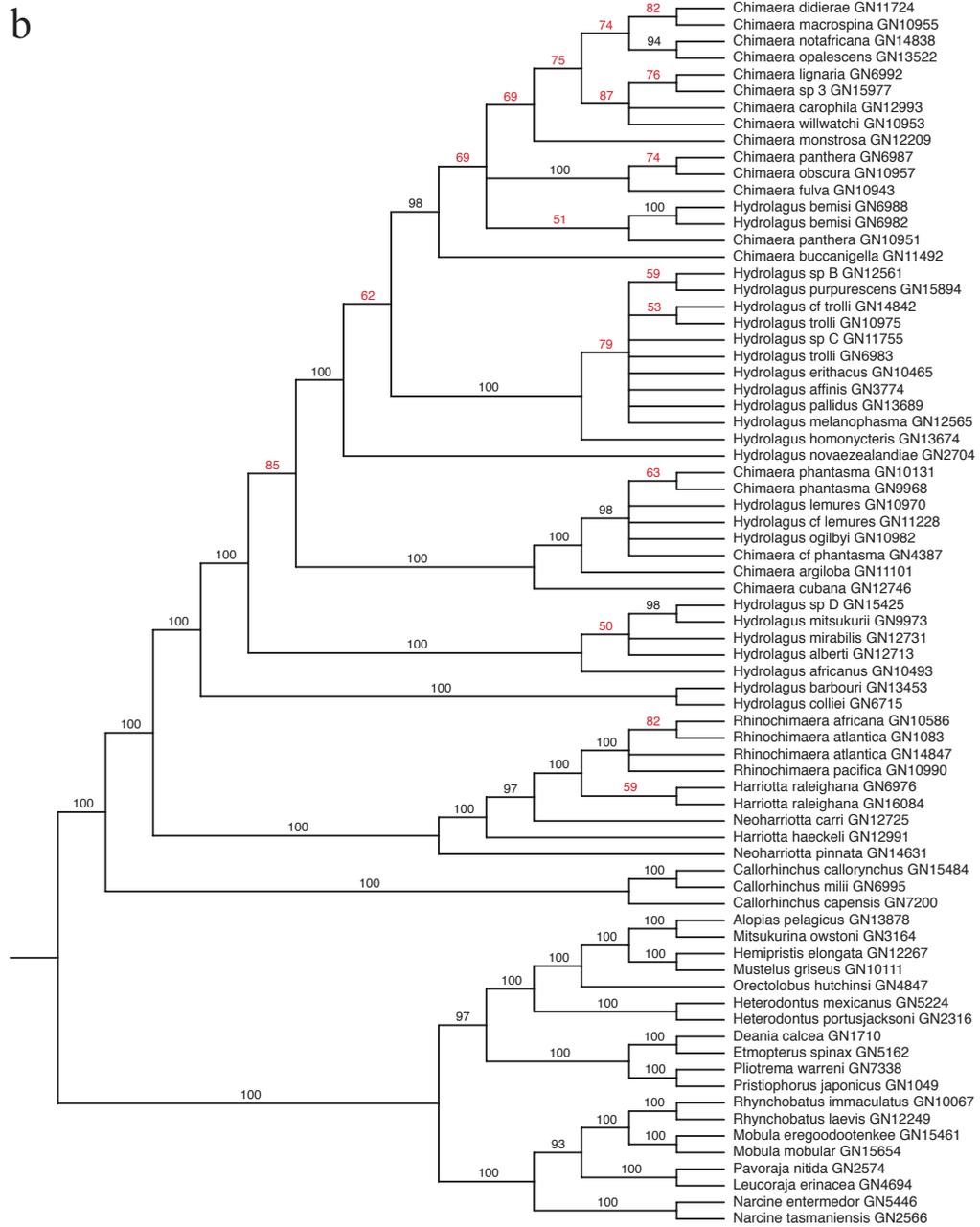
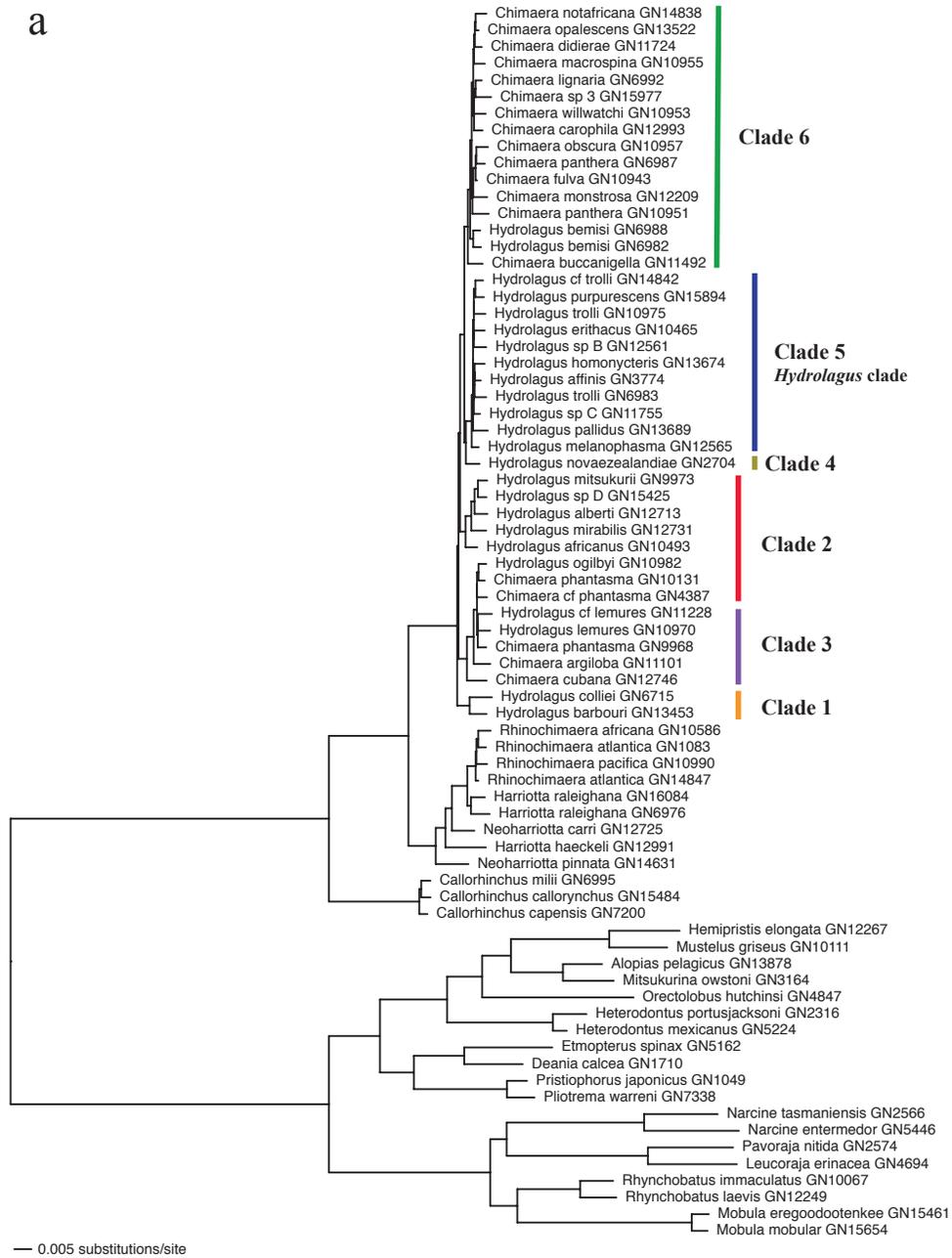


Figure 2.17 The a) maximum likelihood tree topology of the 50 exon nuclear nucleotide data set partitioned by gene+codon under the GTR+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

a



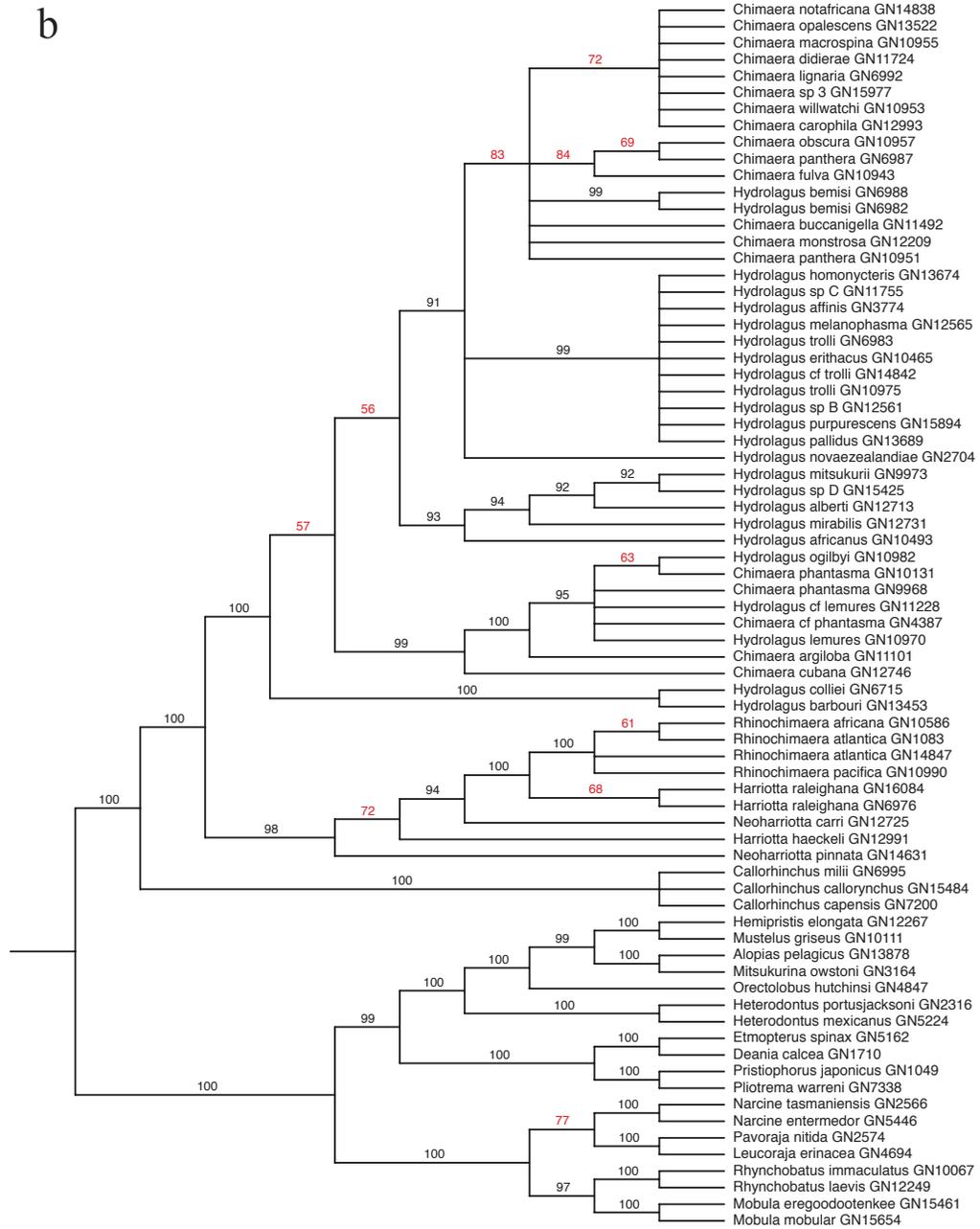


Figure 2.18 The a) maximum likelihood tree topology of the 50 exon nuclear amino acid data set with no partitioning scheme under the JTT+ Γ +I model and b) majority rule consensus tree with bootstrap support for the resulting node relationships (1000 replicates). Nodes with less than 90% bootstrap support are highlighted in red. Polytomies indicate bootstrap support < 50%.

Bayesian Inference Analyses

Tree topologies recovered from the four BI analyses were mainly congruent in that they contained the same eight major clades and species within them (Figures 2.19-2.22). The two major clades, holocephalans and elasmobranchs were each recovered as monophyletic, and highly diverged from one another. The Callorhynchidae clade was well supported with a posterior probability of one in all analyses, and recovered as monophyletic and basal. The species-level relationships were the same among the BI analyses, with *C. callorynchus* and *C. milii* sister to one another, and identical to ML results. The Rhinochimaeridae clade was monophyletic, sister to the Chimaeridae clade, and well supported. Relationships within Rhinochimaeridae were identical among the three analyses with nucleotide character data (Figures 2.19-2.21), but the amino acid data resulted in different relationships within the genus *Rhinochimaera* (Figure 2.22), which showed a lower posterior probability. The Rhinochimaeridae clade relationships of the nucleotide data sets under BI were identical to the ML relationships. The Chimaeridae clade was highly supported and consisted of same six well-supported major clades across data sets (Figures 2.19-2.22).

The major difference within the Chimaeridae clade between the four BI analyses was the placement of clades 2 and 3 (Figures 2.19-2.22). The clock-like filtered data set recovered clades 2 and 3 as sister groups (Figure 2.19). The 50 exon nucleotide data set with no partitioning scheme and a codon partition recovered clade 2 as a sister group to the group containing clades 4, 5, and 6 (Figures 2.20-2.21), while the 50 exon amino acid data set had the opposite pattern with clade 3 sister to the group of clades 4, 5, and 6 (Figure 2.22). Many of the species-level relationships were highly supported with a posterior probability of one in the different topologies,

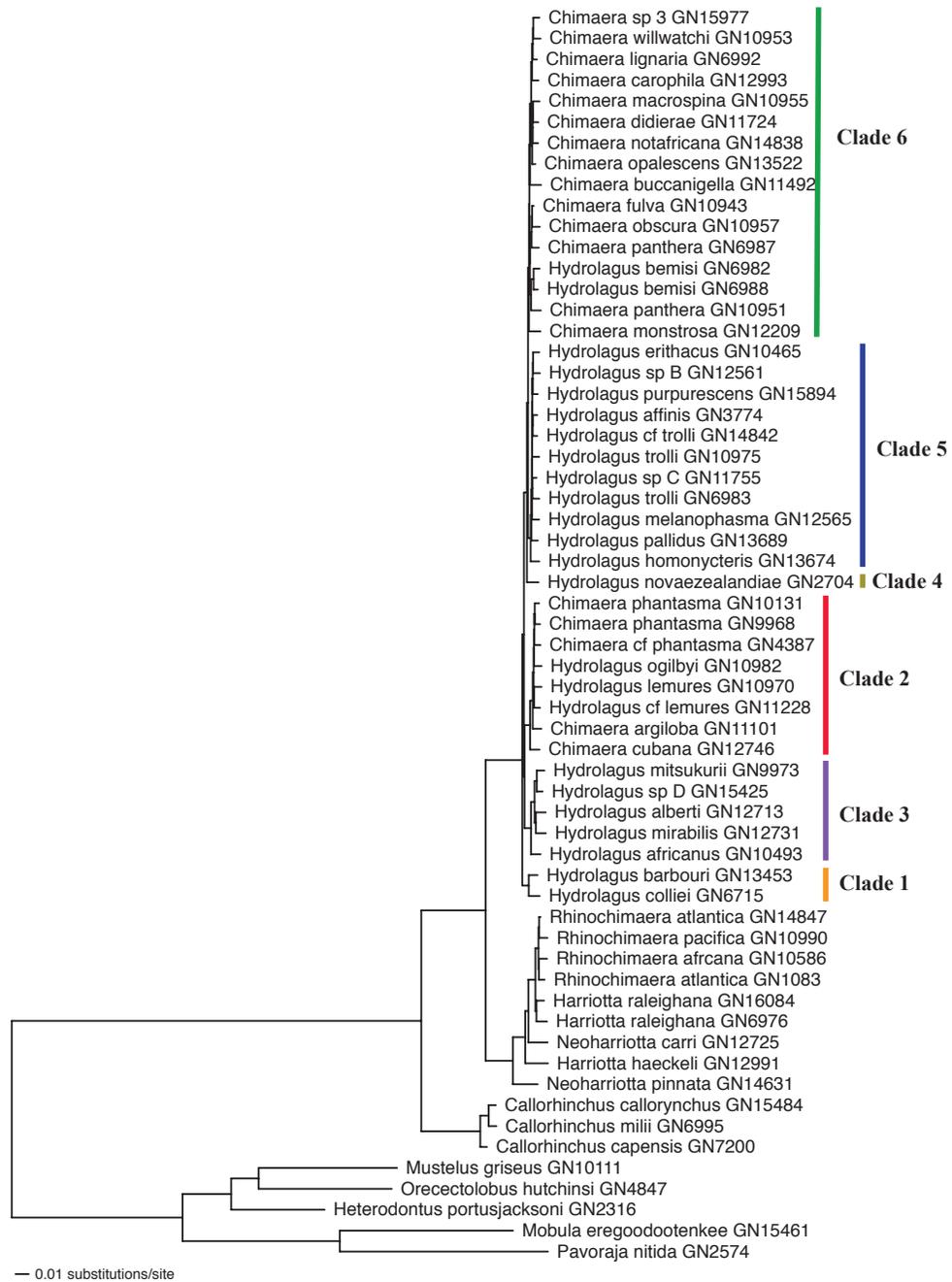


Figure 2.19 The Bayesian inference majority rule consensus tree topology of the clock-like filtered exon nuclear nucleotide data set partitioned by codon under a GTR+ Γ +I and SYM+G model. All nodes had a posterior probability between 0.98 and 1.

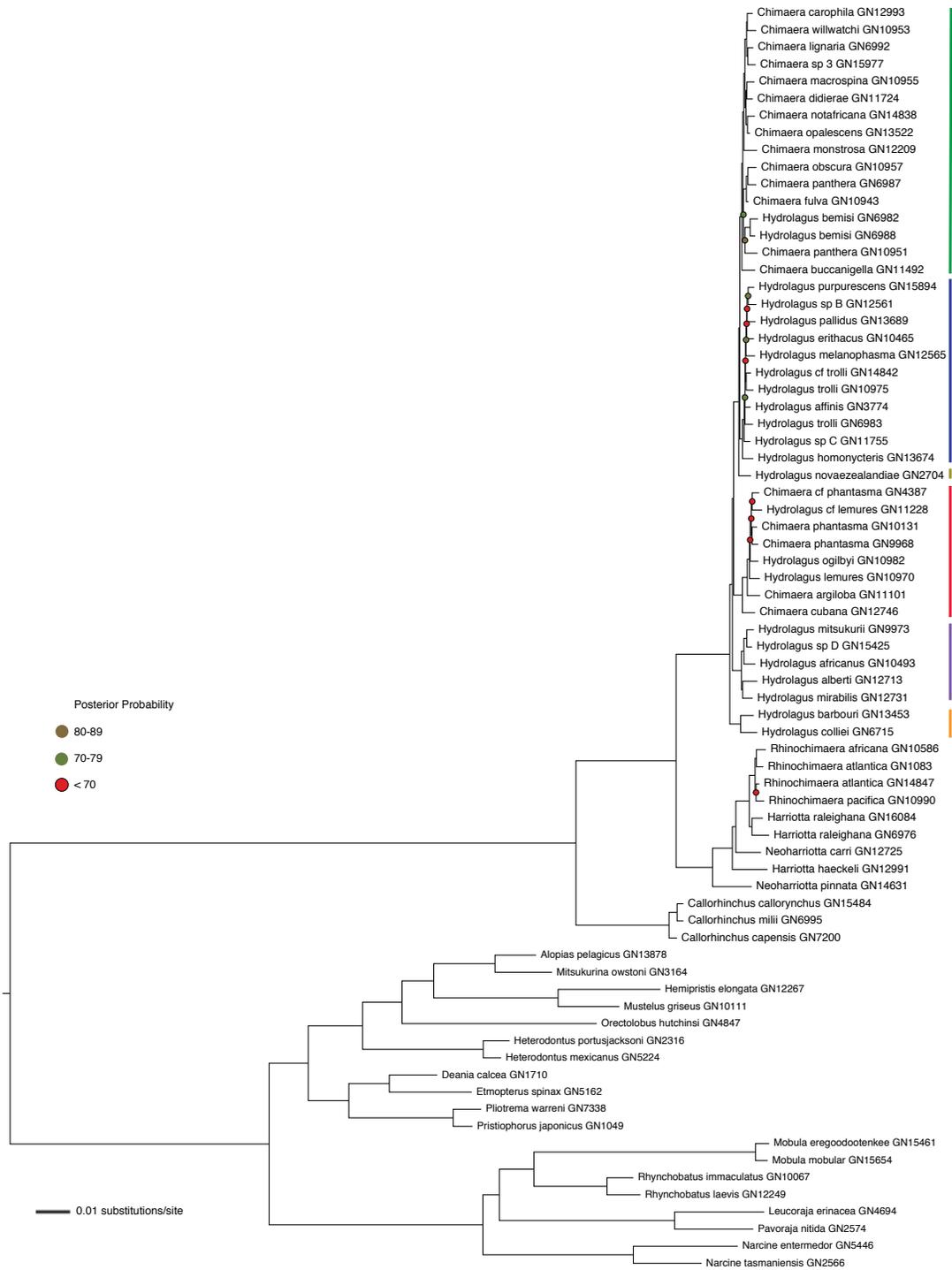


Figure 2.20 The Bayesian inference majority rule consensus tree topology of the 50 exon nuclear nucleotide data set with no partitioning scheme under a GTR+ Γ +I. Nodes without a colored circle represent a posterior probability > 90%.

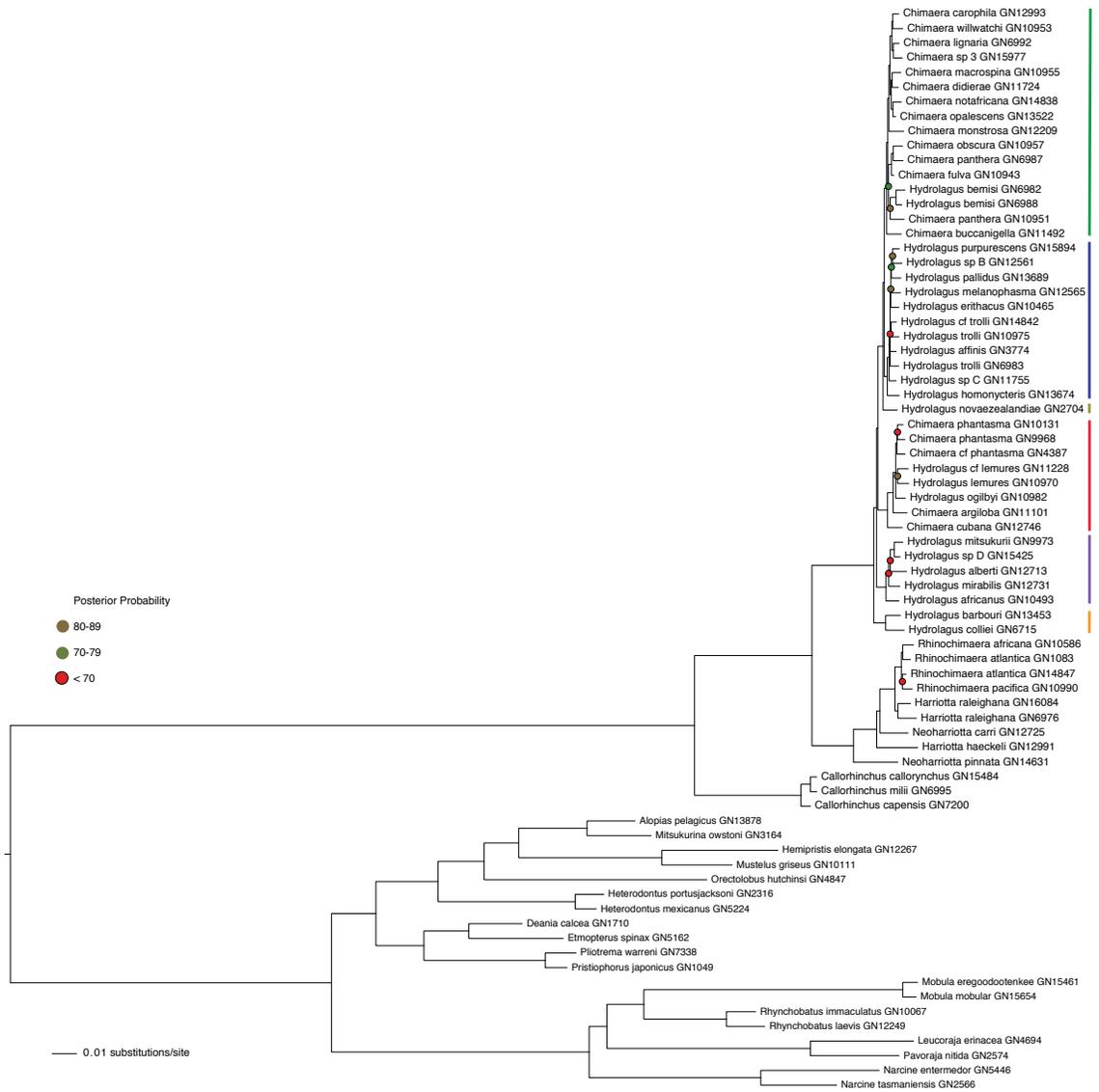


Figure 2.21 The Bayesian inference majority rule consensus tree topology of the 50 exon nuclear nucleotide data set partitioned by codon under a GTR+ Γ +I. Nodes without a colored circle represent a posterior probability > 90%.

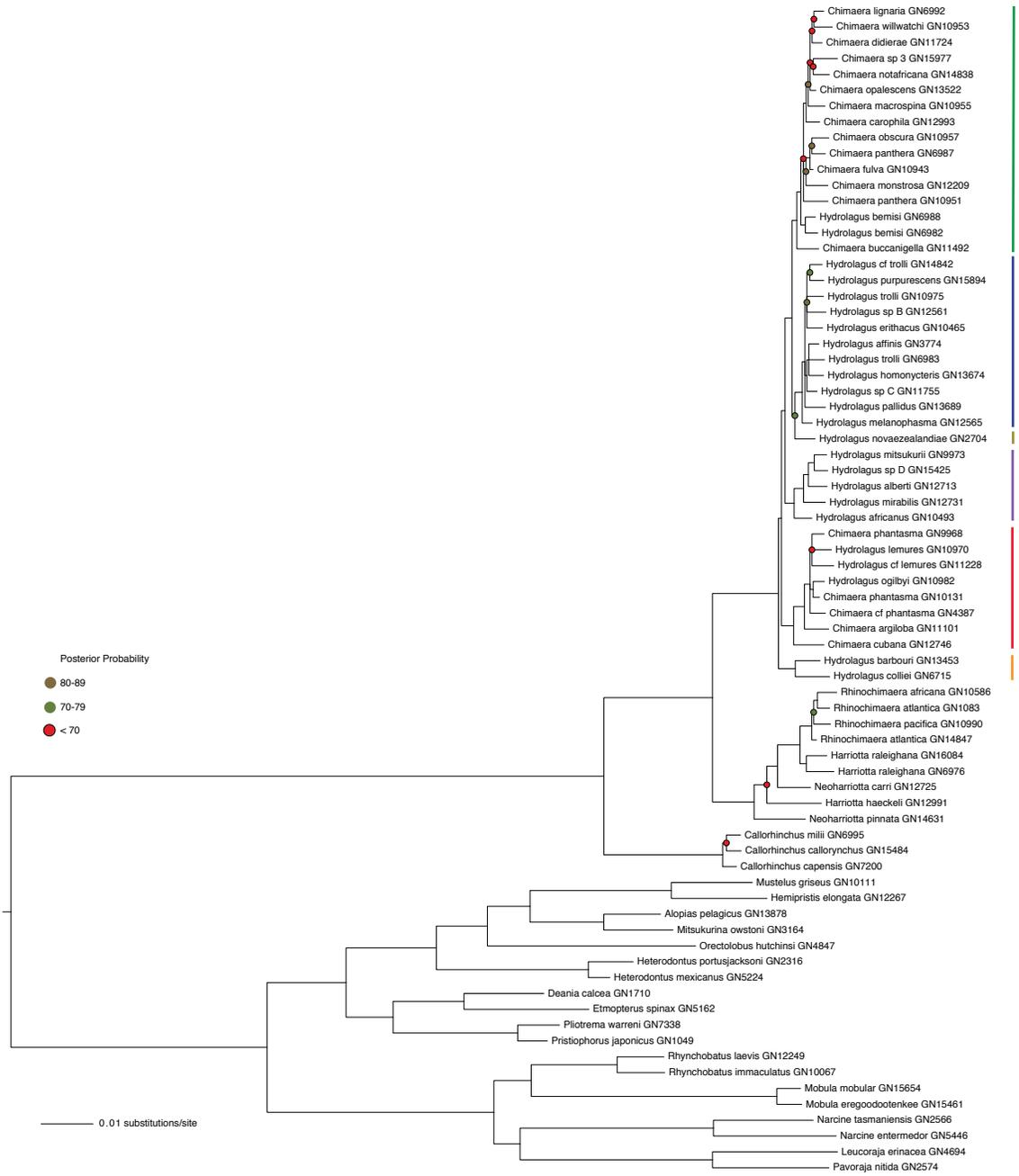


Figure 2.22 The Bayesian inference majority rule consensus tree topology of the 50 exon nuclear amino acid data set with no partitioning scheme under a JTT+ Γ +I. Nodes without a colored circle represent a posterior probability > 90%.

but some of the shallower splits have lower posterior probabilities indicating less reliability for those relationships, especially in the amino acid result. The three families show high divergence from one another, however, within the families, branch lengths are very short, indicating little genetic variation and sequence divergence. This exacerbates the problem of resolving placement of clades to one another, species-level relationships, and species delimitation.

The clock-like filtered BI topology was identical to the ML nucleotide codon topology, and all nodes were highly supported with >0.98 posterior probability (Figure 2.19). The 50 exon BI of nucleotide data with no partitioning (Figure 2.20) and a codon partition (Figure 2.21) produced similar results. There was high support for the major clades and some of the within-clade relationships, but lower posterior probabilities were seen at several of the species-level splits. They differed in the relationships within clades 2, 3, and 5, with the other clades congruent. The resulting BI topologies were similar to their counterpart ML topologies, with the same relationships among the major clades, with differences in the relationships towards the tips of the trees (i.e. clades 2, 3, and 5). The 50 exon amino acid BI resulted in a similar topology (Figure 2.22), but resulted in overall less support than nucleotide BI results. The overall BI amino acid topology was similar to the ML topology, especially in the placement of clades 2, 3, and 4, but differed in shallower node splits within clades 2, 5, and 6.

In general, BI topologies showed better posterior probability support for nodes than the more conservative bootstrap with ML inference (Hillis & Bull, 1993). Branch lengths were similar between BI analyses and the ML analyses. Only one other study has used BI to estimate the phylogenetic relationships among chimaeroids, where 14 species were included (Licht et al., 2012). While the tree topologies between the ML and BI analysis were similar here, incongruences were also noted. The common patterns between the two methods provide

additional support for those phylogenetic relationships, especially among the major clades, and the species that fall within them.

Divergence Time Estimation

Based on the divergence dating of holocephalans using multiple nuclear exons, holocephalans diverged from a common ancestor with elasmobranchs in the Devonian period of the Paleozoic era, as speculated in other studies (Didier, 1995; Pradel et al., 2009; Grogan et al., 2012). Grogan et al. (2012) stated that holocephalans reached their peak diversity during the Carboniferous, and that they suffered extinctions during the end of the Permian. An ancestral lineage must have survived this period until the Callorhynchidae family diverged in the Triassic to Jurassic periods of the Mesozoic. Then, in the late Jurassic period, the remaining two families, Rhinochimaeridae and Chimaeridae diverged, leading to the present day diversity.

The tree topology from the combined BEAST runs using the 50 exon data set under a codon partitioning scheme (Figure 2.23) was mainly congruent with the ML and BI topologies under the same data set and partitioning scheme. The only differences lay in a few of the species-level relationships within clade 2 in the ML analysis and clade 5 in both the ML and BI analyses. The majority of node splits had a posterior probability of one or near one. However, there were several shallower internal and terminal nodes that had lower posterior probabilities. This is likely a reflection of the smaller data set and less signal available in the data to resolve the nodes, or potentially the result of homoplasy.

Divergence time estimates are given as a range, equivalent to the 95% credible interval, and posterior means provided as a point estimate. The chondrichthyans diverged from a common ancestor with bony vertebrates between 418.66 Ma in the Devonian and 737.98 Ma in the Neoproterozoic era, with a posterior mean of 566.75 Ma (Neoproterozoic era of the Precambrian;

Figure 2.23). The mean estimate is much older than currently considered based on fossil evidence. Fossil information places a soft maximum bound of 462.5 Ma on this split (Benton et al., 2009). A molecular dating study estimated a divergence between chondrichthyans and bony vertebrates at 464-443 Ma (Inoue et al., 2010). They placed an upper and lower constraint on this node, while in this study, no fossil calibration was used for this divergence. Thus, no prior information was used to place soft bounds on the posterior distribution for this deep node age. This is likely a major reason why such a large credible interval, and older point estimate was estimated here. It has been shown that divergence time estimates rely heavily on the calibration of nodes (Tamura et al., 2012). Several studies have concluded that calibration priors set on deeper nodes, like the root of the tree, yield more precise estimates (Hug & Roger, 2007; Sanders & Lee, 2007; Ho & Phillips, 2009; Mello & Schrago, 2014), with a similar result for accuracy (Mello & Schrago, 2014). Lack of a calibration for the deepest node may lead to more uncertainty in age estimation, and thus, a larger credible interval, as estimated here.

It also has been thought that deeper nodes compared to younger nodes may be more difficult to resolve because they are older, and thus, there may be more uncertainty in the molecular clock (Yang & Rannala, 2006), a greater effect of rate variation (Rannala & Yang, 2007), and more substitution saturation (Schwartz & Mueller, 2010; Lukoschek et al., 2012). Even though the model of molecular evolution employed in the analysis takes into account substitution rate variation and substitution saturation, these potential issues may not be modeled properly leading to biases. Homoplasy is likely a problem at deeper nodes where more time has passed, and lots of substitutions have occurred independently among lineages. Rates of substitution saturation can often be incorrectly estimated, biasing branch length estimates (Yang, 1996; Xia et al., 2003; Brandley et al., 2011). Additionally, while the data set was originally filtered for clock-like exons, those that do not appear to have accelerated rates of evolution across lineages, this was

only done for the original data containing only the holocephalans and outgroup elasmobranchs. The additional bony vertebrate taxa were later added, after the selection of exons. Thus, it may be that rate variation between the ingroup holocephalans and the bony vertebrates may also be contributing to the older estimation and large variance. The evolutionary model may not hold for these evolutionarily distance groups. As the rate of evolution becomes larger between groups, so does the variance (Tamura et al., 2012), which is likely contributing to the large credible interval estimated here. It would be beneficial in the future to re-run the analysis, placing a fossil constraint on the chondrichthyan-bony vertebrate split based on fossil information, and evaluate how that affects this divergence age estimate, as well as shallower node estimates.

The holocephalans diverged from a common ancestor with the elasmobranchs between 288.45 Ma in the Permian and 510.79 Ma in the Cambrian (395.83 Ma of the Devonian), within the Paleozoic (Figure 2.23). This deeper node also showed a large 95% credible interval. Previous molecular divergence time dating estimates this split between 410 Ma to 494 Ma (Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012; Renz et al., 2013). These estimates are congruent with fossil evidence of chondrichthyans, where holocephalans and elasmobranchs had diverged by at least 410 Ma (Coates & Sequiera, 2001). The timing of the split estimated in this study is slightly younger than these previous estimates, but the credible interval includes all of the previous estimated divergence times, fossil dating, and falls within the Paleozoic era. Previous estimates mainly utilized mitochondrial genes, with one study including a nuclear gene. In general, the mitochondrial genes are considered to evolve at a faster rate than nuclear genes. This study used conserved, slow-evolving exons, which would likely have a slower rate of evolution than mitochondrial genes, and less saturation at deeper nodes. Thus, time estimates based on mitochondrial genes may lead to overestimates compared to nuclear genes. This is a potential reason for a younger point estimate and credible interval. The large credible interval

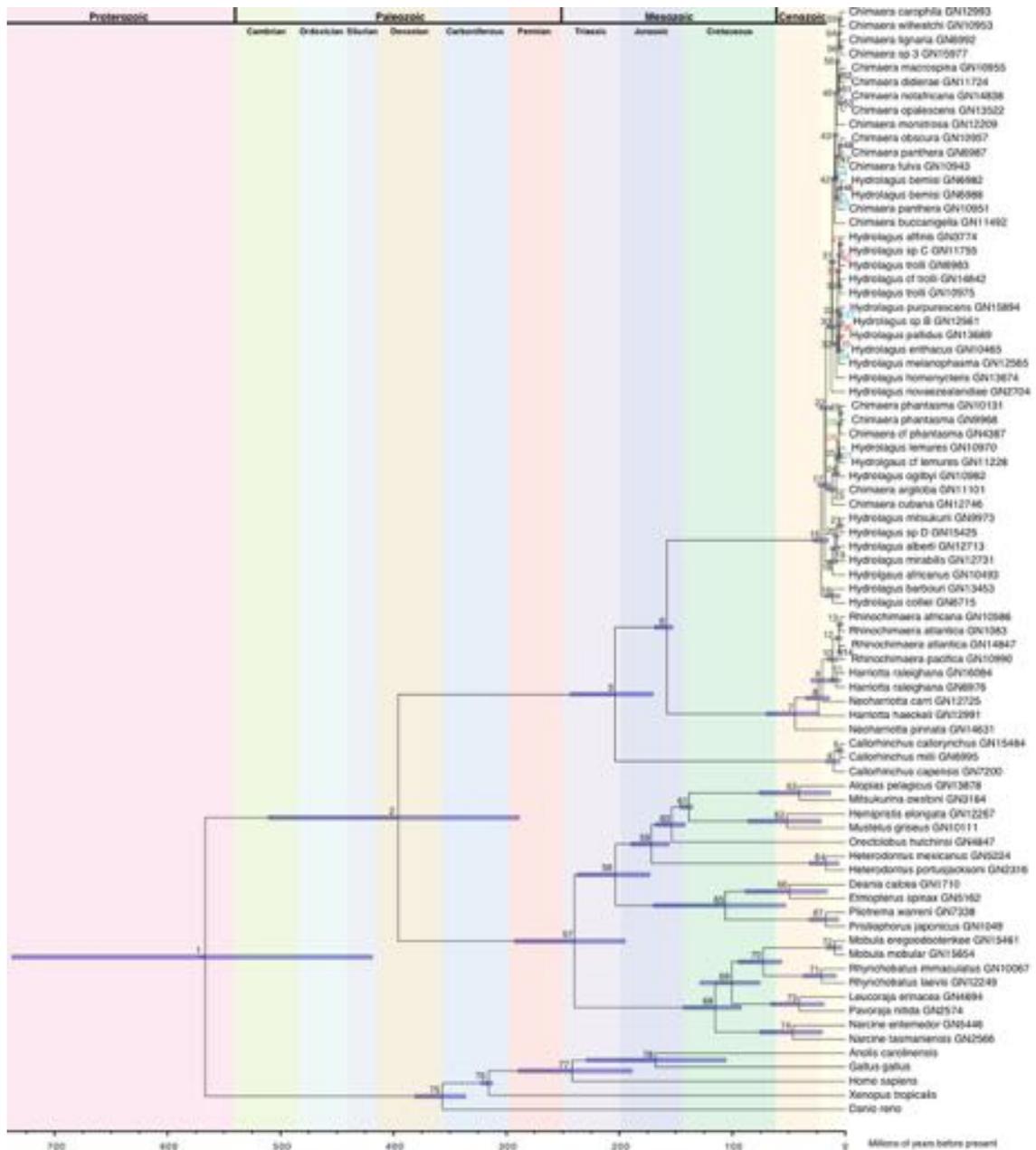


Figure 2.23 Divergence time tree of 50 exon nuclear nucleotide data set under a codon partition. Bars represent 95% credible intervals of node age estimates. Nodes are numbered, posterior mean and credible intervals of ages given in Table 2.6. Node number color represent posterior probabilities: black = 0.90-1, blue = 0.80-0.89, green = 0.70-0.79, red < 0.70. Axis scale represent millions of years before present.

could be an effect of the potential reasons outlined above (i.e., deep node, inconsistent molecular clock, rate variation, saturation). Additionally, the lack of a calibration on the root, the bony vertebrate outgroup, and the likely discrepancies in rate variation among this outgroup and holocephalans, may have effects on downstream node estimates, leading to more variance, especially in this adjacent node. Also, the fossil calibration for this node had a large bound, 280 to 462.5 Ma, whereas previous calibrations for this node placed a hard minimum bound at 410 Ma, with no upper bound. The prior distribution for this node may be inappropriate, leading to the large variance in the estimated node time. It would be advisable to investigate how altering the prior distribution would affect the age estimate and credible interval, in order to reduce the uncertainty in the divergence time estimate.

Callorhinchidae split off during the Mesozoic era between 169.97 Ma of the mid-Jurassic and 243.28 Ma of the mid-Triassic (203.86 Ma; Figure 2.23). The three species of *Callorhinchus* are quite young, diverging from one another relatively recently in the early Miocene to Pleistocene of the Cenozoic era. Others have also found that the three species are quite young (Inoue et al., 2010; Licht et al., 2012). This estimate is similar to previous divergence dating estimates that had a credible interval from 125 Ma to 320 Ma, with point estimates of 167 Ma (Inoue et al., 2010), 187 Ma (Licht et al., 2012), and 220 Ma (Heinicke et al., 2009). The mean estimate in this study is older than two of the previous estimates, but the credible interval is close to these previous dates.

Rhinochimaeridae and Chimaeridae diverged from a common ancestor between 152.5 Ma and 168.54 Ma in the late to middle Jurassic (157.85 Ma; Figure 2.23). This is a slightly older estimate than previous works at 122 Ma (146-98 Ma; Inoue et al. 2010), and 107 Ma (182-51 Ma; Heinicke et al., 2009), but in concordance with an estimate of 159 Ma (164-156 Ma; Licht et al., 2012). The rhinochimaerids are relatively young, diverging in the Cenozoic era.

Neoharriotta pinnata diverged from the other members in the Eocene. The *Rhinochimaera* diverged from the other members in the Miocene.

Extant species of Chimaeridae, including *Hydrolagus* and *Chimaera* appear to be relatively young, with recent divergences since their MRCA in the Cenozoic, from the Oligocene forward (Figure 2.23). Previous estimates of divergence times within Chimaeridae only included a few of the species (Inoue et al., 2010; Licht et al., 2012). The estimates in this study for species within Chimaeridae are much younger than previous work, all being quite recent. Since *Hydrolagus* and *Chimaera* were found to be paraphyletic, a divergence date for the two genera cannot be estimated. Caution should be taken when interpreting the age estimates, however, due to the lower support for several of the nodes within the Chimaeridae family.

Within the elasmobranchs, the split between neoselachian sharks and batoids was estimated between 194.81 and 292.58 Ma (239.71 Ma; Figure 2.23). This estimate is largely congruent, yet slightly older, with the known fossil record for the first appearance of neoselachians and batoid in the middle Triassic (Cuny et al., 2001; Underwood, 2006). The remaining splits within sharks and rays provide reasonable age estimates based on known fossil records for respective lineages.

The outgroup, five bony vertebrates, node estimates were found to be younger than fossil evidence suggests (Figure 2.23). The split between birds and mammals was estimated at between 188 and 289 Ma (242 Ma). However, fossil evidence, and the node calibration ages used in the analysis suggests a range of 312.3 and 330.4 Ma. The node age estimate for the MRCA of the toad-bird-mammal split was between 312 and 322 Ma (315.8 Ma), which is slightly younger than fossil data suggests at 330.4 and 350.1 Ma. Younger molecular divergence dates than previously estimated or based on fossil evidence could be a result of either (1) incorrect fossil calibration dates, or (2) under-sampling of taxa (Near et al., 2005; Schulte, 2013). Inaccurate fossil dating

due to an incomplete fossil record can lead to underestimation of molecular divergence ages (Marshall, 1990; Springer, 1995; Near et al., 2005). Also, if a fossil was assigned to the incorrect node (Benton & Ayala, 2003), or there was an error in the dating of the fossil (Conroy & van Tuinen, 2003), these could contribute to biases in the estimation of divergence times. However, in this case, the bird-mammal split is a widely used calibration (Benton et al., 2009), and unlikely that the fossil dating is inaccurate, but not out of the question. This would lead one to question whether the taxon sampling of bony vertebrates was insufficient to accurately estimate the divergence times. Schulte (2013) found that clades with few taxa sampled had younger node time estimates than clades with greater taxon sampling. Since only five taxa were sampled from within bony vertebrates, this may have contributed to the younger age estimates within the outgroup found in this study. A lack of adequate phylogenetic signal in the data and systematic error could also contribute to this issue. Divergence age estimates for all lineage splits are given in Table 2.6.

Several biases exist that may affect divergence age estimates, including paralogy issues in nuclear data, substitution rate heterogeneity between lineages, genes, sites within genes, and over time (Wray et al., 1996; Bromham & Hendy, 2000; Levinton, 2008), fast early rates of molecular evolution (Bromham & Hendy, 2000), fossil calibrations (Levinton, 2008), model of molecular evolution (Levinton, 2008), and inaccuracy of molecular clock (Bromham & Hendy, 2000). The original set of exons were chosen because they were orthologous among the model vertebrates, however, paralogs among these exons could potentially be present within Chimaeriformes. Rate heterogeneity is taken into account in the model of molecular evolution, however, if the model is not appropriate for the data, this could lead to inaccurate estimates. This is especially true for evolutionary rate variation between the evolutionarily distant lineages of chondrichthyans and bony vertebrates, which may have had an effect on divergence time

estimates of the nuclear data. Model selection was used in effort to choose the most appropriate model for the sequence data. Thus, efforts were made to minimize these biases, but it is unlikely given the taxon sampling, data set, and number of parameters estimated based on the data, that all bias had been minimized, and one or more of these may attribute to the under and over-estimation and variance in node ages seen in this analysis.

Conflicting age estimates across different studies can be attributed to several factors, including different phylogenetic estimation software, different fossil calibration points and priors (Marshall 1990, 2008; Ho & Phillips, 2009; Inoue et al., 2010; Dornburg et al., 2011), gene marker selection, taxon sampling, substitution rate heterogeneity (Solitas et al., 2002; Dornburg et al., 2012), molecular clock model (Drummond et al., 2006; Ho & Larson, 2006), and homoplasy (Dornburg et al., 2014). Future studies would benefit from the use of multiple different estimation programs to compare results, inclusion of more sequence data (e.g., full nuclear data set), which would require more computational resources, and changing the fossil calibrations to see how they affect the resulting age estimates.

Overall, the common ancestral lineage of holocephalans originated in the Devonian of the Paleozoic era. The lineage persisted for a long time, through major extinction events of the Carboniferous and Permian-Triassic periods. Holocephalans are thought to have reached their greatest diversity during the Carboniferous, with wide spread extinction by the end of the Permian (Grogan et al., 2012). The modern day families started to diverge in the late Triassic to early Jurassic period, with the Callorhynchidae lineage and the lineage leading to Rhinochimaeridae and Chimaeridae. This period of time was also characterized by the Triassic-Jurassic mass extinction event. The lineages must have survived this period to persist and lead to modern day species, potentially by seeking out the deep sea, as many extant species are deep sea

Table 2.6 Divergence ages of nodes estimated from the 50 exon nuclear data set in BEAST. Node ages are represented as both a point estimate (posterior mean) and a range (95% credible interval). Ages are in millions of years (Ma). Nodes numbers are shown in Figure 2.16.

Node	Posterior Mean (Ma)	95% Credible Interval	
		Lower (Ma)	Upper (Ma)
1	566.75	418.66	737.98
2	395.83	288.45	510.79
3	203.86	169.97	243.28
4	10.12	4.4	17.06
5	4.37	1.52	7.89
6	157.85	152.5	168.54
7	44.29	23.44	69.95
8	23.91	14.04	35.03
9	20.19	11.45	30.23
10	10.73	6.19	15.94
11	8.23	3.95	13.11
12	5.71	3.1	8.65
13	4.56	2.08	7.27
14	4.43	1.87	7.23
15	21.27	14.76	28.63
16	11.11	4.54	17.98
17	18.44	13.05	24.5
18	11.18	6.69	16.04
19	8.83	5.09	13.07
20	7.42	3.93	11.26
21	4.27	1.68	7.29
22	17.06	12.05	22.77
23	11.56	7.46	16.04
24	8.26	5.18	11.65
25	6.45	3.95	9.15
26	5.84	-	-
27	4.8	2.42	7.37
28	4.6	2.4	6.93
29	3.91	1.79	6.14
30	12.35	8.75	16.28
31	11.17	7.97	14.64
32	8.28	5.51	11.33
33	6.49	4.21	8.91
34	5.53	3.42	7.79
35	5.16	3.11	7.33
36	4.8	2.77	6.96
37	3.84	1.92	5.94
38	5.66	3.54	7.85
39	4.55	2.32	6.82
40	5.11	3.05	7.37

41	4.11	1.98	6.32
42	9.44	6.71	12.38
43	8.56	6.07	11.2
44	7.86	5.41	10.47
45	6.72	4.21	9.34
46	3.71	1.66	6.01
47	5.22	2.74	7.75
48	4.05	1.87	6.41
49	7.47	5.16	9.96
50	6.44	4.34	8.69
51	5.21	3.11	7.36
52	4.35	2.24	6.53
53	3.58	1.54	5.74
54	5.12	3.13	7.23
55	4.35	2.31	6.44
56	3.92	1.92	5.96
57	239.71	194.81	292.58
58	203.47	172.94	237.37
59	171.68	156	189.89
60	153.65	141.65	167.96
61	138.81	135	146.31
62	51.63	21.36	86.06
63	40.8	12.88	76.19
64	17.08	5.58	31.74
65	106.37	52.76	169.71
66	49.27	15.75	88.59
67	17.28	5.6	31.79
68	115.15	92	143.34
69	100.41	75.41	128.6
70	72.51	56	94.63
71	20.99	7.73	36.94
72	9.26	3.05	16.54
73	40.66	18.62	65.93
74	46.53	20.13	75.81
75	356.57	336.08	381.04
76	315.77	312.3	322.66
77	242.07	188.61	289.29
78	167.96	105.64	229.22

dwellers. It was then in the late Jurassic period that the remaining extant families, Rhinochimaeridae and Chimaeridae diverged from one another.

A very long time passed between the divergence of the Callorhinchidae lineage and the extant diversity. It wasn't until the Miocene, approximately 10 Ma, that modern callorhinchids started to diverge. These three recognized species all reside in the southern hemisphere. It is likely that the ancestor of these three species resided in the southern hemisphere, in what would today be considered the Southern Ocean. By this time, Pangea had broken up into modern day continents, and present day oceans were formed. From a biogeographical perspective, the ancestral lineage was likely present in the Southern Ocean then one population migrated to South Africa landmass, one to South America, and the other to Australia and New Zealand regions. Extant species of Rhinochimaeridae began to diverge in the Eocene (~ 44 Ma), and the remaining diversity followed in the Miocene. Again, it would appear that the ancestral lineage likely occurred in the Southern Ocean region, as *N. pinnata* occurs in the Indian Ocean, and *H. haeckeli* occurs in the south Pacific Ocean. The ancestor of *N. carri* and the remaining lineages may have moved towards the South Atlantic, as several remaining lineages occur in the North Atlantic region. However, there also appears to be some migration to the Pacific Ocean, probably between Antarctica and South America. Interestingly, the closely related lineages of *R. africana* and *R. atlantica* both occur in the Northern Hemisphere, but in different ocean basins, while *R. atlantica* (GN14847) and *R. pacifica* occur in the Southern Hemisphere, but in different ocean basins. Extant Chimaeridae species share a MRCA approximately 21.27 Ma in the Miocene. It too would appear that the ancestral lineage began in the Southern Ocean region. The basal clade, *H. barbouri* and *H. colliei* appear to have migrated upwards to the North Pacific Ocean, with one staying in the western North Pacific and the other migrating to the eastern North Pacific. The ancestor of the remaining lineages was probably present in the Southern Ocean, where it diverged

into the lineages leading to clades 2 and 3, and clades 4, 5, and 6. Clades 2 and 3 likely share an ancestor that was present in the South Atlantic Ocean, as *H. africanus*, *H. alberti*, *H. mirabilis*, and *C. cubana* all occur in a region of the Atlantic Ocean. This would still make it feasible for divergences that lead to species distributions in the Indian and Pacific Ocean of the remaining lineages within clades 2 and 3, as these regions were open, without major landmasses to block migration. The common ancestor of clades 4, 5 and 6 was also likely to be present in the Southern Ocean or Indian Ocean, as the majority of the extant lineages occur in the Indian Ocean or Pacific Ocean. A few species occur in the North and South Atlantic oceans (i.e., *H. affinis*, *H. pallidus*, *C. opalescens*, *C. notafriana*, *C. monstrosa*), indicating that throughout species diversification, some lineages likely migrated from the Southern Hemisphere of the Indian Ocean to the South and North Atlantic. Overall, based on the phylogenetic reconstruction, divergence time estimation, and distribution of extant species, the majority of ancestral lineages that lead to extant species likely began in the Southern Ocean of the Southern Hemisphere, with migration outwards towards the Atlantic and Pacific Oceans, as well as upwards into the Indian Ocean.

Conclusions

This study represents the first comprehensive phylogeny of chimaeroid fishes, which included multi-exon nuclear data and dense taxon sampling. The inclusion of 55 chimaeroid lineages makes this the first phylogenetic reconstruction of this group to include the majority of taxa and all genera, in order to estimate a species-level tree topology. Overall, the three families of Chimaeriformes were each monophyletic with full support. Callorhynchidae was the basal clade, with identical species-level relationships in all resulting topologies. Rhinochimaeridae and Chimaeridae formed a monophyletic group. Within Rhinochimaeridae, there was generally high

support for the deeper level relationships, with *N. pinnata* the most basal species. There were different species-level relationships among analyses within *Rhinochimaera* as well as lower support. While *Rhinochimaera* appears to be a monophyletic genus, *Neoharriotta* and *Harriotta* were not recovered as monophyletic genera. Within Chimaeridae, the two genera were found to be paraphyletic. There were six major Chimaeridae clades recovered in all of the analyses, and the same species were found in each of the clades throughout. However, the relationships within the clades showed differences among analyses, and the relationships between a few of the clades also differed. The resulting genera relationships have wider reaching implications for the taxonomy of this group.

Overall, the full and clock-like filtered nuclear data sets resulted in the most highly supported topologies, with the major difference being between nucleotide and amino acid topologies. A strict consensus tree topology between the full and clock-like filtered nuclear data sets is presented in Figure 2.24, and represents the estimate of the phylogenetic relationships within Chimaeriformes for this nuclear exon data. Biogeographical patterns can be investigated using this consensus tree, where it would appear that the ancestral lineage of holocephalans likely resided in the Southern Ocean with outward and upward migrations into the other ocean basins within families over time. Present day species are known from all ocean basins, except the Arctic, but a large portion of chimaeroid diversity resides in the Indian and Pacific oceans.

Resulting estimated topologies within a data set, among the different exon sets, partitioning schemes, and methods, were similar. In general, lower support for branching patterns were present in estimates based on smaller sized data sets and amino acid data sets. Amino acid sequences are generally considered to be more conserved because they represent the functional information of a gene, and may be under more constraints. Nucleotide sequences are considered to hold more information because they can vary more due to the fact that synonymous

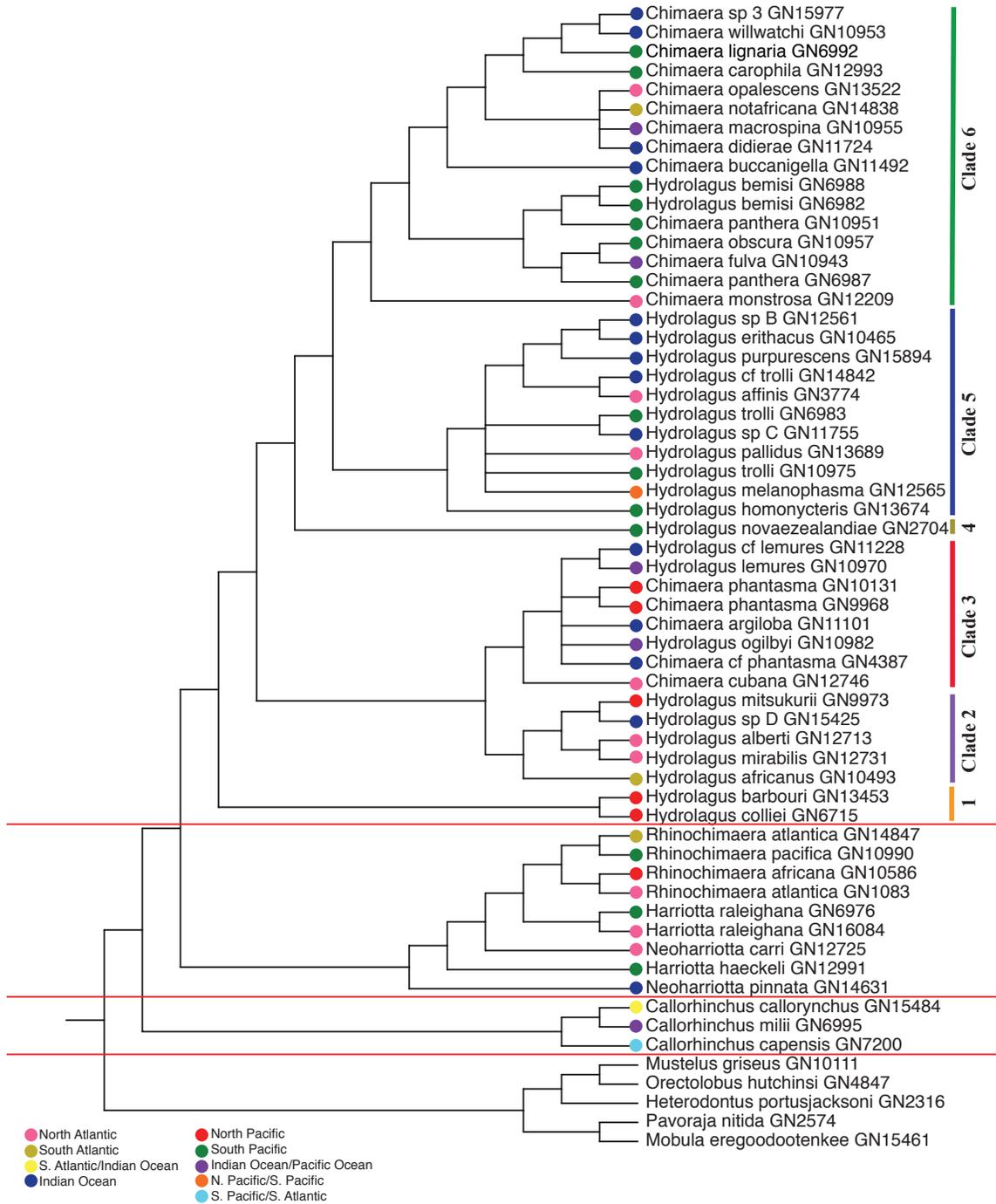


Figure 2.24 Strict consensus tree topology based on the full nuclear 1264 exon and clock-like filtered nuclear exon data sets maximum likelihood trees under a codon partition. Polytomies indicate relationships that were incongruent between topologies. Terminal node colors represent geographic distribution of sampled individuals and/or species distributions.

substitutions do not change the amino acid. Thus, nucleotide sequences provide more characters for phylogenetic inference that may not be present at the level of amino acid. This provides one rationale for why the topologies based on amino acid data may show less support. Additional sequence data would be necessary to try and resolve these relationships with higher support. The 50 exon data set also was found to show less support compared to the other data sets that held much greater sequence lengths, potentially due to the reduction in informative characters, or model mis-specification. While it appears that the data sets which contain more sequence data and informative characters (e.g., full nuclear exon set) produced more reliable results, the use of the reduced exon data sets provide evidence that the methods and models are rather robust for these data sets since all the topologies shared the same clades and species within these clades. When different methods (e.g. ML and BI, nucleotide and amino acid characters) result in similar topologies, it provides confidence that the topologies represent the true history of the data (Carranza et al., 2002; Holland & Hadfield, 2004). Thus, the relationships that were consistently resolved across data sets and methods provide support that these are true evolutionary relationships for these nuclear exons.

There will almost always be biases and error associated with phylogenetic inference, but it is thought that with increasing data that the stochastic error may be reduced (Avice, 2004; Ballard & Whitlock, 2004). However, as alignment size and number of taxa increases, phylogenetic analyses become more computationally demanding. Not to mention that ML and BI inferences, themselves, are computationally cumbersome. One limitation in this study was the lack of unlimited computational resources to investigate the various data sets with respect to model choice, partitioning schemes, and methodology. Since the larger data sets (i.e., sequence length) resulted in much higher support for the estimated topologies, future work should use the larger set of exons (e.g., 1,264 exons), and explore the full range of ML and BI analyses under

different partitions and models of molecular evolution, using both nucleotide and amino acid sequence data. This will provide a better understand of how changes in methods affect estimates, and likely gain a better understanding of the evolutionary history of this group of animals, as seen through congruence and incongruence in topologies. Overall, it may provide a more precise and accurate estimate of the phylogeny of Chimaeriformes using this nuclear exon data.

The total branch lengths for all terminal lineages were quite similar among holocephalans, indicating that they are evolving at a similar rate, and in a clock-like manner, with respect to the sampled exons. Callorhynchidae appears to have a somewhat slower rate, and Rhinochimaeridae and Chimaeridae have a similar slightly faster rate. Estimated internal and terminal branch lengths within the respective family clades were quite short for the majority of lineages. This may indicate little genetic variation between lineages within clades. This makes interpreting species boundaries and their sister relationships challenging. A discussion of the taxonomic implications of the molecular phylogenies will be presented in Chapter 3. The low genetic variation among some of the species may be a consequence of the sampled exons and what appears to be a much slower rate of evolution compared to elasmobranchs and bony vertebrates. These exons were chosen because they are slowly evolving protein-coding genes that are shared across vertebrates, and thus, highly conserved. With a generally very slow rate of evolution among holocephalans, and quite recent divergence times among extant lineages, it is possible that these exons have evolved little since the species' have diverged. On the other hand, such little genetic diversity among sets of lineages may also indicate population variation. Low genetic variation among populations has been attributed to small population sizes, restricted gene flow, and genetic drift (Furlan et al., 2012). However, no data exists on the population sizes and population genetics of chimaeroid species. It is also a possibility that within the Chimaeridae clades and *Rhinochimaera*, that their respective ancestral lineages lead to a burst of species

divergences, given the divergence times of many of these lineages, and while evolving independently since this time, the young ages and a very slow rate of evolution, little genetic variation has yet to accumulate. Different genes or DNA regions can vary with respect to their rates of molecular evolution. Therefore, different molecular markers can vary in their capacity to be informative at differing tree depths (Avice, 2004). The nuclear exons sampled in this study may represent sequence data that are evolving at a rate that is too slow to resolve many of the shallower nodes on the tree, as evidence by low support and differences among analyses. Future work may benefit from selecting different nuclear markers that have a higher substitution rate in Chimaeriformes, providing more variability to better resolve species-level relationships. This would also be necessary to recognize if some of these relationships in the tree represent population variation or unique species.

The evolutionary process is inherently stochastic, with rates, mode, and direction of evolution varying within and between characters, genes, populations, species, and through time. Although there have been advances in methodology and models to incorporate and deal with this variability, one cannot expect that all error and bias is remedied. Thus, this is a major limitation of phylogenetic analysis. While more parameter-rich models may be used to help incorporate the unique patterns in the data, the true history is not known and thus, the true phylogeny will not be error-free.

There are two main types of error, random error due to limited sampling, and systematic error (Swofford et al., 1996; Rokas et al., 2003; Jeffroy et al., 2006; Philippe et al., 2011). This study used dense taxon sampling and large multi-locus nuclear data sets in an effort to reduce random error and recover an accurate phylogeny. Systematic error is the result of incorrect assumptions or model mis-specification of DNA or amino acid sequence evolution (Swofford et al., 1996; Felsenstein, 2004), which results in a non-phylogenetic signal (Ho and Jermiin, 2004,

Philippe et al. 2005a). There are several causes of systematic error including base composition heterogeneity (Foster, 2004, Galtier & Gouy, 1995, Lockhart et al., 1992; Philippe et al., 2005a; Rodriguez-Ezpeleta et al., 2007), across-site rate variation (Yang, 1994; Lopez et al., 2002; Philippe et al., 2005a), heterotachy (site-specific rate heterogeneity through time; Kolaczkowski & Thornton, 2008; Philippe et al., 2005a; Philippe et al., 2005b; Spencer et al., 2005), site non-independence (Robinson et al., 2003; Rodrigue et al., 2006), and site heterogeneous nucleotide or amino acid replacement (Lartillot & Philippe, 2004; Pagel & Meade, 2004), all due to substitutions not accurately modeled by the method (Philippe et al., 2005a). One way to minimize this type of error is by using appropriate evolutionary models for the data, but again, these models may not fully represent the variability and history of the data, resulting in incongruence and a misleading phylogeny.

Other sources of incongruence can come from factors that cause different histories among the characters being analyzed. Orthology is a very important assumption, and when violated due to mechanisms such as gene duplication which causes paralogy (Maddison, 1997; Philippe et al., 2005a), horizontal gene transfer (Maddison, 1997; Bergthorssen et al., 2003; Philippe et al., 2005a), lineage sorting (Maddison, 1997; Satta et al., 2000; Philippe et al., 2005a), or hybridization introgression (Maddison, 1997), the gene no longer has an identical history to the taxa, and can lead to biases and confound the true phylogenetic signal. Here, putatively single-copy exons were sequenced and used for phylogenetic reconstruction. However, it should be noted that while the exons are single-copy across the original model vertebrates interrogated by Li et al. (2013), that does not mean that a duplication did not occur within the Chimaeriformes lineage. This could be difficult to detect if little changes have occurred between the ortholog and paralog.

Divergence time estimation indicated that the holocephalans and elasmobranchs shared a MRCA in the Devonian period. A rather long period of time went by before modern day lineages diverged, with the Callorhynchidae family diverging in the mid-Triassic to mid-Jurassic. Rhinochimaeridae and Chimaeridae shared a common ancestor in the late Jurassic period. All extant species diverged relatively recently in the Cenozoic era. This analysis was limited by the use of the 50 exon data set and a lack of computational resources to run the analysis with a larger data set.

Holocephalans have the potential to be an informative reference for studies aimed at understanding vertebrate genome evolution, as they represent one of the basal lineages of jawed vertebrates. This makes them one of the most distantly related gnathostomes to bony vertebrates. Additionally, it appears that they have a genome organization more similar to mammals, than mammals to teleost fishes, as well as the least derived vertebrate genome sequenced to date, which would provide support for using these fishes in comparative studies with mammals. These fishes can be used to generate a hypothesis about the ancestral condition within holocephalans of characters of interest. However, while a whole genome for *C. milii* has provided insight into vertebrate evolution in several studies, it is actually more representative of the ancestral state to include more diversity than one species, as well as provide more detailed information about the ancestor-descendant relationships among lineages. Since these fish have been evolving independently for millions of years, one species does not necessarily represent the ancestral holocephalans condition better than another. However, by including the majority of diversity within the group, we can aim to get a more precise and accurate estimate, as more information is available to estimate ancestral and derived character states among taxa. This chapter filled an important gap in our understanding of the evolutionary relationships among chimaeroid fishes. A comprehensive phylogeny of these relationships, along with their divergence times, estimated

here can be used to explore patterns and changes in traits of interest within holocephalans by mapping such traits across the tree, as well as estimate the ancestral state of particular traits. This information can then be used across bony vertebrates to explore vertebrate evolution, along with lability, timing and rate of evolution of traits.

Future directions for this study should aim to use the full set of nuclear exons on a High Performance Computing platform, which showed the greatest support and reliability for tree relationships among data sets. It would also be pertinent to select additional multiple nuclear markers that show a greater rate of evolution in Chimaeriformes. A full suite of analyses should be performed, including ML and BI, utilizing nucleotide and amino acid characters, as well as exploring different evolutionary models and partitioning schemes. Comparisons of the slow-evolving versus faster-evolving nuclear markers may help to resolve some of the species-level relationships or population variation that cannot be fully addressed in this study. While congruence between the two independent data sets will provide confidence in relationships within this group, incongruences can also help to better understand their evolutionary history. Future work should also use these independent data sets in divergence time analysis. These analyses should explore the effect of adding and removing outgroup taxa, vary calibration nodes and priors, and use different estimation software to better understand how these factors influence the estimated divergence times.

CHAPTER 3

Molecular Phylogeny of Holocephalans Inferred from Mitochondrial Genomes

Introduction

Using different sources of data in molecular studies can be helpful, as they may be informative at different time depths (Avice, 2004). Also, independent data (i.e., nuclear and mitochondrial DNA) can provide confidence that identical or similar patterns in phylogenetic reconstruction represent true evolutionary relationships (Cunningham, 1997; Rubinoff & Sperling, 2002; Rubinoff & Holland, 2005). Mitochondrial markers have been widely used and show great utility in phylogenetic studies at the species-level and population-level, due to a faster rate of evolution, no recombination, and maternal inheritance, compared to nuclear genes (Pereira, 2000; Avice, 2004; Rubinoff & Holland, 2005; Galtier et al., 2009; Patwardham et al., 2014). The faster substitution rate and more variable sites, along with a smaller effective population size, means alleles are fixed at a faster rate in the mitochondrial genome (Ballard & Whitlock, 2004; Rubinoff & Holland, 2005), and may provide more information for recent time scales and closely related species. Protein-coding regions also may be more useful as their rate of evolution is thought to be more clock-like than non-coding regions (Non et al., 2006; Galtier et al., 2009). The objective of this chapter was to use mitochondrial genomic information and dense taxon sampling to estimate the evolutionary relationships among the majority of chimaeroid lineages, estimate divergence times within the group, and compare to the nuclear data estimates from Chapter 2.

Materials and Methods

Taxon Sampling

The same set of 55 chimaeroid lineages selected for in Chapter 2 were used here. Five elasmobranch species were used as an outgroup to root the phylogenetic trees as in the maximum likelihood and Bayesian inference analysis, as defined in Chapter 2. Fourteen additional elasmobranch species were chosen for divergence time analysis (See Table 2.1 from Chapter 2).

Mitochondrial Genome Capture and Sequencing

The cross-species gene capture method employed in Chapter 2 was used to collect whole mitochondrial genomes of chimaera and elasmobranch species. Custom RNA baits were designed based on several shark species using the coding region of 99 complete or nearly complete mitogenome sequences and 430 full or partially complete D-Loop sequences using the MYBaits system (MYcroarray, Ann Arbor, MI) (Yang, in prep). This platform provides specificity, flexibility, and cost-effectiveness in targeting genomic regions for hybridizations. The genomic libraries prepared for nuclear gene capture were used for mitochondrial genome capture. Targeted mitogenome capture followed the protocol of Li et al. (2013), but only included one round of enrichment, where DNA libraries were hybridized with the shark mitogenome RNA baits. Unique indexes were incorporated during library amplification. Up to 96 indexed samples were pooled in equimolar ratios. Pooled libraries were quantified and diluted, followed by 2 x 300 bp sequencing on an Illumina MiSeq (Illumina Inc, San Diego, CA). Sequence reads were demultiplexed to sort reads into their respective samples by indices. Adapter sequences and low quality reads were trimmed using Cutadapt and FastQC within Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed sequences were

imported into Geneious® v.7.1.9 (Biomatters Ltd., Auckland, NZ). Duplicate reads were removed and unique reads were mapped to a reference sequence of a closely related species (downloaded from GenBank: *Callorhinchus milii* NC_014285, *Chimaera fulva* NC_014288, *Harriotta raleighana* NC_014292, *Hydrolagus lemurs* NC_014290). The *NADH2* sequences obtained in Chapter 2 for each of the 55 chimaeroid lineages were used to validate the mitogenome sequences by comparing the *NADH2* sequence from Sanger sequencing to the one from the mitochondrial capture. In all cases, the *NADH2* sequences collected by the gene capture method matched the sequence from Sanger sequencing. A consensus whole mitogenome for each sample was obtained from mapped reads.

Mitochondrial Data Set

The 13 protein-coding genes (*ND1*, *ND2*, *COI*, *COII*, *ATPase 8*, *ATPase 6*, *COIII*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*, *Cytb*) of the mitogenome of each sample were concatenated and nucleotide and translated amino acid sequences aligned using the Geneious® algorithm, which uses a progressive pairwise method based on a neighbor-joining guide tree. The complementary strand sequence was used for *ND6*. Stop codons were excluded from the amino acid alignment. PAUP* v4.0a152 (Swofford, 2002) was used to determine number of constant, variable, and parsimony informative sites.

Model Choice and Partitioning

The best-fitting partitioning scheme and model of molecular evolution for nucleotide and amino acid alignments was selected in PartitionFinder v.1.1.0 (Lanfear et al., 2012) using the Bayesian information criterion (BIC). Both maximum likelihood and Bayesian inference methods were explored. The greedy algorithm was used, and branch lengths were “linked”.

When branch lengths are ‘linked’, a single set of branch lengths are estimated for the tree, but each partition is given its own rate multiplier, which can shrink or stretch all branch lengths (Lanfear et al., 2012). This allows each partition (e.g., codon position) to have a different overall substitution rate, but assumes that relative rates are constant among lineages (Lanfear et al., 2012). While ‘unlinked’ branch lengths allow for partitions to have different rates among lineages, accounting for more variation, they typically too many parameters to be estimated from the data, and are often not preferred over ‘linked’ schemes (Frandsen et al., 2015). Thus, ‘linked’ branch lengths were chosen here. Partitioning schemes included no partitioning of sites, by codon position, by gene, and by gene + codon position.

Maximum Likelihood Analyses

Maximum likelihood (ML) analyses were conducted in RAxML v.8.0.26 (Stamatakis 2014). Table 3.1 presents a list of all ML analyses conducted along with model choice and partitioning schemes. Briefly, nucleotide alignments used the general time reversible (GTR) substitution model (6 substitution rate parameters, allows for unequal base frequencies) + gamma parameter (Γ), which models substitution rate heterogeneity over sites + invariable sites (I), which is the fraction of sites assumed to be invariable. The amino acid alignments used the MTMAM substitution model (Cao et al., 1998; Yang et al., 1998) + Γ + I and HIVB substitution model (Nickle et al., 2007) + F. The ML tree was estimated under the partitioning and model schemes using 200 or 1,000 inferences. Nonparametric bootstrap support values for nodes were obtained using 1,000 replicates in RAxML. Trees were visualized in FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Table 3.1 Maximum likelihood analyses conducted on the concatenated 13 mitochondrial protein-coding genes data set. All analyses were conducted in RAxML.

Data Set	Model	Partition	Runs	Bootstrap Replicates
Nucleotide	GTR+I+G	None	1000	1000
Nucleotide	GTR+I+G	Codon	1000	1000
Nucleotide	GTR+I+G	Gene	200	1000
Nucleotide	GTR+I+G	Gene + Codon	1000	1000
Amino Acid	MTMAM+I+G	None	1000	1000
Amino Acid	MTMAM+I+G; HIVB+F	Gene	200	1000

Table 3.2 Bayesian inference analyses conducted on the concatenated 13 mitochondrial protein-coding genes data set. All analyses were conducted in MrBayes.

Data Set	Model	Partition	Runs	Generations / run	Sample Frequency
Nucleotide	GTR+I+G	None	2	20,000,000	1000
Nucleotide	GTR+I+G	Codon	2	20,000,000	1000
Nucleotide	GTR+I+G	Gene + Codon	2	20,000,000	1000
Amino Acid	MTMAM+I+G	None	2	15,622,000	500
Amino Acid	MTMAM+I+G; MTMAM+G; JTT+G	Gene	2	15,309,000	500

Bayesian Inference

Bayesian inference (BI) was implemented in the program MrBayes 3.2.6 (Ronquist et al., 2012) within the Cyberinfrastructure for Phylogenetic Research (CIPRES) gateway (Miller et al., 2010) to approximate the posterior probabilities of the phylogenetic trees. Table 3.2 presents a list of all BI analyses conducted on the mitogenome data set. Metropolis-coupled Markov chain Monte Carlo (MC)³ algorithm was used with two parallel runs, each with one cold and seven heated chains. Heated chains are used to help mixing by flattening out the posterior distribution. Thus, the heated chains can move more freely around the parameter space. At intervals, an attempt is made to swap the cold chain and a randomly selected heated chain. If it is accepted, the cold chain can move great distances and increase mixing. The limit of heated chains for use in CIPRES was chosen, as more chains can help improve mixing and convergence, especially with large data sets. Chains ran for approximately 15 million to 20 million generations, depending on data set, with a burn-in of 25%, and sampled every 500th to 2,000th generation. The starting tree was random with arbitrary values for branch lengths and model parameters. Prior distributions for all model parameters were set as the default in MrBayes, which indicated no prior knowledge of parameters. Topology prior was set as uniform distribution; branch lengths set as unconstrained, gamma-dirichlet (1, 0.1, 1, 1); four stationary nucleotide frequencies set as dirichlet (1, 1, 1, 1); six substitution rates set as dirichlet (1, 1, 1, 1, 1, 1); shape parameter set as exponential (1); proportion of invariable sites set at uniform (0, 1). MrBayes was used to summarize samples of model parameters and summarize tree samples. Tracer v1.6 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to visualize (MC)³ output, the parameter values sampled from the chain, check mixing and convergence, and evaluate the burn-in. Samples were summarized using histograms, trace plots, means, and credible intervals to assess mixing and convergence in Tracer. All estimated parameters from each run showed good mixing

and convergence onto their respective posterior distributions as evident by ESS values > 200 , Potential Scale Reduction Factor (PSRF) approaching one, and visualization of trace plots and distributions. A 50% majority rule consensus tree was used to summarize trees in MrBayes. The tree, along with associated posterior probability values and branch lengths was visualized in FigTree.

Divergence Time Estimation

Divergence times of the sampled lineages were estimated using a Bayesian relaxed molecular clock method under an uncorrelated lognormal distribution for lineage-specific rate heterogeneity in the program BEAST v2.4.3 (Bouckaert et al., 2014). The nucleotide data set was partitioned by codon, as described earlier. BEAUti v2.4.3 (Bouckaert et al., 2014) was used to generate input files for BEAST.

Thirteen fossil calibration priors were assigned to appropriate nodes based on confidence of fossil record (Underwood, Pers. Comm.). Fossil calibration ages and prior settings can be found in Table 3.3. Calibration fossils were used to place soft upper and lower bounds on respective nodes, which dates the most recent common ancestor (MRCA) of that lineage.

The best scoring ML tree under a codon partitioning scheme for the nucleotide data set was used to make a starting tree. The R package ‘ape’ v.3.4 (Paradis et al., 2004), using the ‘chronos’ function, was used to create an ultrametric tree from the ML tree. Figtree was used to visualize the branch lengths of the nodes used in fossil calibrations. These branch lengths were used with the Marshall method (Marshall, 2008), to define the calibration lineage and calculate F_{95} maximum bounds for lineages that do not have reliable bounds known from the fossil record. The program Mesquite v3.2 (Maddison & Maddison, 2017) was used to re-scale the branch

Table 3.3 Fossil calibrations used in divergence time estimation. Fossil node number refers to the corresponding node on the divergence time tree where the fossil calibration was placed. Minimum and maximum soft age constraints are given in millions of years (Ma). The exponential prior distribution used in the analysis for each fossil is given by the exponential offset and mean. Min. = Minimum

Fossil	Lineage	Minimum age (Ma)	Maximum age (Ma)	Exponential Offset	Exponential Mean	Min. age reference
1	Chimaeriformes	280	577.2 ^a	280	99	Coates et al. 2017
2	Callorhynchidae	145	256.3 ^a	145	37	Underwood, Pers. Comm.
3	Rhinochimaeridae	152.5	192 ^a	152.5	13.1	Underwood, Pers. Comm.
3	Chimaeridae	112	192 ^a	112	26.7	Nesov & Averainov 1996
67	Lamniformes	135	178.6 ^a	135	14.5	Rees 2005
67	Carcharhiniformes	41	178.6 ^a	41	45.8	Underwood et al. 2011
65	Heterodontiformes	156	230.7 ^a	156	24.8	Underwood 2002
71	Pristiophoriformes	85	191.4 ^a	85	35.3	Cappetta 1980
71	Squaliformes	129	191.4 ^a	129	20.7	Thies 1981
59	Mobulidae	56	180 ^b	56	41.5	Cappetta 2012
59	<i>Rhynchobatus</i>	52	180 ^b	52	42.6	Cappetta 2012
58	<i>Narcine</i>	92	180 ^b	92	29.3	Claeson et al. 2013
57	Rajiformes	72	180 ^b	72	36	Siverson & Cappetta 2001

^a Maximum age calculated using Marshall (2008) method

^b Maximum age based on fossil record of Batoidea, from Jurassic, Toarcian Holzmaden, ~180 Ma (Duffin Pers. Comm.)

lengths on the ultrametric tree so calibration fossils fit appropriately on the tree. Each partition was analyzed under a separate GTR+ Γ +I model, while clock and tree models remained linked, respectively. Substitution rate, shape, proportion of invariant sites, GTR substitution model parameters, and base frequencies were set to "estimate", and the gamma category count set to 4. The relaxed lognormal clock model was selected, and clock rate set to "estimate". The relaxed lognormal clock model was selected, and clock rate set to "estimate". The birth-death model was used for the tree prior to model branching rates on the tree. Tree model priors were set as default: birthRate2.t:tree uniform distribution, initial value 1.0, [0, 1000] lower and upper bound; relativeDeathRate2.t:tree uniform distribution, initial value 0.5, $[-\infty, \infty]$. Default evolutionary model parameter priors were used: gamma shape parameter with an exponential distribution, initial value 1, $[-\infty, \infty]$; proportion invariant sites had a uniform distribution, initial value 0.05, [0, 1.0]; substitution rates had a gamma distribution, initial value 1.0, [0, ∞]. Hyperpriors on the clock model were also set as default: uclMean.c:Clock uniform distribution, initial value 1.0, $[-\infty, \infty]$; uclStdev.c:Clock gamma distribution, initial value 0.1, [0 ∞]. An exponential time to most recent common ancestor (tmrca) prior distribution was set for each of the fossil calibration nodes and monophyly was enforced. A soft minimum age constraint (exponential offset) was used for each prior, and a soft maximum age constraint was used by setting the exponential mean value so that 95% of the distribution lies between the minimum and maximum age constraint. The four batoid calibration fossils have both a reliable minimum and maximum bound set by fossil data. However, the remaining nine fossils only have reliable minimum ages, and thus, the method of Marshall (2008) was used to calculate a soft maximum age.

Four independent Markov Chain Monte Carlo (MCMC) BEAST runs were conducted, each with 200 million generations, sampled every 1,000th generation. Each individual run was checked in Tracer to assess proper mixing and convergence on to the posterior distribution. An

effective sample size (ESS) value > 200 is widely regarded as being accurate for sampling from the posterior distribution for each parameter. Tree and log files from the four independent runs were combined in LogCombiner v2.4.3 (<https://github.com/CompEvol/beast2/releases>), resampling states at a lower frequency of 4000, with a 25% burn-in. The combined log file was visualized in Tracer to assess model parameter values, node-height estimates, summary statistics, and trace files of parameters. TreeAnnotator v.2.4.3 (<https://github.com/CompEvol/beast2/releases>) was used to summarize the posterior probability density of the combined tree file as a maximum clade credibility tree. FigTree was used to visualize the mean and 95% highest posterior density (HPD) limits of node heights (divergence time estimates) and the posterior probabilities of the nodes.

Results and Discussion

Mitochondrial Genome Organization

The modified capture method collected mitochondrial genomes for all chimaeroid and outgroup samples. For the ingroup chimaeroids only, the percentage of reads on target was highly variable between 14.6% and 93.4% (Table 3.4; Figure 3.1). The mean sequence depth ranged from 1,226 to 19,834 (Table 3.4; Figure 3.2). However, it should be noted that the D-loop region is highly variable, and is not considered reliably accurate. The total length of genomes ranged from 16,758 bp to 21,631 bp, with variation likely due to D-loop region (Table 3.4). Previously, only eight chimaera species mitogenomes had been sequenced (*H. lemures*, *C. monstrosa*, *C. fulva*, *H. raleighana*, *R. pacifica*, *C. milii*, *C. capensis*, *C. callorynchus*). This adds 47 new mitogenomes, 31 of which are from described species, and 16 other potentially unique lineages.

Table 3.4 Mitochondrial genome capture statistics.

	Mean	Median	Minimum	Maximum
Total non-duplicate reads	658182	718674	229124	939057
Number of reads on target	306642	288937	78202	582025
Percentage of reads on target (%)	47.3	46.4	14.6	93.4
Mean sequence depth	4711	3881	1226	19834
Total length of mitogenome (bp)	19698	20090	16758	21361

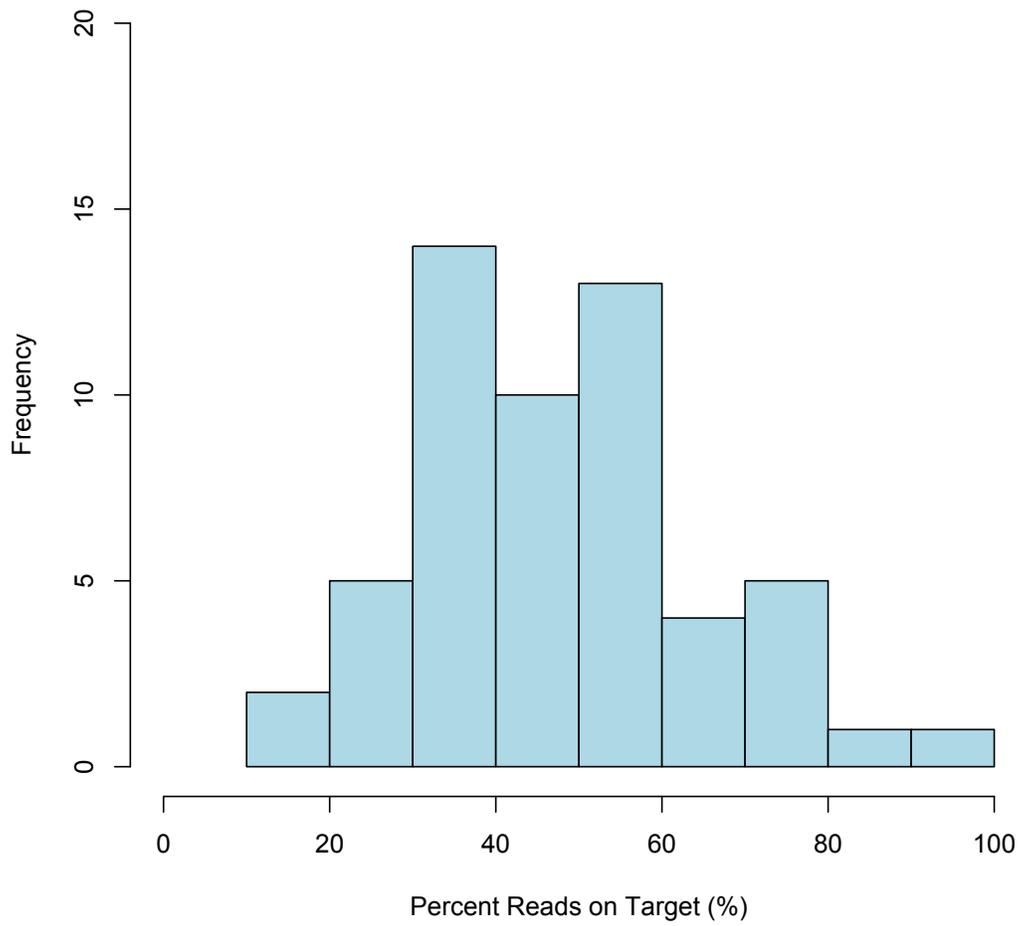


Figure 3.1 Percent of mitochondrial genome capture sequencing reads on target for all chimaeroid lineages sampled.

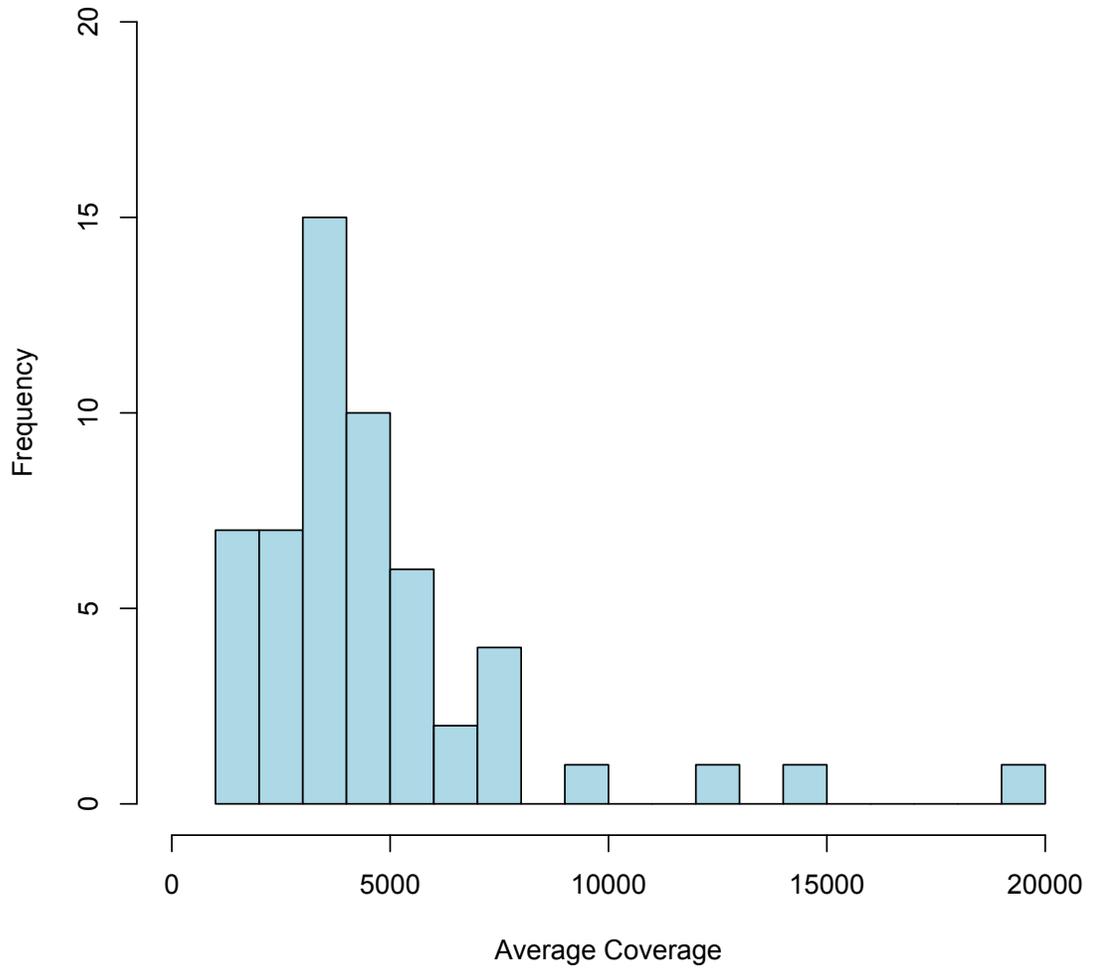


Figure 3.2 Average coverage of the mitochondrial genome capture for all samples chimaera lineages.

The Callorhynchidae mitochondrial genome organization was consistent with the previous *C. milii* mitochondrial genome. The Chimaeridae and Rhinochimaeridae lineages were atypical of vertebrate mitochondrial genome organization, in that a long non-coding region lies between tRNA^{Thr} and tRNA^{Pro}, which can be seen by the much longer genome length. This non-coding region has been identified in the mitogenomes of other members of Chimaeridae and Rhinochimaeridae (Arnason et al., 2001; Inoue et al., 2010). Thus, it is likely a feature found in the ancestor that lead to Rhinochimaeridae and Chimaeridae.

The 13 concatenated protein-coding gene alignment was a total of 11,439 base pairs in length. Table 3.5 contains characteristics of each of the 13 protein-coding genes in the alignment including length, number of constant and variable sites, and the number of parsimony informative sites. The base composition for the entire alignment of holocephalans was A: 28.7%, C: 26.3%, G: 13.4%, and T: 31.7%, with a GC content of 39.5%.

Table 3.5 Characteristics of the 13 protein-coding genes of the mitochondrial alignment used in phylogenetic analyses. From and To indicate base pair position in the alignment, No. = number. PIS = parsimony informative sites.

Gene	From (bp position)	To (bp position)	Total No. bp	No. constant sites	No. variable sites	No. PIS
ND1	1	978	978	466	512	474
ND2	979	2025	1047	394	653	581
COI	2026	3567	1542	891	651	613
COII	3568	4257	690	384	306	277
ATPase8	4258	4425	168	57	111	100
ATPase6	4426	5109	684	293	391	349
COIII	5110	5895	786	416	370	339
ND3	5896	6246	351	148	203	190
ND4L	6247	6543	297	108	189	166
ND4	6544	7923	1380	546	834	746
ND5	7924	9771	1848	763	1085	958
ND6	9772	10296	525	192	333	296
Cytb	10297	11439	1143	520	623	563

Maximum Likelihood Analyses

The resulting topologies for all ML analyses were mainly congruent, with the same overall topology and eight major clades. Differences between topologies were identified in some of the shallower node and species-level relationships. Each of the three families were recovered as monophyletic with strong bootstrap support, with Callorhinchidae basal, and Rhinochimaeridae and Chimaeridae forming a monophyletic group (Figures 3.3-3.5). A strict consensus tree topology of the nucleotide and amino acid no partitioned analysis was generated (Figure 3.6), and will be used below to describe the major details of the phylogenetic relationships among chimaeras.

The family Callorhinchidae was recovered as the basal clade, highly diverged from Rhinochimaeridae and Chimaeridae. This is consistent with previous morphological and molecular work (Didier et al. 1995; Heinicke et al. 2009; Inoue et al. 2010; Licht et al. 2012), and the nuclear data from this study (see Chapter 2). Overall, the relationships among the three species could not be resolved, as the different character sets and partitioning schemes produced differing results. The nucleotide data set analyses resulted in *C. milii* sister to *C. callorynchus* (Figure 3.3). However, the nucleotide codon+gene partition (Figure 3.4) and the amino acid data set (Figure 3.5) resulted in *C. milii* sister to *C. capensis*. There was poor bootstrap support across all topologies for these relationships. Very short terminal branch lengths (Figures 3.3-3.5) indicate little sequence variation between the three recognized species. This was also found in previous molecular work on *Callorhynchus* species (Inoue et al., 2010; Licht et al., 2012).

Rhinochimaeridae, which consists of three genera, was recovered as a monophyletic group, sister clade to Chimaeridae, to the exclusion of Callorhinchidae, like previous studies have concluded (Didier, 1995; Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012). The

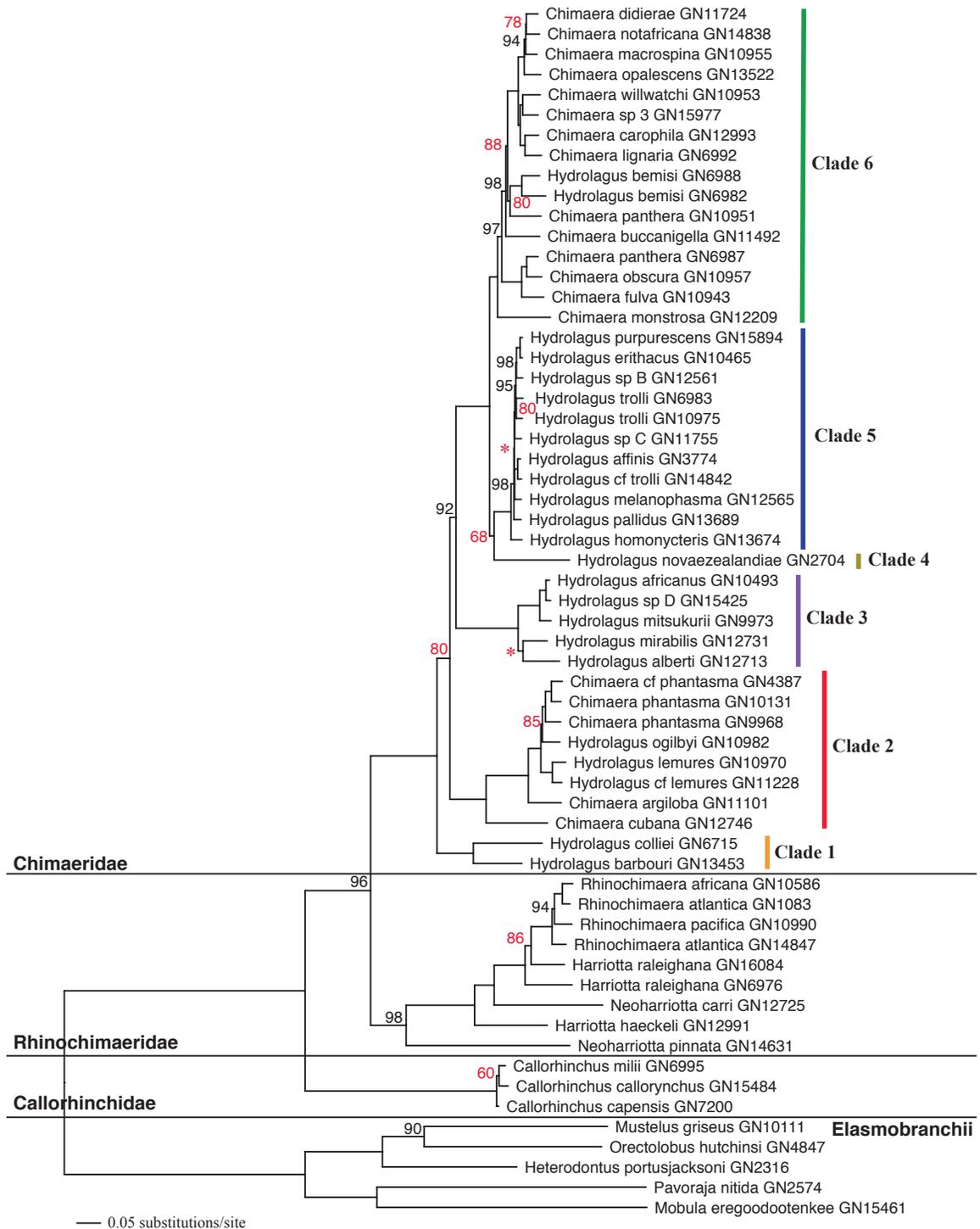


Figure 3.3 Maximum likelihood tree topology of the mitochondrial nucleotide data set partitioned by codon under the GTR+ Γ +I model. Nodes with less than 100% bootstrap support values shown on tree. * = <50% bootstrap support

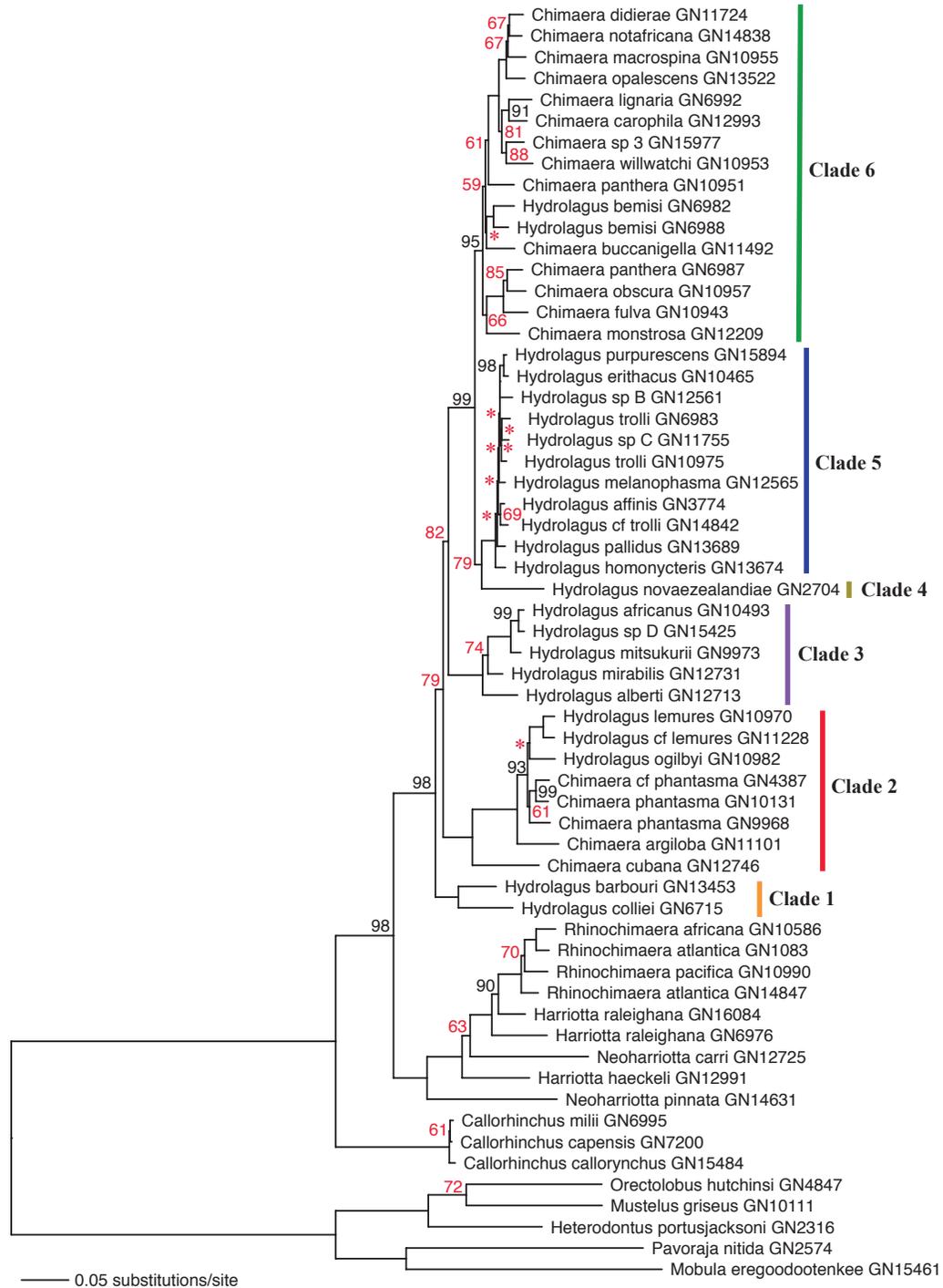


Figure 3.5 Maximum likelihood tree topology of the mitochondrial amino acid data set with no partitioning scheme under the MTMAM+ Γ +I model. Nodes with less than 100% bootstrap support values shown on tree. * = <50% bootstrap support.

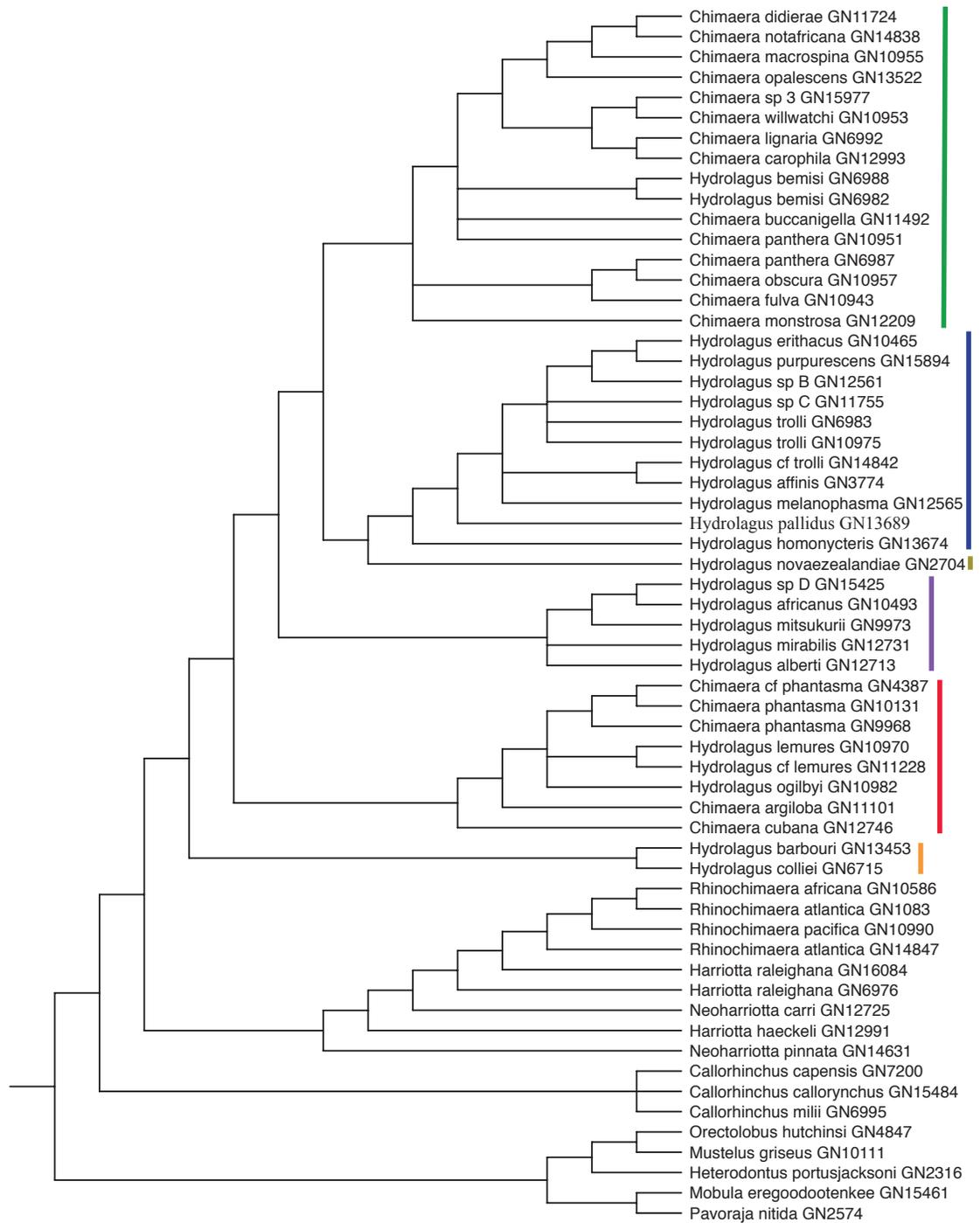


Figure 3.6 Strict consensus tree topology of the mitochondrial nucleotide and amino acid data set with no partitioning scheme.

species-level relationships within this clade were identical in all analyses and showed high bootstrap support (Figures 3.3-3.6). The genus *Rhinochimaera* was recovered as monophyletic, however, neither *Harriotta* nor *Neoharriotta* were supported as monophyletic genera. *Neoharriotta pinnata* was retrieved as the basal lineage of the clade, highly diverged from the other species. *Harriotta haeckeli* represented a unique lineage, as did *N. carri*. *Harriotta raleighana* was represented by two distinct lineages, but not sister taxa. One lineage, a specimen collected from the north Atlantic Ocean, was more closely related to the *Rhinochimaera* lineages, than to the other *H. raleighana* from New Zealand. The two *H. raleighana* lineages appear to be divergent with a high degree of genetic variation between one another, indicating they are likely different species. Within *Rhinochimaera*, two *R. atlantica* species, one from off the east coast of the United States, and the other from South Africa, were retrieved as separate lineages, not more closely related to one another. *Rhinochimaera africana*, collected from off Japan, was found to be most closely related to *R. atlantica* from the Atlantic Ocean (USA). *Rhinochimaera pacifica*, from Australian waters, then clusters with *R. africana* and *R. atlantica* (USA) to the exclusion of *R. atlantica* (South Africa). There are no biogeographical patterns to the relationships within *Rhinochimaera*. Previous molecular work lacked adequate taxon sampling to resolve the relationships within Rhinochimaeridae. However, one study found that *H. raleighana* and *R. pacifica* were sister taxa, to the exclusion of *N. pinnata* (Licht et al. 2012), which is similar to the pattern retrieved in this study. This is the first study to include a majority of the diversity within Rhinochimaeridae.

The family Chimaeridae was recovered as a monophyletic group with high bootstrap support in all analyses (Figures 3.3-3.5). The two genera, *Chimaera* and *Hydrolagus*, were not recovered as monophyletic. All tree topologies resulted in the same major clade relationships, which were highly supported. However, some of the relationships within the six major clades

differed among analyses. The most basal Chimaeridae clade (clade 1) in all analyses, which had 100% bootstrap support, contained two species, *H. colliei* and *H. barbouri*. The split between clade 2 (*C. cf phantasma*, *C. phantasma*, *H. ogilbyi*, *H. lemures*, *H. cf lemures*, *C. argiloba*, *C. cubana*) and the remaining clades had lower support, from 79% to 86% (Figures 3.3-3.5), but was consistent across character sets and partitions (Figure 3.6). Clade 2 was well supported and consisted of both *Chimaera* and *Hydrolagus* species. The majority of species-level relationships were resolved within clade 2 (Figure 3.6), except for the placement of *H. ogilbyi* which differed between the nucleotide and amino acid analyses (Figures 3.3-3.5). The next split between clade 3 (*H. africanus*, *H. sp. D*, *H. mitsukurii*, *H. mirabilis*, *H. alberti*), and the remaining clades had better support in the nucleotide analyses (>90%), but lower support in the amino acid analyses (82%), but was also consistent across resulting topologies. Clade 3 was well supported and consisted of only *Hydrolagus* species, and was mostly consistent across topologies. The placement and relationship between *H. mirabilis* and *H. alberti* differed between nucleotide and amino acid data sets, and was not well supported in any analysis (Figures 3.3-3.6).

Clade 4 contained one species, *H. novaezealandiae*, and was recovered as a highly divergent sister taxon to clade 5. Clade 5 consists of 11 described and unidentified *Hydrolagus* species (*H. homonycteris*, *H. pallidus*, *H. melanophasma*, *H. cf trolli*, *H. affinis*, *Hydrolagus* sp. C, *H. trolli*-GN10975, *H. trolli*-GN6983, *Hydrolagus* sp. B, *H. erithacus*, *H. purpurescens*). The split between these two clades was not well supported in any analysis. While a few of the species-level relationships were consistent across resulting topologies, several of the relationships were not resolved (Figure 3.6), and showed poor support, especially in the amino acid tree topology (Figures 3.3-3.5). Clade 5 resulted in extremely short estimated internal and terminal branch lengths. This would indicate very little genetic variation among the lineages in this clade.

The last Chimaeridae clade 6 (*C. monstrosa*, *C. fulva*, *C. obscura*, *C. panthera*-GN6987, *Chimaera buccanigella*, *C. panthera*-GN10951, *C. lignaria*, *C. carophila*, *Chimaera* sp. 3, *Chimaera willwatchi*, *C. opalescens*, *C. macrospina*, *C. notafriana*, *Chimaera didierae*, *H. bemisi*-GN6988, *H. bemisi*-GN6982) was recovered as sister to clades 4 and 5, was well supported, and included mainly *Chimaera* species, but also *Hydrolagus* (Figures 3.3-3.6). The relationships within the clade were mostly consistent and resolved among resulting topologies (Figure 3.6), except for between nucleotide and amino acid analyses (e.g. *C. monstrosa*, *C. panthera*-GN10951). Terminal branches showed divergence among the lineages, however, some of the internal branches were quite short. In several of these cases, the support for these nodes was low (Figures 3.3-3.5), but the consistence across character sets and partitioning schemes lends support that they are true relationships for this data set.

Partitioning of the data sets did not have any major effects on the results. The analyses using the nucleotide data under no partition, codon partition, and gene partition produced identical trees, with all major clades, and many of the shallower nodes well supported (Figure 3.3; only codon partition tree shown). The codon + gene nucleotide data set differed from the other nucleotide data sets in only the relationships between the three species of *Callorhinchus* (Figure 3.4). The analyses using the amino acid data with no partition and a gene partition produced identical topologies with high bootstrap support for major clades (Figure 3.5; only no partition tree shown).

The differences among topologies was mainly evident between the two character types (nucleotide and amino acid). While the major clades and their relationships were identical, the species-level relationships within Callorhinchidae and the some of the Chimaeridae clades did differ (clades 2, 3, 5, 6). Also, several species relationships at the tips of the tree saw less bootstrap support in the amino acid analyses than in the nucleotide analyses. This may be a

function of a reduced number of character sites in the amino acid data set. Overall, the ML nucleotide and amino acid analyses shared the same major clades, however, some of the species-level relationships within a clade differed (Figure 3.6).

Current nucleotide substitution models assume a stationary process, including homogeneity of nucleotide composition among lineages in a phylogeny (Felsenstein, 1988; Nabholz et al., 2011; Sheffield, 2013; Betancur-R et al., 2013). It assumes that across the tree, base composition is constant, and that all branches and ancestral sequences have the same equilibrium base composition designated in the substitution rate model (Galtier & Gouy, 1998). However, this assumption of stationarity in nucleotide frequencies across taxa is often violated. In fact, mitochondrial DNA has been shown to have nucleotide composition heterogeneity (Meyer, 1994; Ballard & Whitlock, 2004; Rubinoﬀ & Holland, 2005). When heterogeneity in nucleotide frequencies across lineages is present, but not accounted for in phylogenetic reconstruction, this can lead to model mis-specification and mislead phylogenetic inference (Galtier & Gouy, 1995; van den Bussche et al., 1998; Foster & Hickey, 1999; Foster, 2004; Ho & Jermiin, 2004; Nabholz et al., 2011). One such affect that has been shown is the grouping of unrelated lineages with similar base compositions, despite their evolutionary relationships (Lockhart et al., 1994; Galtier & Gouy, 1995; Phillips et al., 2004; Delsuc et al. 2005; Blanquart & Lartillot, 2008; Sheffield et al., 2009; Som, 2014). However, some studies have indicated that base composition bias may not affect phylogenetic methods, where the signal is strong enough to overcome violation of stationarity, and that only extreme bias may mislead inference (Conant & Lewis, 2001; Rosenberg & Kumar, 2003). Nucleotide characters compared to translated amino acids, in general, show more variation and substitution saturation. Heterogeneity in base composition is most often found at 3rd codon positions, because nucleotide substitutions at this position are usually synonymous (Meyer, 1994). Those sites that are most saturated (3rd positions)

are more biased (Rodriguez-Ezpeleta et al., 2007), which can affect substitution rate estimation and the subsequent phylogeny (Galtier & Gouy, 1995). This may be an explanation for the differences observed between nucleotide and amino acid data sets. A stationary process is assumed by the nucleotide substitution model used here, however, the process is actually non-stationary, resulting in model mis-specification. On the other hand, the amino acid data set is not affected by this base composition bias, and thus, different relationships are resolved. Nonetheless, it is possible that codon bias is present within amino acid data. Only minor differences were observed between the data sets, at some shallow internodes and terminal nodes. It is possible that base composition bias in the nucleotide data grouped some species within a clade together due to similarities in their composition, while slightly different relationships were observed in the amino acid data. But, given that major differences were not observed between the data sets, and no erroneous placement of taxa was inferred, it is likely that base composition bias is not having a significant effect on the inference. However, a future direction could be to explore base composition heterogeneity within the data set, and explore approaches to minimize bias and model non-stationarity to infer the phylogeny. This can then be compared to the resulting topologies in this study to determine whether a bias may be affecting the phylogenetic reconstruction.

Bayesian Inference Analyses

Tree topologies recovered from BI analyses were mainly congruent between character set and partitioning schemes, with the same major clades and species within (Figures 3.7-3.9). The most basal and monophyletic clade, Callorhinchidae, was well supported with a posterior probability of one in all analyses. Species-level relationships within Callorhinchidae differed among data set and partition scheme, and showed lower posterior probabilities for these

relationships. Short terminal branch lengths were also estimated by the BI analyses. This pattern was also evident in the ML analyses, with unresolved species relationships, low reliability, and little genetic variation among species. The Rhinochimaeridae clade was monophyletic, sister to the Chimaeridae clade, well supported, and identical across all BI analyses. The species-level relationships also were identical to those in the ML analyses. The Chimaeridae clade was highly supported and consisted of six well supported major clades, as seen in the ML analyses. Most of the species-level relationships were highly supported with a posterior probability of one, but some of the relationships had lower probabilities. Mainly those lineages that showed different topologies between nucleotide and amino acid analyses had lower support, and similar to the differences seen between character sets in the ML analyses. The relationships between the six major Chimaeridae clades were identical across BI analyses, with differences among analyses coming from within clade relationships. Overall, the branch lengths were similar between BI analyses and the ML analyses.

The topology of the nucleotide data with no partition and partitioned by codon was identical except for the relationships among the three *Callorhinchus* species (Figure 3.7; only codon partition tree shown). The no partitioned tree topology compared to the codon + gene partitioned topology differed in the species-level relationships among the large *Hydrolagus* clade 5 (Figure 3.8). The Bayesian inferences based on the amino acid data produced identical results for no partition and a gene partitioning scheme (Figure 3.9; only no partition tree shown). The major clades and the species that they contain remained consistent between nucleotide and amino acid analyses, however, species-level relationships differed, similar to the ML analyses.

Comparisons of ML and BI topologies for the same partitioning schemes revealed identical or very similar topologies. Codon partitioning resulted in the same topology and both amino acid partitioning schemes had the same tree topologies and similar branch lengths between

methods. No partitioning of the nucleotide data between methods was similar except for the relationships between the *Callorhinchus* species, and codon + gene partitioning showed differences among the clade with very similar *Hydrolagus* species (clade 5). The congruence of the major clades and species relationships between ML and BI analyses provide additional confidence that these are true relationships for this data set.

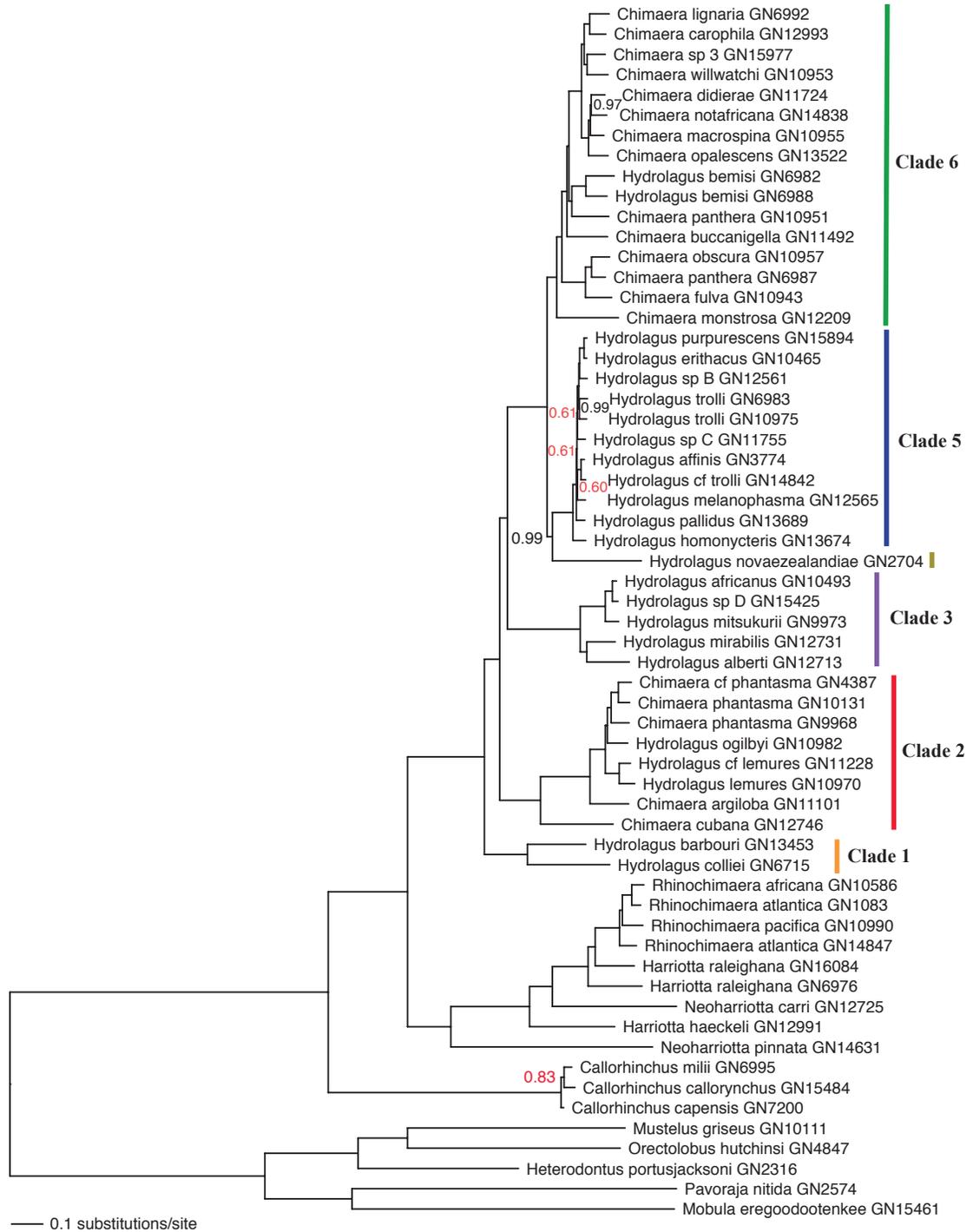


Figure 3.7 The Bayesian inference majority rule consensus tree topology of the mitochondrial nucleotide data set partitioned by codon under a GTR+ Γ +I model. Posterior probabilities < 1 are shown on the tree.

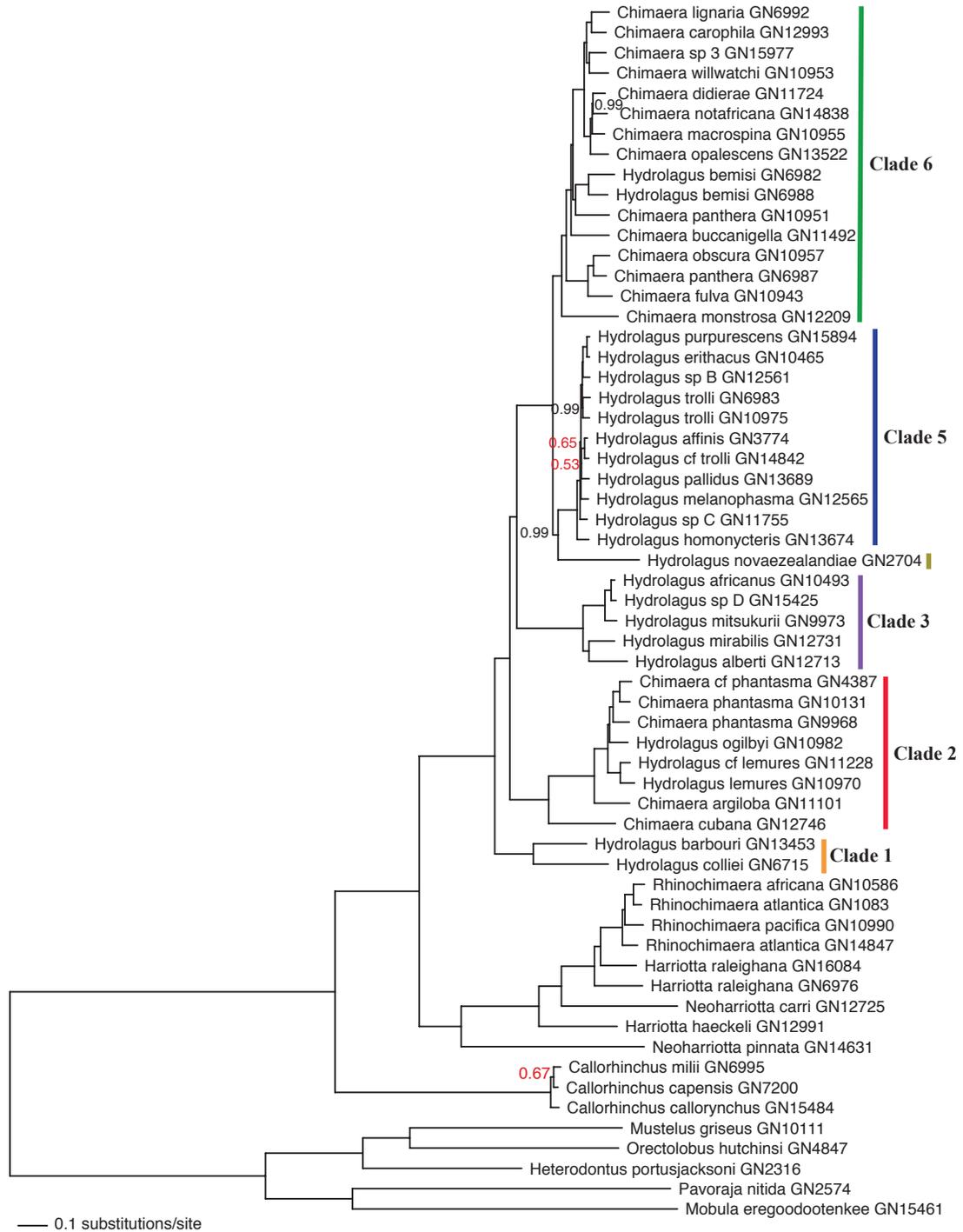


Figure 3.8 The Bayesian inference majority rule consensus tree topology of the mitochondrial nucleotide data set with a codon + gene partition under a GTR+ Γ +I model. Posterior probabilities < 1 are shown on the tree.

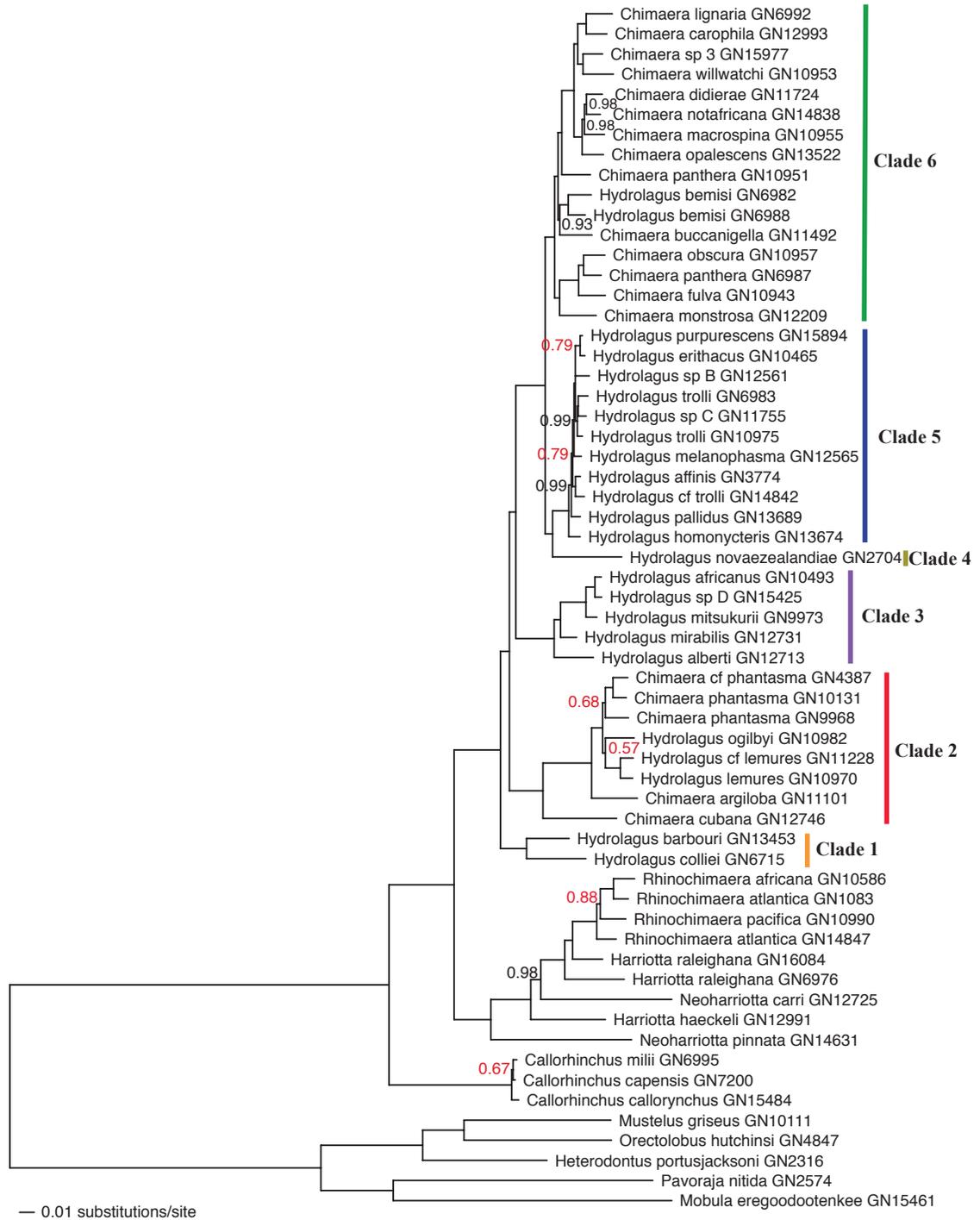


Figure 3.9 The Bayesian inference majority rule consensus tree topology of the mitochondrial amino acid data set with no partitioning scheme under a MTMAM+ Γ +I model. Posterior probabilities < 1 are shown on the tree.

Divergence Time Estimation

Divergence time dating of holocephalans based on mitochondrial markers revealed that chimaeras diverged from a common ancestor with elasmobranchs in the Paleozoic era, as considered by others (Didier, 1995; Pradel et al., 2009; Grogan et al., 2012). A long period of time past before the modern Callorhynchidae lineage diverged from a common ancestor with Rhinochimaeridae and Chimaeridae in the late Triassic to early Jurassic periods of the Mesozoic. The divergence of the families Rhinochimaeridae and Chimaeridae occurred in the late Jurassic period of the Mesozoic when they last shared a common ancestor. Modern day lineages of all families diverged to some extent in the Cretaceous of the Mesozoic era, but mainly relatively recently in the Cenozoic era.

The resulting tree topology from the BEAST analysis (Figure 3.10) was almost identical to the BI and ML analysis under a codon partitioning scheme. The only difference was found in the species-level relationships among the three species of *Callorhynchus*. The BEAST topology was identical to the BI analysis with no partitioning scheme. The majority of node splits had a posterior probability of one, with only a few splits showing a lower probability of <90% within clade 5.

Divergence time estimates are given as a range, equivalent to the 95% credible interval, and posterior means provided as a point estimate. The extant holocephalans diverged from a common ancestor with the elasmobranchs between 329.47 Ma in the Carboniferous and 463.06 Ma in the middle to late Ordovician, with a posterior mean of 395.22 Ma (middle Devonian), all within the Paleozoic (Figure 3.10). Previous estimates of the divergence between holocephalans and elasmobranchs placed the date around 410 Ma to 494 Ma (Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012; Renz et al., 2013). Fossil evidence suggests that the two groups had

diverged by 410 Ma (Coates & Sequiera, 2001). The point estimate in this study is slightly younger than previous estimates and fossil information, however, the credible range includes the other estimated divergence dates, fossil evidence, and is located within the Paleozoic. These previous estimates used a hard minimum bound of 410 Ma. This deep node had a wide credible interval, which indicates a large degree of uncertainty around the estimate. First, deeper nodes can be affected more by substitution saturation (Schwartz & Mueller, 2010; Lukoschek et al., 2012). When rates of saturation are not correctly modeled, this can bias estimates, particularly in mitochondrial genes that may have a relatively fast substitution rate (Yang, 1996; Xia et al., 2003; Brandley et al., 2011). Also, the calibration fossil for this node had quite a broad bound, 280 to 577.2 Ma. This wide bound indicates uncertainty, and it has been considered that overly broad priors may result in excessively large uncertainty in estimates (Saladin et al., 2017). This is likely contributing to the large variance in the age estimate at this node. Divergence time estimates can be very sensitive to prior distributions (Ho & Phillips, 2009). Thus, this prior distribution may be leading to the large uncertainty estimated for this divergence age. Another potential explanation for the large variance is the data violate one or more of the assumptions (e.g., stationarity of base frequencies) of the model of molecular evolution, which could lead to a greater uncertainty in the estimate. It may be the case that there is base composition bias between holocephalans and elasmobranchs, which would not be modeled appropriately, leading to a large credible interval on the divergence between these two groups.

Callorhynchidae diverged from the lineage that lead to Rhinochimaeridae and Chimaeridae in the Mesozoic era, between 179.47 Ma (early Jurassic) and 226.98 Ma (late Triassic), with a mean of 202.07 Ma (late Triassic; Figure 3.10). The three species of *Callorhynchus* are quite young, diverging from one another relatively recently in the late Miocene to early Pliocene of the Cenozoic era. Previous divergence dating has estimated the

Callorhinchus species to be quite young as well (Inoue et al., 2010; Licht et al., 2012). Previous estimates had a credible range from 125 Ma to 320 Ma (Heinicke et al., 2009; Inoue et al., 2010; Licht et al., 2012). The mean and credible interval estimated in this study is very similar to the previous point estimates, and fall within their credible range.

Rhinochimaeridae and Chimaeridae diverged from a common ancestor between 152.5 Ma to 166.22 Ma in the late to middle Jurassic of the Mesozoic era (157.25 Ma; Figure 3.10). This estimate is similar to Licht et al. (2012) estimate of 164 Ma to 156 Ma (159 Ma), but slightly older than others at 122 Ma (146-98 Ma; Inoue et al. 2010), and 107 Ma (182-51 Ma; Heinicke et al., 2009). This node had two fossil calibrations based on information for both Rhinochimaeridae and Chimaeridae that placed a soft bound of 112 Ma to 192 Ma, which may have attributed to an older estimate here compared to the other two studies with younger estimates. *Neoharriotta pinnata* and *H. haeckeli* diverged from the other rhionchimaerids in the early to late Cretaceous of the Mesozoic (Figure 3.10). The remaining lineages diverged during the Cenozoic era. The genus *Rhinochimaera* is quite young, diverging in the Oligocene to Eocene.

Extant diversity of Chimaeridae share a common ancestor in the mid- to late Cretaceous of the Mesozoic era, with modern day lineages quite young, diverging in the Cenozoic (Figure 3.10). Previous estimates of divergence times within Chimaeridae only included a few species (Inoue et al., 2010; Licht et al., 2012), making it difficult to compare with this study, but in general were relatively similar, with divergences in the late Cretaceous to Cenozoic.

The divergence between sharks and batoids was estimated at 221 Ma to 285 Ma (252 Ma) in the late Paleozoic to early Mesozoic eras (Figure 3.10). A study by Klug (2010) examined the extinct †Synechodontiformes, and concluded that they were a monophyletic sister group to modern sharks. This puts the origin of modern neoselachians at ~ 250 Ma. The estimate here is in concordance with the age of the stem-group neoselachians, but may be slightly older than fossil

record for definite presence of neoselachians in the middle Triassic (Cuny et al., 2001). The remaining node divergence time estimates within sharks and rays were mainly congruent with fossil evidence and prior information. Divergence age estimates for all lineage splits are given in Table 3.6.

The ancestral lineage of modern day holocephalans originated in the Paleozoic era, with a point estimate in the Devonian, and a credible range from the Ordovician to Carboniferous. This ancestral lineage continued for a longer period of time, surviving the Carboniferous extinction event, and Permian-Triassic extinction event. The ancestor leading to the three modern day families diverged in the late Triassic to early Jurassic period. These lineages also likely had to survive the mass extinction event of the Triassic-Jurassic boundary. The Callorhynchidae lineage, which diverged in Triassic-Jurassic period, has seen a long period of time without leaving extant taxa, until the late Miocene to early Pliocene of the Cenozoic, where the three known species diverged. Rhinochimaeridae and Chimaeridae diverged in the middle to late Jurassic. Extant taxa of Rhinochimaeridae began diverging in the early to late Cretaceous of the Mesozoic era, with the majority of lineages in diverging in the Cenozoic era. The *Rhinochimaera* is the youngest genus, diverging in the Oligocene to Eocene. Extant diversity of Chimaeridae share a common ancestor in the middle to late Cretaceous, with modern day lineages diverging within the Cenozoic era. This historical biogeography for this group was detailed in chapter 2, and because of the overall congruence in the phylogenetic reconstruction between the mitochondrial and nuclear data sets and divergence time estimates, the same overall patterns are evident for the mitochondrial tree. The radiation of modern day families and species likely came from an ancestor that resided in the Southern Ocean region, with subsequent migrations outward and upward to the Atlantic, Pacific and Indian oceans over time.

Table 3.6 Divergence ages of nodes estimated from the mitochondrial data set in BEAST. Node ages are represented as both a point estimate (posterior mean) and a range (95% credible interval). Ages are in millions of years (Ma). Nodes numbers are shown in Figure 3.10.

Node	Posterior Mean (Ma)	95% Credible Interval	
		Lower (Ma)	Upper (Ma)
1	395.22	329.47	463.06
2	202.07	179.47	226.98
3	157.25	152.50	166.22
4	6.34	4.44	8.39
5	4.65	3.01	6.41
6	127.64	112.42	142.63
7	77.38	63.44	91.39
8	62.05	49.90	74.95
9	36.73	29.10	44.71
10	31.06	24.05	38.36
11	15.22	11.76	18.89
12	13.21	9.90	16.58
13	7.01	4.75	9.40
14	92.27	78.48	106.64
15	53.92	38.15	70.37
16	81.95	69.22	94.84
17	56.11	44.90	67.45
18	27.96	22.19	33.96
19	17.71	14.40	21.17
20	8.88	6.12	11.74
21	16.17	12.94	19.44
22	13.32	10.38	16.42
23	8.25	5.81	10.89
24	75.76	63.49	88.56
25	27.83	21.25	34.76
26	23.09	16.57	29.84
27	9.03	6.50	11.85
28	3.58	2.28	4.95
29	42.93	36.76	49.41
30	37.90	31.43	44.51
31	10.06	7.93	12.33
32	7.20	5.88	8.56
33	6.69	5.48	7.94
34	6.25	5.02	7.52
35	3.06	1.99	4.22
36	6.33	5.15	7.54
37	5.44	4.36	6.55
38	4.90	3.78	6.01
39	4.85	3.79	5.95
40	2.02	1.30	2.79

41	36.35	31.00	42.03
42	32.63	27.80	37.85
43	16.37	12.35	20.55
44	11.64	8.02	15.40
45	28.91	24.44	33.61
46	27.27	22.95	31.83
47	25.01	20.57	29.42
48	15.82	11.50	20.24
49	17.98	14.95	21.15
50	12.52	9.90	15.28
51	11.00	8.41	13.57
52	10.27	7.71	12.86
53	15.99	13.07	19.06
54	13.62	10.53	16.62
55	11.87	8.93	14.94
56	252.62	221.89	285.41
57	202.43	173.54	232.33
58	187.55	159.09	216.39
59	156.65	127.30	186.53
60	42.68	28.82	57.25
61	16.58	10.69	22.91
62	113.43	91.05	136.93
63	98.13	69.16	127.96
64	193.14	176.97	210.37
65	174.06	160.75	188.16
66	156.22	146.09	167.20
67	138.37	135.00	144.89
68	68.55	48.07	89.74
69	66.94	48.98	84.87
70	36.51	23.62	50.67
71	138.55	129.00	153.07
72	104.34	82.90	125.98
73	45.53	30.76	61.22

Comparisons with Nuclear Phylogenetic Analyses

The overall tree topologies recovered for the various analyses across both mitochondrial and nuclear data sets were broadly similar, with the majority of differences lying in the species-level relationships within the major clades. Three monophyletic families of holocephalans were highly supported in both character sets. Callorhinchidae was recovered as the basal clade, and the relationships of the three species were identical and well supported in all nuclear analyses, but were not well supported in the mitochondrial analyses. However, in the strict consensus tree between the ML mitochondrial codon partitioning tree and ML full nuclear codon partitioning tree, the relationships within Callorhinchidae were identical (Figure 3.11). The families Rhinochimaeridae and Chimaeridae were highly supported as each monophyletic and sister clades, to the exclusion of Callorhinchidae. Within Rhinochimaeridae, both nuclear and mitochondrial character sets indicated that *Neoharriotta* and *Harriotta* were not monophyletic genera, while *Rhinochimaera* appears to be monophyletic. Relationships within this clade were consistent in the mitochondrial data, but the relationships within *Rhinochimaera* and between the two *H. raleighana* lineages differed between mitochondrial and nuclear topologies, as indicated by the polytomies in the consensus trees (Figure 3.11). Within the family Chimaeridae, there were six consistent clades between mitochondrial and nuclear character sets, however, the placement of *H. novaezealandiae* differed between the two. It was recovered as a sister lineage to clades 5 and 6 in many of the nuclear tree topologies, whereas in other nuclear and all mitochondrial topologies it was placed as the sister lineage to clade 5, to the exclusion of clade 6. The polytomy present for the placement of the clades 2 and 3 shows the disagreement in these clade relationships between the nuclear and mitochondrial data (Figure 3.11). Many of the species-level relationships were inconsistent between character sets. However, all the species within a clade remained the same. Both nuclear and mitochondrial tree topologies supported that

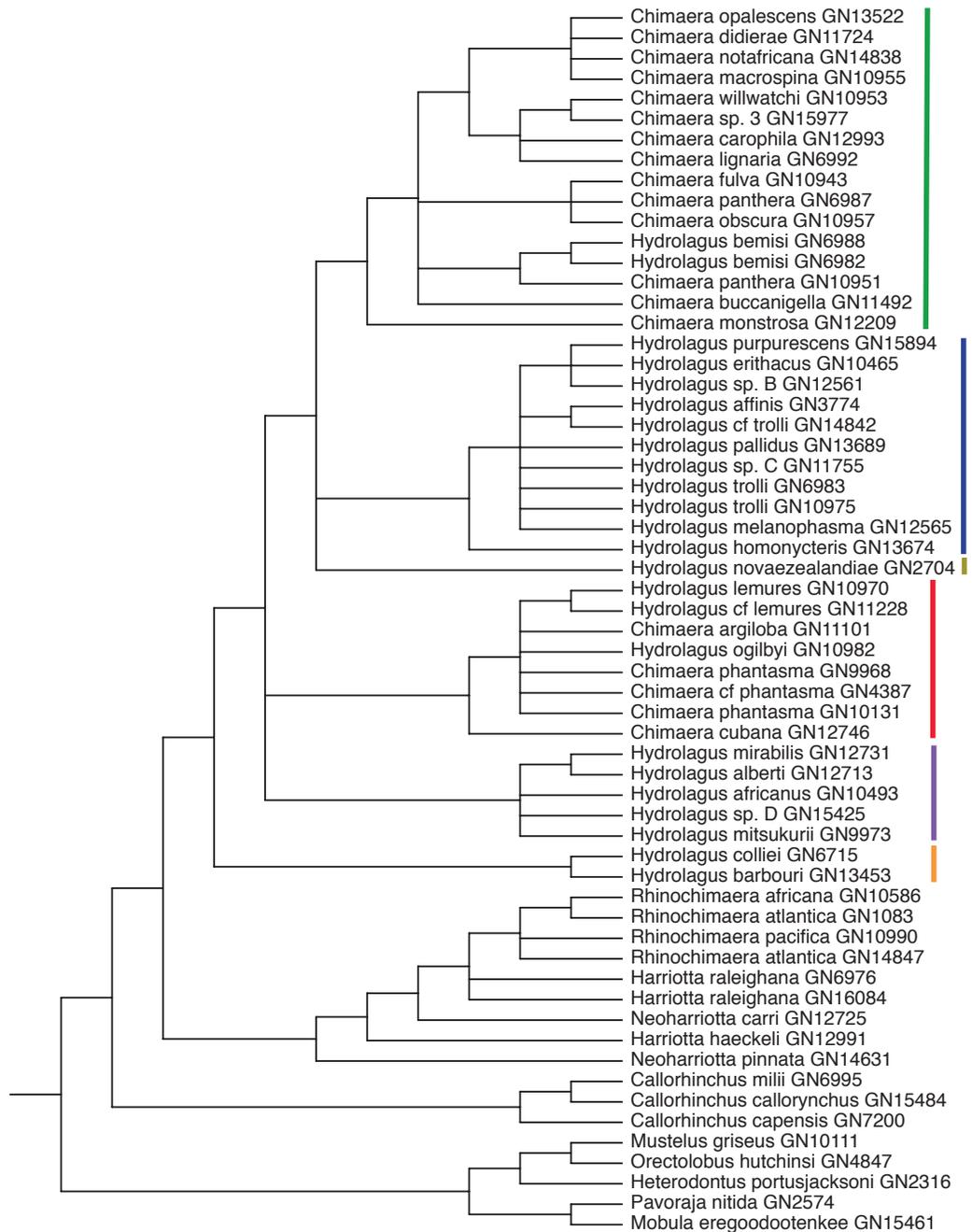


Figure 3.11 Strict consensus tree topology of the nucleotide mitochondrial data set and full nuclear data set partitioned by codon.

Hydrolagus and *Chimaera* are not monophyletic genera. Incomplete lineage sorting is one of the major sources of error in tree reconstruction that can cause incongruence in phylogenies, especially with more recent divergences (Maddison & Knowles, 2006; Philippe et al., 2011). Lineage sorting of ancestral polymorphisms with differential retention of those alleles can produce a gene phylogeny that differs from true species phylogeny (Page & Holmes, 1998; Pamilo & Nei, 1988). This may be one major source of incongruence between the nuclear and mitochondrial data, particularly at shallower nodes.

The estimated branch lengths between mitochondrial and nuclear tree topologies differed, with the mitochondrial topologies typically having longer branch lengths, while very short branch lengths were estimated for many of the node splits within the nuclear topologies. This may be due to differences in substitution rates between the slow-evolving nuclear exons and the mitochondrial protein-coding genes used in the analyses. Mitochondrial DNA is typically considered to have a faster substitution rate compared to nuclear DNA due to its smaller effective population size, which would fix alleles in the population quicker (Ballard & Whitlock, 2004; Rubinoﬀ & Holland, 2005; Galtier et al., 2009), as well as less efficient DNA repair machinery and exposure to oxidative stress (Song et al., 2005; Tuppen et al., 2010).

Divergence time dating of the major splits for holocephalans were congruent between nuclear and mitochondrial character sets. The MRCA of the split between holocephalans and elasmobranchs had a point estimate of 395.22 Ma in the mitochondrial data and 395.83 Ma in the nuclear data, with credible intervals of 329.47-463.06 Ma and 288.45-510.79 Ma, respectively. Both data sets included the same soft minimum age bound, but differed in their maximum age bound for the divergence between holocephalans and elasmobranchs. These large credible intervals were estimated at a deep node in the phylogeny. Deeper nodes can be more difficult to resolve, which may introduce error, and result in a greater variance in the estimate. Also, deeper

nodes are more affected by homoplasy, which could mislead the phylogenetic method. Rate variation among lineages, particularly chondrichthyans and bony vertebrates in the nuclear data, may be present, which may affect the estimation method. Thus, error associated with model misspecification in both analyses could result in the uncertainty in this node estimate, especially because it is a deep node. In both analyses, wide constraint times were used, which translated to a broader prior distribution than a more constrained bound. Thus, this arbitrary prior distribution may not be quite appropriate, resulting in a larger degree of uncertainty in the estimate. An overly broad prior distribution may produce large uncertainty in time estimates (Saladin et al., 2017). The split of Callorhinchidae from the lineage leading to Rhinochimaeridae and Chimaeridae had estimates of 202.07 Ma and 203.86 Ma for mitochondrial and nuclear data, respectively, and very similar credible intervals. This node also had a fossil calibration, but the prior information differed between the data sets. The age of the MRCA of Rhinochimaeridae and Chimaeridae was 157.25 Ma and 157.85 Ma, with similar credible intervals, for mitochondrial and nuclear data, respectively. Again, this node had the same lower bounds, but different upper bounds for the two analyses. Thus, a similar, but not identical prior distribution was set for the different data sets. It is possible that the prior distributions are constraining the estimations in a similar manner, leading to similar age estimates. The extreme congruence in these estimates between the independent data sets lends support that these are consistent estimates for these divergences in holocephalans. Both data sets estimated very recent divergences of the three species of Callorhinchidae in the Cenozoic. Differences between data sets lied in the estimates of the divergences within Rhinochimaeridae and Chimaeridae. The mitochondrial data set produced older estimates for some of the splits within these families. For example, the MRCA for Chimaeridae had an estimated age in the Cretaceous (92.27 Ma) for mitochondrial data set, while the age was in the Cenozoic (21.27 Ma) for the nuclear data set. Similarly, in the

Rhinochimaeridae, all lineage divergences occurred recently in the Cenozoic for the nuclear data set, while some of the divergences occurred in the Cretaceous of the Mesozoic era in the mitochondrial data set. There were no reliable fossil calibrations that could be used for these shallower internodes. Also, the low sequence variation, especially as seen in the nuclear data, between closely related lineages could make inferring these divergence times with reliability difficult.

Overall, the high degree of congruence between the mitochondrial and nuclear gene trees with respect to the major clades, provides confidence that these are likely true relationships for the group, and the different data sets are reflective of the same evolutionary history. The low resolution at shallower internal and terminal nodes is the result of the different relationships recovered by the two independent data sets. These relationships could not be fully resolved based on the mitochondrial and nuclear data sets.

Taxonomic Implications

The family Callorhynchidae contains three recognized species, *C. milii* from New Zealand and Australia, *C. capensis* from South Africa, and *C. callorhynchus* from South America. There are no morphological characters that have been found to distinguish the species apart. The color pattern was put forward as a character for distinguishing species, however, coloration can be variable within a species, and is not a good character to base species delimitation. Currently, the geographic locality of each species is used to separate them as unique species. These three species were described over 150 years ago, before extensive morphological comparisons among similar specimens was conducted in order to appropriately describe a new species. Based on the morphological similarity of these species, some authors have proposed that the three species may represent one species (Bigelow & Schroeder, 1953; Krefft, 1990). However, differences in the

egg cases could imply unique species (Didier et al., 2012). The three species were found to be very closely related as shown by the very short terminal branch lengths in both nuclear and mitochondrial data sets, which implies low sequence variation. With the low genetic variation and no consistent morphological characters to differentiate species, it could be hypothesized that the three species represent one species with population structure captured in the data based on locality. This shows the importance of diligently examining specimens that are similar in morphology when investigating a potentially new species. Species designations based on geographic location alone is not appropriate for describing new species. It is suggested here that these three species may represent one wide-ranging species in the southern hemisphere. However, morphological and molecular data should be collected on multiple specimens of each of these species. In the case that they would become one species, they would take the species name *C. callorynchus*, synonymizing *C. milii* and *C. capensis*.

The rhinochimaerids are a family characterized by an elongate, fleshy snout. It contains three genera and eight recognized species, which occur in the Atlantic, Indian, and Pacific Oceans. Morphologically, the genus *Rhinochimeara* has smooth tooth plates, lacking hypermineralized tissue, and mature males develop tubercles on the dorsal caudal fin (Didier, 1995). The genera *Harriotta* and *Neoharriotta* both have tooth plates with raised hypermineralized tissue and lack tubercles (Didier, 1995). The major differences between these two genera is the presence of an anal fin in *Neoharriotta*, and absence in *Harriotta*. While Didier (1995) placed these genera into two groups, the Rhinochimaerinae with *Rhinochimaeridae*, and Harriottinae with *Harriotta* and *Neoharriotta*, she stated that there were no synapomorphies present to join these two genera, that were not present in *Rhinochimaera*. This study sampled seven of the eight recognized species. In all analyses, *Rhinochimaera* was recovered as a monophyletic genus. Two separate lineages identified as *R. atlantica* were sampled, and in no analysis, were these two lineages recovered as

most closely related. In fact, *R. atlantica* sampled from the Atlantic Ocean, off the USA was most closely related to *R. africana* sampled from Japan. Whereas the *R. atlantica* sampled from South Africa relationship within the group differed based on analysis. Branch length estimates within *Rhinochimaera* were very short, especially in the nuclear data, and to a degree in the mitochondrial data. This could be an indication of population variation instead of species-level variation. On the other hand, if these species are valid, it would indicate that a unique, currently undescribed lineage is present. While studies have indicated morphological differences between the three species (e.g., Compagno et al., 1990), the low sequence variation, and two *R. atlantica* lineages not more closely related to one another, raises questions about the relationships and validity of species. It is recommended that an extensive morphological examination of *Rhinochimaera* specimens be undertaken, along with molecular data analysis, in order to resolve species boundaries and relationships within this genus.

The two genera *Neoharriotta* and *Harriotta* were not recovered as monophyletic. Two of the three *Neoharriotta* species were sampled, missing *N. pumila*. *Neoharriotta pinnata* was recovered as a unique and divergent lineage in all analyses. Since it is the first recognized species in the genus, it would retain its genus and species name. *Neoharriotta carri* was also found to be a unique, divergent lineage, and thus, considered a valid species. However, it is suggested that *N. carri* may need to be placed into its own genus, as molecular evidence suggests it is not most closely related to the other member of the genus, *N. pinnata*. *Harriotta haeckeli* was recovered as a unique species, divergent from other lineages. However, it too is not more closely related to other members of *Harriotta*. Since the genus *Harriotta* was first described under *H. raleighana* in 1895, *H. haeckeli* also would need to be placed into a new genus. *Harriotta raleighana* was represented by two sampled lineages, one from the Atlantic Ocean (Scotland), the other from New Zealand. In the nuclear data sets, these two lineages are sister to one another, and more

closely related to *Rhinochimaera* than other *Harriotta* or *Neoharriotta*. In the mitochondrial data set, the two lineages are not sister, in fact, *H. raleighana* from the Atlantic is sister to the *Rhinochimaera* clade. The branch length in the nuclear data show quite low estimated sequence variation between these two species, whereas in the mitochondrial data set, these two lineages show a high degree of variation, divergent from one another and the *Rhinochimaera* clade. It is suggested here that these two lineages likely represent unique species. In that case, one of the lineages would retain the *H. raleighana* species name, mostly likely the lineage collected from the Atlantic Ocean, where the original specimen used in the species description was collected. The other lineage would require a new species description and name, and could potentially require a new genus. A morphological investigation along with additional molecular data of *H. raleighana* specimens collected from throughout its range (worldwide) is prudent to determine if multiple unique species are present, examine population structure, and determine their relationships to other members of the family. The anal fin character that is present in *Neoharriotta* but absent in *Harriotta* and *Rhinochimaera* appears to be a plastic trait. If it was the ancestral trait in this group, then the trait would have been lost in the *H. haeckeli* lineage, and also lost in the ancestor of *Rhinochimaera* and *H. raleighana*. If it was not the ancestral condition in the group, the two *Neoharriotta* lineages would have independently gained the anal fin.

The family Chimaeridae has historically been represented by two genera, *Chimaera* and *Hydrolagus*. In *Chimaera*, there is the presence of an anal fin web separated from the caudal fin by a small notch, which lacks cartilaginous support (Didier, 1995). The small anal fin separated from the caudal fin is absent in *Hydrolagus*. This is the only morphological character that separates species into the two genera. All analyses across independent data sets revealed that the two genera are paraphyletic. Similar to the anal fin character of *Neoharriotta*, the anal fin seems to be a plastic trait in Chimaeridae, with several instances of *Hydrolagus* lineages within a group

of *Chimaera* lineages. Since this is the only character separating the two genera, and molecular data does not provide evidence that the character represents two unique groups, it is suggested that species of Chimaeridae belong to one genus, *Chimaera*.

Six distinct clades were recovered within Chimaeridae across both nuclear and mitochondrial data sets and analyses. Clade 1 contained two species, *H. colliei* and *H. barbouri*, congruent in all topologies, divergent from one another. Both species have a patterned coloration and reside in the north Pacific Ocean, *H. colliei* from northeastern Pacific and *H. barbouri* from the northwestern Pacific. They would remain their respective species, however, their genus would change to *Chimaera*.

Clade 2 consisted of eight lineages, including both *Chimaera* and *Hydrolagus* species. The only congruent relationships across data sets was the placement of *C. cubana* as the basal species, and that *H. lemures* and *H. cf lemures* were always sister taxa. The relationships among the remaining lineages in this clade were not resolved. The estimated terminal branch lengths within the nuclear data sets indicated little sequence variation, which makes interpreting their species boundaries difficult. The mitochondrial data set showed more sequence divergence. In both data sets, *C. cubana* appears to be a unique lineage from the remaining lineages. It is the only species in this lineage that occurs in the Caribbean, while the other members occur in the north Pacific and Indian Oceans. Under the mitochondrial data set, the remaining clade members could be interpreted as each being a unique lineage. However, the nuclear data set does not corroborate this, with extremely short terminal branches, the relationships cannot be distinguished between species and population variation. In this case, it is recommended that additional specimens are collected and both morphological and molecular data are analyzed to determine whether these lineages represent unique species, population variants, or a combination within the clade. Two lineages were identified as *C. phantasma* from Taiwan, yet in the mitochondrial data

they were not most closely related, but were sister taxa in the nuclear data, which may allude to unique lineages. The fact that *H. lemures* from Australia is always sister to *H. cf lemures* from Indonesia, and not more closely related to those lineages with similar distribution, could also suggest that there may be some species separation. However, several of these lineages are morphologically very similar (i.e., *H. ogilbyi*, *H. lemures*), and further investigation is warranted.

Clade 3 consisted of five lineages of *Hydrolagus* from varying geographic regions. Based on the consensus between nuclear and mitochondrial data, the only congruent relationship was the sister grouping of *H. mirabilis* and *H. alberti*, both from the Gulf of Mexico. This relationship shows high divergence in the mitochondrial data set as unique lineages, while less variation is evident in the nuclear data. However, it is suggested here that these two lineages remain their respective species, but the genus name should be changed to *Chimaera*. The relationships among the other three lineages differs between the character sets. The mitochondrial data would suggest that *H. mitsukurii* (NW Pacific) is likely a unique lineage. The low sequence variation between *H. africanus* (SE Atlantic-South Africa) and *H. sp. D* (Indian Ocean-India) could indicate a single species. However, the nuclear data reveals a different pattern, with *H. sp. D* sister to *H. mitsukurii*, and *H. africanus* appearing to be a unique lineage. Either way, it is considered that *H. mitsukurii* and *H. africanus* remain species, pending further investigation. Again, with *Hydrolagus* being synonymized, *Chimaera* would become the genus name for all *Hydrolagus* species.

Clade 4 contained one species, *H. novaezealandiae*, which was highly divergent in the mitochondrial data set, and showed a little sequence variation in the nuclear data set. This lineage represents a unique lineage from New Zealand waters. The genus name is suggested to change to *Chimaera*.

Clade 5 contained 11 identified lineages of *Hydrolagus*. The species-level relationships were variable both within and between mitochondrial and nuclear data sets. A few consistent patterns were present, like the grouping of *H. sp. B*, *H. erithacus*, and *H. purpurescens*, all from the Indian Ocean, *H. cf trolli* (Indian Ocean-South Africa) and *H. affinis* (North Atlantic), and the basal position of *H. homonycteris* (Australia). It should be noted that the specimen identified as *H. purpurescens* was collected from the Indian Ocean, while the type locality of this species is from the Pacific Ocean (Hawaii). While photo image of this specimen shows overall similarities to *H. purpurescens*, here this specimen is not taken to represent the type *H. purpurescens*, until further investigation of the specimen is undertaken. In both mitochondrial and nuclear data sets, there is extremely little sequence variation within the clade, which could indicate population variation. Also, morphologically, these species appear to be quite similar, attain large body sizes, and have an overall dark brown to black coloration. Only *H. homonycteris* does not fit this entirely, as it typically attains smaller total lengths than other members of the clade, and may represent a unique species. However, further collection and study of this species is needed to determine if it is unique or should be synonymized. Based on the results of both character sets of low genetic variation and similarities in general morphology, it is proposed that these lineages represent one large, wide-ranging species. This species would have a distribution in the Atlantic, Indian, and Pacific Oceans. In this case, *H. erithacus*, *H. pallidus*, *H. trolli*, and *H. melanophasma* would be synonymized. The species name would revert to the first recognized species in the clade, *H. affinis*. However, it would become *Chimaera affinis*, due to the elimination of the genus *Hydrolagus*.

Clade 6 consists of 16 identified lineages of *Chimaera* and *Hydrolagus*. The nuclear data set estimated very low sequence variation within the clade, which makes interpreting the species boundaries based on this data set difficult. The mitochondrial data set showed greater sequence

variation among the lineages. Taxonomic suggestions will be based on the mitochondrial data set. *Chimaera monstrosa* was highly divergent and basal to all other lineages, and represents a unique species. All the remaining lineages also appear to show sequence variation along with morphological differences that would indicate unique species. Those lineages currently recognized as valid species would remain their respective species. Undescribed lineages (i.e., *C.* sp. 3) require formal species descriptions. There were two lineages identified as *C. panthera*, GN10951 represents the originally described species, as evident by a vouchered specimen. The other *C. panthera* lineage, GN6987, was recovered as more closely related to *C. obscura*. This specimen is also vouchered, and a photo image does not provide evidence to conclude that it is the true *C. panthera*. The specimen is quite small and has lost its coloration, which is one of the obvious characters for this species. Thus, this lineage likely represents a unique species and requires further investigation of the vouchered specimen, and collection of specimens around New Zealand to describe the new species. Two lineages were identified as *H. bemisi*, both from New Zealand, and they appear to be divergent from one another. Both of these specimens are vouchered in a museum collection, and photo images show overall similarities. Future work needs to be done on this species complex around New Zealand to identify if these two lineages are truly unique species. Additionally, the species name *H. bemisi* would need to be altered to reflect the change in genus name.

There are several examples from the phylogenetic reconstruction where species that are separated based on minimal morphological differences, and to some extent geographic distribution, show very little genetic variation with respect to both the nuclear and mitochondrial data, indicating that these 'species' may actually represent only one species. For example, there are multiple lineages of the *Rhinochimaera* species that show little genetic variation, but are described in the literature based on morphological characters as well as differences in their known

geographic localities. Another example includes the *Hydrolagus* species within clade 5, which are morphologically quite similar, but do show some morphological differences across the identified lineages, as well as both overlap and separation in distribution. In these cases, the resulting phylogenetic tree provides us with a hypothesis that these lineages may actually represent only one species, respectively. A potential explanation why defining species based on limited morphological variation and/or geographic distribution may not be accurate is the concept of epigenetics. Epigenetics refers to heritable chemical modification to DNA that does not involve changes to the DNA sequence, but does involve potential changes in gene transcription and function (Peaston & Whitelaw, 2006). There are several examples that have shown a link between phenotypic variation and epigenetic modification, especially when related to environmental cues (e.g., Kooke et al., 2015; Triantaphyllopoulos et al., 2016). Thus, a species may have phenotypic variability due to differences in gene expression that is affected by epigenetic modifications to DNA, and these could be due to exposure to different environments by particular populations. However, these phenotypic changes should not be taken as a means to designate a new species. In the examples listed above, these particular species may be exposed to a variety of environments that lead to heritable effects on gene function, which in turn causes phenotypic variability. Thus, care should be taken when describing new species based on little morphological variability and even geographic distribution. Additionally, DNA sequences can help provide additional data when trying to accurately identify new species or population variants, particularly when morphological differences are weak.

Conclusions

This study sequenced mitochondrial genomes of 55 chimaeroid lineages. Previously, only eight species had mitogenome data available. Thus, 47 additional lineages, not necessarily valid species, were contributed. This is the first molecular phylogeny of chimaeroid fishes to include the majority of the known diversity within the group, using the 13 protein-coding genes of the mitogenome. The three families were recovered as monophyletic, with Callorhinchidae the basal clade, and Rhinochimaeridae and Chimaeridae sister groups. Overall, eight clades were recovered in all analyses based on both ML and BI methods. Within Callorhinchidae, the relationships among the three species were not resolved, as different analyses produced different relationships, with poor support. Rhinochimaeridae relationships were identical in all resulting topologies, *Rhinochimaera* a monophyletic genus, with *Neoharriotta* and *Harriotta* paraphyletic. Chimaeridae resulted in 6 consistent clades, with differences among analyses found only within clades. The genera *Chimaera* and *Hydrolagus* were found to be paraphyletic. Within ML analyses, the resulting topologies were mainly congruent, with differences found at shallow nodes and species-level relationships. Partitioning did not have a major effect. Character set (nucleotide versus amino acid) did result in minor differences at the species-level. One explanation is the potential for base composition bias and substitution saturation present in the nucleotide data set. When these biases are not modeled properly, especially due to the violation of stationarity, model mis-specification can lead to incorrect inferences and incongruence among trees. Amino acid data does not translate the base composition bias, but may have codon bias, and is less likely to have high homoplasy. Thus, conflicting signals in the two data sets and model violations may have produced these minor incongruences. Given the potential for base composition bias, it would be valuable for future studies to explore this topic within this mitochondrial data set. Several approaches have been developed that allow for non-stationarity in

base frequencies, including distance models (Lake, 1994; Lockhart et al., 1994; Galtier & Gouy, 1995), models that assign base frequencies to each branch using maximum likelihood methods (Yang & Roberts, 1995; Galtier & Gouy, 1998), and Bayesian methods (Foster, 2004). However, these models can become quite complex, computationally demanding, may not scale well to large amounts of data, and can result in over-parameterization of the model. Additionally, other approaches to minimize base composition bias have been used including, RY-coding (Phillips et al., 2004), removing 3rd codon positions from the analysis, and translating nucleotide data to amino acids (Betancur-R et al., 2013). A combination of identifying base frequency bias, minimizing bias, and incorporating models of non-homogeneity into the phylogenetic estimation can be explored and compared to the findings in this study. Bayesian inference recovered mainly congruent trees between partitions and character sets. When comparing ML and BI of the same partition and character type, they recovered identical or very similar results.

Divergence time dating indicated that holocephalans and elasmobranchs shared a common ancestor between the Ordovician and Carboniferous periods of the Paleozoic era. Modern day families shared a common ancestor in late Triassic to early Jurassic, when Callorhynchidae and the lineage leading to Rhinochimaeridae and Chimaeridae last shared a common ancestor. The Rhinochimaeridae and Chimaeridae lineages split during the late Jurassic period. Extant lineages diverged relatively recently, mainly in the Cenozoic, with a few lineages estimated in the Cretaceous. It would be beneficial for future studies to explore divergence time estimation using these mitochondrial markers by altering calibration fossils, prior information, and the number and taxa used for the outgroup to understand how these factors may affect the age results. Also, different estimation methods should be explored, as well as potentially using other mitochondrial markers (i.e. non-coding regions), to examine similarities and differences among the different data sets, methods, models, and assumptions.

In general, the overall topologies resulting from the mitochondrial and nuclear data sets were similar, with the same eight clades, and species within them. Incongruences were mainly evident in the placement of some of the Chimaeridae clades relative to one another, and relationships within the major clades. The fact that two independent data sets estimated the same major evolutionary relationships lends support that the results reflect the true history of this group for the given data sets. Additionally, divergence time dating estimated highly similar ages for the major splits within the group, with differences mainly in times within the Rhinochimaeridae and Chimaeridae at shallower nodes.

The results from both the nuclear and mitochondrial phylogenetic analyses suggest that there is a need for future investigation into some of the lineages to further resolve taxonomic questions. It is recommended that both morphological and molecular data be collected for those lineages outlined above in order to better understand phenotypic and genetic variation within and between species. It appears that there are likely some new species that need to be studied and described, as well as some instances where currently valid species should be synonymized under another valid species as they represent the same species.

Incongruence between data sets is likely a product of error and violation of assumptions. There are several potential reasons for incongruences, but overall, while incongruences were evident among data sets, there were no substantial disparities. There are two main types of error, random error that is due to limited sampling, and systematic error that is due to incorrect assumptions and non-phylogenetic signal in the data (Swofford et al., 1996, Rokas et al., 2003, Jeffroy et al., 2006, Philippe et al., 2011). Stochastic error is the major limitation of phylogenies based on individual genes (Philippe et al., 2005a), and several studies have shown that phylogenies from single gene analyses can be highly incongruent (Rokas et al., 2003, Phillips et al., 2004, Delsuc et al., 2005, Jeffroy et al., 2006). This is the main limitation in this study as the

mitochondrial genome represents a single, non-recombining locus, so all sites are linked (Avice, 2004; Ballard & Whitlock, 2004; Galtier et al., 2009). The use of only a single genetic marker can lead to incongruence in gene tree - species tree reconstructions (Rubinoff & Holland, 2005). However, while the use of mitochondrial markers alone is not ideal, utilizing mitochondrial markers along with nuclear markers only helps to support congruence, identify potential reasons for incongruences, and gain a better understanding of the overall evolutionary history of the taxa.

Systematic error due to incorrect assumptions or mis-specification of the model of molecular evolution, can also produce a non-phylogenetic signal in the data, leading to incongruence in data sets (Swofford et al., 1996; Felsenstein, 2004; Ho and Jermiin, 2004, Philippe et al. 2005a). There are many causes of systematic error that have been investigated and studied; those with respect to the inability of the method to properly model the evolutionary process of the data include base composition heterogeneity (Foster, 2004; Galtier & Gouy, 1995; Lockhart et al., 1992; Philippe et al., 2005a; Rodriguez-Ezpeleta et al., 2007), across-site rate variation (Yang, 1994; Lopez et al., 2002; Philippe et al., 2005a), heterotachy (site-specific rate heterogeneity through time; Kolaczkowski & Thornton, 2008; Philippe et al., 2005a; Philippe et al., 2005b; Spencer et al., 2005), site non-independence (Robinson et al., 2003; Rodrigue et al., 2006), and site heterogeneous nucleotide or amino acid replacement (Lartillot & Philippe, 2004; Pagel & Meade, 2004). These result from multiple substitutions that are not accurately modeled by the method (Philippe et al. 2005a). Mitochondrial DNA has been shown to have base composition and codon bias (Meyer, 1994; Ballard & Whitlock, 2004; Rubinoff & Holland, 2005); substitution rate variation within and among genes and between lineages (Rubinoff & Holland, 2005); early substitution saturation or homoplasy due to a high substitution rate (Meyer, 1994; Ballard & Whitlock, 2004; Rubinoff & Holland, 2005; Galtier et al., 2009). While appropriate models are chosen based on the data to incorporate this variation, no model likely

perfectly reflects the history of the data, making this another limitation of phylogenetic analysis, which can produce incorrect phylogenies. Future work would benefit from exploring if and how different models of molecular evolution affect resulting tree topologies. Also, it would be pertinent to explore base composition heterogeneity across lineages, and if extreme, explore methods to either reduce this bias, or use non-stationarity approaches to estimate the phylogeny, and compare with classical models used here to determine if this may be affecting the inference.

The differential retention of ancestral polymorphism or incomplete lineage sorting (ILS) is another major cause of incongruence in phylogenetic studies (Maddison, 1997; Satta et al., 2000; Philippe et al., 2005a; Rubinhoff & Holland, 2005), as well as hybridization/introgression (Maddison, 1997; Naumov et al., 2000; Rubinhoff & Holland, 2005). In these instances, the gene no longer has an identical history to the taxa, and error in phylogenetic estimation can arise. Incomplete lineage sorting is particularly important with more recent divergences, or when there have been short periods of time between divergences (Rubinhoff & Holland, 2005; Maddison & Knowles, 2006; Philippe et al., 2011). It is likely that ILS is in part responsible for differences seen between the nuclear and mitochondrial data sets at the species-level.

Different DNA markers are useful over varying divergence depths, due mainly to evolutionary rate variation (Avice, 2004). Mitochondrial DNA, in general, has a higher substitution rate compared to nuclear DNA, so there are more variable sites (Ballard & Whitlock, 2004; Rubinhoff & Holland, 2005; Galtier et al., 2009), which makes it a good marker for inferring shallower nodes and closely related species. Whereas slower evolving genes are better for deeper divergences, since there is less homoplasy. This study chose to use the protein-coding genes of the mitochondrial genome, as they are conserved making alignment and homology reliable, and are considered to have a faster evolutionary rate than nuclear genes. This makes them a good candidate marker for inferring relationships at the species-level. These protein-coding genes are

thought to be more consistent for phylogenetic analysis, compared to other mitochondrial regions, because they are more conserved, slow-evolving, have a relatively long alignment length, and assumed to correspond to a molecular clock (Finnila et al., 2001; Silva et al., 2002).

Other mitochondrial genome regions may also show potential for use in resolving evolutionary relationships within chimaeras, as detailed below. Thus, future work would benefit from investigating these regions, and utilizing them in a similar framework to estimate species-level relationships among chimaeroid fishes. In particular, since these protein-coding genes are likely quite conserved, there may not be enough informative variation in these genes to fully resolve these species and population-level questions. Mitochondrial ribosomal RNA (rRNA) genes are present in all organisms (Smit et al., 2007), making it a good target for phylogenetic analysis. However, they tend to be slow-evolving due to their functional importance, which would make them more useful for deeper divergences instead of species-level divergences (Avisé, 2004; Patwardhan et al., 2014). One problem with rRNA is that alignment across organisms can be difficult, while protein-coding gene alignments tend to be much more accurate (Avisé, 2004). The control region of the mitochondrial genome contains highly variable regions, and has been frequently used in population level studies (Pereira, 2000; Non et al., 2006). However, the high mutation rate can lead to higher levels of homoplasy, limiting its usefulness in accurate tree reconstruction (Finnila et al., 2001). These highly variable regions can also have problems with alignment accuracy and homology (Meyer, 1994). The last non-coding regions of the mitochondrial genome are transfer RNAs (tRNAs). Transfer RNAs are one of the most ancient groups of sequences (Widmann et al., 2010), which would make them valuable for phylogenetic studies. However, they are considered to be poor markers for the use in phylogenetic reconstruction due to short sequence lengths, difficulty with alignment, they are highly conserved, involved in horizontal gene transfer, can easily change specificity, and have been subject to gene

duplications (Widmann et al., 2010). Thus, tRNAs are likely poor representatives of the true evolutionary history of a group of organisms.

Again, the mitochondrial genome is limited by the fact that it represents only a single molecular marker, and does not contain independent loci of which can be used to replicate the data for the evolutionary history of taxa. However, the mitochondrial genome, while representing only a very small fraction of a species genome, serves as a fundamental component of an organism, making its patterns and history important. A future direction of this work could be to use the non-coding regions of the mitochondrial genome to investigate their usefulness in resolving the relationships among lineages, especially at the species level. In particular, the control region would be the first region to target, as it has a higher mutation rate, and would be more helpful at shallower nodes, where the protein-coding genes may not contain enough variation at this level to resolve some of these relationships. In order to do this, the control region for the lineages would need to be re-sequenced, as the gene capture method used here did not provide highly reliable results. However, the control regions for the lineages captured here can be used to design better baits to target this region, specific for chimaeras. Then hybridization capture can be used to target and sequence this region, followed by a suite of phylogenetic analyses. The results should be interpreted independently and compared to the protein-coding genes tree topologies. Additionally, the rRNA and tRNA genes could be aligned across the lineages and analyzed to look at congruence across the topologies of all of these regions. This may provide a more comprehensive picture of the evolutionary relationships of these fishes, as they each provide a unique temporal range and potentially history of the taxa.

This chapter provides a comprehensive phylogenetic reconstruction of holocephalans fishes using mitochondrial markers, as well as the timing of diversification of the major lineages. The congruence between the mitochondrial and nuclear phylogenies provides confidence that

major relationships represent the true evolutionary history for this group. That is, the resolution of the major clades within the families, as well as genera relationships and to an extent species-level relationships. Taxonomic implications were evident, and future work is necessary to resolve species complexes and biodiversity. These phylogenies provide a foundation on which to further explore patterns and transformations in character traits within holocephalans, as well as ancestral state reconstruction. Holocephalans can be an important reference for comparative studies with bony vertebrates given their basal phylogenetic position in the jawed vertebrate tree of life. Also, their genomic architecture is more similar to mammals, than mammals are to teleost fishes, which may make them better for comparative studies with humans than teleost fishes. The phylogeny can be used to provide a more accurate estimate of the ancestral state of a particular trait of interest, which can then be used in studies with other vertebrates to better understand genome evolution and identify patterns of genomic, physiological, and developmental transitions over time.

CHAPTER 4

Identification of Sex-Specific Genetic Markers in Chondrichthyan Fishes

Introduction

Sexual Determination

Sex determination is a fundamental process for proper development and reproduction in sexually reproducing organisms. It results from an initial genetic or environmental signal that determines whether an embryo develops as male or female; embryos contain an undifferentiated bipotential primordium, which matures into either an ovary or testis depending on the primary sex-determining signal (Kobayashi & Nagahama, 2009; Siegfried, 2010; Uller & Helanterä, 2011). Primary sex determination is thought to be a hierarchical cascade, with two main types: 1) genetic sex determination (GSD), and 2) environmental sex determination (ESD). In GSD, sex is determined by inheritance of a single gene on a heteromorphic or homomorphic sex chromosome, or several sex-related genes on multiple sex chromosomes. In ESD, sex is determined by extrinsic factors such as temperature (TSD) or social interactions. It is thought that many of the processes and pathways involved in sex determination are conserved across vertebrates (Cutting et al., 2013). However, vertebrates display a wide range of sex-determining mechanisms including several types of genetic signals (e.g., XY, ZW, XO, ZO systems), several types of environmental cues (e.g., temperature, social cues), as well as combinations of genetic and environmental signals (Quinn et al., 2011). Birds, mammals, and even crocodylians show extreme conservation in their sex determination mechanisms (Mank et al., 2006). However, other vertebrate lineages like teleost fishes, amphibians, turtles, and lizards exhibit a wide variety of sex-determining mechanisms within their respective groups (Mank et al., 2006; Bachtrog et al., 2014).

From an evolutionary standpoint, differences between species are due to changes at a molecular level (i.e., genes) that regulate developmental processes (Haag & Doty, 2005). Many critical mechanisms involved in basic vertebrate development have been found to be conserved

across metazoans (Haag & Doty, 2005; Trukhina et al., 2013). Even though sex determination is a critical process for development and reproduction, it may be the least conserved process with lability in the primary signal as well as downstream pathways (Marín & Baker, 1998; Haag & Doty, 2005; Trukhina et al., 2013). To date, eight master sex-determining genes have been identified in jawed vertebrates (Table 4.1). In some cases, closely related species and even populations within the same species have different sex-determining mechanisms, indicating that this process may be rapidly evolving (Bull, 1983; Charlesworth, 1996; Bachtrog et al., 2014).

The high diversity of sex-determining mechanisms among vertebrates suggests a complex evolutionary history of transitions among mechanisms (Janzen & Krenz, 2004; Mank et al., 2006; Quinn et al., 2011). For example, heterogametic sex, the most common mechanism known in animals, has evolved many times and in different groups (Bull, 1983). To understand the process of sex determination, it is necessary to know both the genetic basis underpinning the mechanisms, as well as how the mechanisms have transitioned over time. While several different mechanisms have been found in vertebrates studied thus far, little is known about the ancestral condition or how sex-determining mechanisms transition across vertebrate taxa.

Sex Determination in Teleost Fish

Fishes represent more than one-half of all vertebrates, with over 34,000 species (Nelson et al., 2016). Among these species there is extraordinary diversity in morphology, physiology, habitat-associations, and behavior, which has allowed them to adapt to diverse aquatic environments (Nelson et al., 2016). Not only have they adapted to various habitats, but they also have a diverse array of sex-determining mechanisms for sexual reproduction. The underlying phenotype of sexual reproduction, male or female, in fish is governed by the two main mechanisms, GSD and ESD.

Table 4.1. Known sex-determining genes in jawed vertebrates (adapted from Trukhina et al., 2013).

Species	Sex-determining gene	Gene Info
Placental mammals & marsupials	<i>SRY</i>	Transcription factor; testis-determining factor; <i>SOX3</i> duplication
Birds	<i>DMRT-1</i>	Transcription factor; DM-domain gene; testis-determining factor; dosage dependent
<i>Xenopus laevis</i>	<i>DM-W</i>	Transcription factor; DM-domain gene; ovary-determining factor; <i>DMRT1</i> duplication
<i>Oryzias latipes</i> <i>Oryzias curvinotus</i>	<i>dmy</i>	Transcription factor; DM domain gene; testis-determining factor; <i>DMRT1</i> duplication
<i>Oryzias luzonensis</i>	<i>gsdf</i> [†]	Secretory protein of <i>TGF-β</i> superfamily; duplication of gonadal soma-derived growth factor, <i>gsdf</i>
Patagonian Pejerrey	<i>amhy</i>	Anti-Müllerian hormone (<i>amh</i>) duplication; <i>TGF-β</i> superfamily
<i>Takifugu rubripes</i>	<i>amhr2</i>	<i>amh</i> receptor 2; <i>TGF-β</i> superfamily; missense SNP in kinase domain; ortholog of <i>amhr2</i>
<i>Oncorhynchus mykiss</i>	<i>sdY</i>	Truncated, divergent form of interferon regulatory factor 9 (<i>Irf9</i>)

All varieties of reproductive strategies are present in teleost fishes, which include unisexuality, hermaphroditism and gonochorism (Devlin and Nagahama 2002). Unisexual species are all female and reproduce by parthenogenesis, gynogenesis, or hybridogenesis (Lampert & Scharl, 2008). Hermaphroditism includes individuals that are both phenotypically male and female and produce both egg and sperm at some point within their lifetime. Gonochorism refers to species that have two separate sexes, where individuals express their phenotype throughout their entire life, and is the most common reproductive mode in fishes. From a phylogenetic perspective, hermaphroditism appears to have independently arisen in several different orders of fishes, while gonochorism appears to be present in all fish orders (Avisé & Mank, 2009), indicating the extreme flexibility among fish reproductive modes.

Fishes show all mechanisms of GSD, which includes monogenic systems with heteromorphic or homomorphic sex chromosomes, multiple sex chromosome systems, and polygenic systems with more than one genetic factor located on different chromosomes, that leads to the determination of sex and gonad differentiation (for detailed reviews of GSD see Devlin & Nagahama, 2002; Volff et al., 2007; Sandra & Norma, 2009; Brykov, 2015; Huele et al., 2014; Martínez et al., 2014). The phylogenetic distribution of GSD mechanisms among teleosts is presented in Mank et al. (2006).

In a monogenic system, sexual development is typically determined by the presence or absence of a single genetic factor located on a sex chromosome (Devlin & Nagahama, 2002). Most fish species with GSD lack morphologically distinguishable sex chromosomes, indicating that they are relatively young, and have not had time to differentiate (Ohno, 1974; Baroiller et al., 2009a). However, there are some examples of fish with heteromorphic sex chromosomes (Ota et al., 2003; Chen et al., 2008a; Chen & Reisman, 1970; Peichel et al., 2004).

There are two main types of monogenic systems, male heterogamety (XX/XY), where the male has two different sex chromosomes, produces two types of gametes, and the sex-determining gene is generally present on the Y chromosome in males, absent from the X chromosome, and female heterogamety (ZZ/ZW), where the female produces two types of gametes and may carry the sex-determining gene. Fishes show all types of monogenic systems; XY male heterogamety (Yamamoto, 1969; Thorgaard, 1977; Schartl, 2004; Baroiller et al., 1999) and ZW female heterogamety (Mair et al., 1991; Volff & Schartl, 2001; Takehana et al., 2008; Chen et al., 2014). In several cases, both XY and ZW systems are present in closely related species, for example, in tilapia species (Cnaani et al., 2008). Chromosome translocations and fusions have led to XX/XO systems (Devlin & Nagahama, 2002; Alves et al., 2006) and multiple sex chromosome systems seen in a variety of species (see Kitano & Peichel, 2012; Gihigliotti et al., 2014). Also, more than one sex chromosome system may be present in the same species (e.g., playfish, *Xiphophorus maculatus*, X, Y, and Z chromosomes; Schultheis et al., 2006; Cioffi et al., 2013).

There are currently five known sex-determining genes that have been recruited in fish, each with a different known function. All but one of the genes is known to be involved in early downstream events of sex determination and differentiation in vertebrates. Most studies to date have focused on trying to identify the master sex-determining gene in fishes with GSD. However, recent studies have shown that potential autosomal-linked minor genetic factors, as well as minor environmental factors may also influence sex determination in species with GSD, even those with a major sex-determining gene (Martínez et al., 2014).

The medaka, *Oryzias latipes*, has male heterogamety (XX/XY); however, sex chromosomes are indistinguishable (Uwa & Ojima, 1981). This was the first teleost fish to have its sex-determining gene isolated, *dmy/dmrt1y* (DM-domain gene on the Y chromosome), which

encodes a putative protein with a DM-domain DNA-binding motif (Matsuda et al., 2002; Nanda et al., 2002). It was further found that *dmy*, located in the sex-determining region of the Y chromosome, was necessary for normal male development (Matsuda et al., 2002). Additionally, Nanda et al. (2002) identified a functional duplicated copy of the autosomal *DMRT1* present in the ~ 280 kb sex-determining locus of the medaka, termed *dmrt1y*. They showed that it was only expressed in male embryos, whereas the autosomal copy was expressed later in development (Nanda et al., 2002). The identified *dmy* and *dmrt1y* are synonymous; it is the only functional gene in the sex-determining locus of the Y chromosome (Kondo et al., 2006), it is necessary for male development, and expressed early in development in only males. The *dmy* gene also has been found to be the sex-determining gene for the closely related Malabar ricefish, *O. curvinotus*, with homologous Y chromosome to *O. latipes* (Matsuda et al., 2003; Kondo et al., 2004). Studies, however, have shown that the *dmy* of *O. latipes* and *O. curvinotus* is not present in other *Oryzias* species (*O. celebensis*, *O. mekongensis*, *O. luzonensis*) nor any other fish species studied thus far (Kondo et al., 2003; Kondo et al., 2004; Tanaka et al., 2007; Myosho et al., 2012).

DMRT1 (DM-related transcription factor 1) represents a gene family, all of which contain the highly conserved DM domain (Zhu et al., 2000). DM domain genes are known to play a role in the sex-determining pathway of *Drosophila* (*doublesex/dsx*) and *Caenorhabditis elegans* (*mab-3*; Shen & Hodgkin, 1988; Burtis & Baker, 1989; Yi et al., 2000). *DMRT1* encodes a transcription factor known to have a role in sex determination (i.e., male development) in a range of vertebrates including mammals, birds, turtles, frogs, and tilapia (Raymond et al., 1998; Nanda et al., 1999; Raymond et al., 1999; Smith et al., 1999; Guan et al., 2000; Kettlewell et al., 2000; Marchand et al., 2000; Shibata et al., 2002), and is the gene most homologous to *dmy* (Matsuda et al. 2002).

Oryzias luzonensis, Luzon ricefish, a closely related species to *O. latipes* and *O. curvinotus*, also possesses an XX/XY sex-determining system with homomorphic sex chromosomes (Hamaguchi et al., 2004). A novel male sex-determining gene in *O. luzonensis* was identified, termed *gsdf^y*, which is a derived version of the gonadal soma derived growth factor (*gsdf*) (Myosho et al., 2012). *Gsdf* is a secretory protein within the *TGF-β* super family (Myosho et al., 2012), implicated in proliferation of primordial germ cells and spermatogonia in *O. mykiss* (Sawatari et al., 2007). There were two alleles, *gsdf^x* specific to the X chromosome and *gsdf^y* specific to the Y chromosome (Myosho et al., 2012). Comparisons between the two alleles revealed 12 nucleotide substitutions, but the amino acid sequences were identical (Myosho et al., 2012). Gene expression profiles for *gsdf^x* and *gsdf^y* revealed higher expression in XY embryos, with higher *gsdf^y* expression at zero days after hatching, and similar expression levels five and ten days after hatching (Myosho et al., 2012). Huele et al. (2014) speculated that male expression may be due to a cis-regulatory sequence mutation of the *gsdf^y* allele, potentially in the steroidogenic factor 1 (*sf-1*) binding site that could cause failure of binding in *gsdf^y*, affecting expression (Myosho et al., 2012). Transgene experiments were conducted to confirm *gsdf^y* as the sex-determining gene in *O. luzonensis*, where *gsdf^y* reverses XX females into fertile XX males.

The Patagonian pejerrey, *Odontesthes hatcheri*, exhibits an XX/XY sex-determining system (Koshimizu et al., 2010; Hattori et al., 2010). A Y-linked, duplicated copy of anti-Müllerian hormone (*amh*), termed *amhy*, was identified as the candidate master sex-determining gene (Hattori et al., 2012). They showed that *amhy* was found in only males and on a single chromosome of XY fish, identifying the presumed Y chromosome, which is morphologically indistinguishable from the X. *Amhy* expression was detected in XY males from six days after fertilization, to hatching, before morphological differentiation of gonads, and during testicular differentiation. Knockdown of *amhy* in XY embryos inhibited testicular development, leading to

up-regulation of female transcripts (*foxl2*, *cyp19a1a*) and ovary development. Their evidence supports *amhy* as a candidate sex-determining gene necessary for testis development in *O. hatcheri*. *Amh* is a secretory protein within the *TGF-β* super family, produced by Sertoli cells, responsible for regression of Müllerian ducts during male sex differentiation of mammals, birds, and reptiles (Josso et al., 2001; Teixeira et al., 2001; Rey et al., 2003; Johnson et al., 2008). This is the second *TGF-β* member to be implicated as a master sex-determining gene in fishes, with previous evidence for involvement in sex differentiation.

The tiger pufferfish, *Takifugu rubripes*, has an XX/XY sex-determining system with homomorphic sex chromosomes (Kikuchi et al., 2007) along with a sequenced genome (Aparicio et al., 2002). Kikuchi et al. (2007) first identified *amh* type II receptor (*amhr2*) gene as a potential candidate sex-determining gene. Subsequently, Kamiya et al. (2012), through the use of genetic and association mapping using the genome, identified a single nucleotide polymorphism (SNP) in the *amhr2* gene that perfectly correlated with phenotypic sex. The SNP, a C > G, was located within the kinase domain in exon nine of the *amhr2* gene, which led to an amino acid change from a histidine to an aspartic acid. The allele with a histidine had reduced function, and when homozygous, lead to development of ovaries. When heterozygous, one histidine allele and one aspartic acid allele, the gonads developed as testis. All males investigated have been heterozygous and females homozygous (Kamiya et al., 2012). Two other species of *Takifugu*, *T. pardalis* and *T. poecilonotus*, also possess the sex-linked SNP in *amhr2*, indicating that it too is the sex-determining gene in these species. On the other hand, it is not present in another pufferfish lineage, *Tetraodon nigroviridis*, indicating that the SNP likely was present in the common ancestor of the *Takifugu* species after divergence from *Tetraodon*. They also provided evidence that the gene is an ortholog of *amhr2* found in other vertebrates, and not a gene duplication. *Amhr2* is type II receptor within the *amh* signaling pathway, belonging to the *TGF-β*

family (Josso et al., 2001). It appears that *amhr2* and *TGF-β* signaling in fishes has a critical role in sex determination and differentiation, and may be more widespread in vertebrates. This would make the third example of a TGF-β family member gene as a master sex-determining gene, instead of a transcription factor.

The rainbow trout, *Oncorhynchus mykiss*, has an XX/XY sex-determining system (Johnstone et al., 1979; Hunter et al., 1982, 1983; Johnstone & Youngson, 1984; Davidson et al., 2009). There is evidence for both homomorphic (Thorgaard et al., 1983; Felip et al., 2005) and heteromorphic sex chromosomes (Thorgaard, 1977). Yano et al. (2012) identified the sex-determining gene, *sdY* (sexually dimorphic on the Y chromosome) in *O. mykiss*. The *sdY* gene encodes a putative protein with sequence similarity to the carboxy-terminal domain of *Irf9* (interferon regulatory factor 9) (Yano et al., 2012). They provided evidence that *sdY* is expressed in early testicular development, localized to epithelial cells of gonads and to some somatic cells around the germ line, no expression in ovaries/females, and strictly male-specific, located in the sex-determining locus. All the evidence indicating that *sdY* is necessary and sufficient for testicular differentiation. All other salmonid species to date have been determined to have an XX/XY sex determining system, however, the Y chromosome across species are not syntenic (Yano et al., 2013). The *sdY* gene was found to be present in 14 other salmonid species, and in all but two species *sdY* was male-specific (Yano et al., 2013). The *sdY* gene encodes a truncated, divergent protein, similar to *Irf9*, an immune-related gene (Yano et al., 2012).

The half-smooth tongue sole, *Cynoglossus semilaevis*, is an important aquaculture species and the first flatfish to have a sequenced genome (Chen et al., 2014). This species has a female heterogametic ZZ/ZW sex-determining system with distinguishable sex chromosomes (Zhuang et al., 2006; Chen et al., 2009; Shao et al., 2010). Due to its importance as a commercially important species, and the fact that females grow much faster than males, the ability to develop

female stocks for aquaculture has been an increased area of study, where a number of female-specific markers have been amplified (Chen et al., 2007; Liu et al., 2007; Chen et al., 2008b; Liu et al., 2008; Chen et al., 2009; Liao et al., 2014). A study found that under a normal temperature (22 °C), sex was mainly determined by GSD, but under high temperature treatment during critical sex-determining period, ~73% of ZW individuals became sex-reversed pseudomales (Chen et al., 2014). This indicated that this species was also sensitive to temperature, and thus, has both GSD and ESD. When these pseudomales were crossed with a normal female, they produced viable offspring with a male bias, that is, 94% of ZW individuals were pseudomales. The Z-linked genes retained their paternal methylation pattern (Chen et al., 2014), indicating a role of methylation in sex determination. All of this pointed the researchers to hypothesize a Z-linked mechanism of sex determination for *C. semilaevis* that results in male development. They searched genes on the Z chromosome, and found one in particular, *DMRT1*, known to be involved in sex determination in vertebrates. They suggested that *DMRT1* may be the Z-linked sex determining gene for this species because the only functional copy resides on the Z chromosome, it is highly expressed in male germ cells and somatic cells surrounding undifferentiated gonad during development, as well as during testis development, and there was demethylation of the promoter region of *DMRT1*, but has yet to be confirmed.

The Nile tilapia, *Oreochromis niloticus*, has evidence of both GSD and influences of temperature (Mair et al., 1991; Cnaani et al., 2008; Baroiller et al., 2009b). Previous studies have indicated an XY system, where sex-linked markers were identified near the *amh* gene (Lee et al., 2003; Eshel et al., 2011). Even more recent, Eshel et al. (2014) identified a male-specific duplication of *amh*, termed *amhy* (herein referred to as *amhy-1*) that differed from *amh* sequence by a 233 bp deletion on exon seven. In 2015, Li et al. identified another Y-linked duplication of *amh*, also termed *amhy* that lies downstream of the previously identified *amhy-1*. *Amhy* coding

sequence is identical to the X-linked copy of *amh*, however, there is a missense SNP, which changes an amino acid, and lacks 5608 bp of the promoter sequence (Li et al., 2015). Expression profiles of *amhy* and *amhy-1* indicated that they were both expressed exclusively in XY gonads at five days post hatching, onward, during the critical period of sex determination and development. Knockout experiments of *amhy* resulted in male to female sex reversal in XY fish, and over expression resulted in female to male sex reversal in XX fish. These same experiments were done with *amhy-1*, which did not induce sex reversal. They conclude that *amhy* is both sufficient and necessary for male development, and thus, a good candidate for the male sex-determining gene in this particular strain of Nile tilapia. It is possible that different strains may have different modes, sex chromosomes, and sex-determining genes (Li et al., 2015), and thus, will need to be examined in more depth in other strains and populations.

Not all species with GSD have a single major sex-determining gene on a particular sex chromosome. A polygenic sex system is one where multiple sex loci or alleles determine sex (Moore & Roberts, 2013), and has been reported for several fish species (Devlin & Nagahama, 2002; Vandeputte et al., 2007). Polygenic sex determination was first observed in a poeciliid fish species, the platyfish, *Xiphophorus maculatus* (Kallman, 1973). *Xiphophorus maculatus* has a multiple sex chromosome system with XY and YY males and XX, XW, and YW females, all of which are homomorphic (Schultheis et al., 2009). There is both male and female heterogamety, and multiple loci are postulated to be linked to the sex-determining region (Schultheis et al., 2009). There are many other examples of fish that use more than one gene for sex determination (Martínez et al., 2014; Cnaani et al., 2008; Ser et al., 2010; Ross et al., 2009; Vandeputte et al., 2007). A polygenic system of sex determination has also been postulated for *D. rerio*, with potentially four different sex-linked chromosomes and environmental factors (Liew et al., 2012; Bradley et al., 2011; Anderson et al., 2012; Shang et al., 2006).

Fish inhabit various aquatic habitats, each exposed and dependent on the environment, including embryos. Thus, changes in environmental conditions could potentially have an effect on embryo sex determination and differentiation. In order for sex determination through environmental factors to be present, there would need to be a difference in a particular cue that would lead to the sexual differentiation pathway for one sex over the other. Sex determination in many fish species is influenced by environmental factors such as temperature, pH, population density, oxygen content, and social interactions (Crews, 1996; Nakamura et al., 1998; Baroiller et al., 1999; Baroiller & D'Cotta, 2001). Both gonochoristic species and hermaphroditic species experience ESD, either as a primary signal, a minor factor, or as a mechanism for sex change. It has been speculated that it may not necessarily be the environmental factor that affect sex ratios but instead their influence on growth rate (Kraak & de Looze, 1993). For example, temperature affects metabolism and therefore growth rate (Baroiller et al., 2009a). For detailed review of environmental sex determination in fishes see Devlin and Nagahama (2002), Baroiller et al. (2009a), Sandra and Norma (2010), and Martínez et al. (2014). A major gap in our knowledge is our lack of understanding of the molecular mechanism by which environmental factors act on the sex-determining cascade.

The most prevalent factor of ESD in fishes is temperature; there are over 60 species of fish, across a divergent array of orders and families that are temperature sensitive during development that can affect sex (Baroiller et al., 2009a). Water pH is another environmental factor that has been shown to affect sex ratios in fish (Rubin, 1985; Römer & Beisenherz, 1996; Baron et al., 2002). Hypoxia was studied in the zebrafish, where it was found that hypoxia resulted in a greater proportion of males compared to normal oxygen conditions, postulating hypoxia as a potential factor that affects sex differentiation (Shang et al., 2006). Density has been suggested as a factor in determining sex ratios for eels (*Anguillidae*), where crowding is

associated with more males (Davey & Jellyman, 2005), and for the Chinese paradise fish, *Macropodus opercularis*, where individuals in isolation developed as males, while females were correlated with density (Francis, 1984).

Social factors affecting sex are found in many hermaphroditic fish species, where a particular social cue leads to a change in sex. The main example of this socially regulated environmental sex change is the presence or absence of a socially dominant individual, which has been found in several species of fishes, mainly among the wrasses, damselfishes, and parrotfishes (Godwin, 2009). When the dominant individual or individuals of a social group are removed, the largest individual of the opposite sex is then stimulated to change sex and become the dominant individual of the group (Godwin, 2009). The Midas cichlid, *Amphilophus citrinellus*, is an example of size-related social control of sex change (Francis, 1990; Francis & Barlow, 1993). The relative size of individuals during the juvenile stage determines sex. Larger fish are more dominant, and become males, while smaller individuals become females (Francis & Barlow, 1993). The interactions between larger and smaller individuals affect growth and ultimately sex differentiation (Francis & Barlow, 1993). However, a study in a wild population from a different area did not see any link between juvenile size and masculinization (Oldfield et al., 2006), and this may be due to the observation in a wild versus a controlled laboratory setting, or that different populations may have evolved different factors for sex change.

There are several examples of fish species that have both a genetic sex-determining system where an environmental factor has the ability to alter sexual development. The most common environmental factor to have an effect on GSD is temperature (Baroiller et al., 1999, 2009a; Baroiller & D'Cotta, 2001; Baroiller & Guiguen, 2001). While it has been known that temperature has the ability to alter sexual phenotype in some species of fish, the link between GSD and ESD is still unknown.

Zebrafish have been hypothesized to have a polygenic sex-determining system with both genetic and environmental factors (Anderson et al., 2012, Liew et al., 2012), with morphologically indistinguishable chromosomes (Pijnacker & Ferwerda, 1995; Daga et al., 1996; Traut & Winking, 2001; Phillips et al., 2006). However, breeding experiments have resulted in sex ratios that would be predicted if sex determination was genetic (Tong et al., 2010; Pelegri & Schulte-Merker, 1999; Liew et al., 2012). Two different studies have identified sex-associated loci, albeit they were different loci predicted in each study (Bradley et al., 2011; Anderson et al., 2012). Environmental factors including temperature and hypoxia have also been thought to influence sex ratios (Pelegri & Schulte-Merker, 1999; Uchida et al., 2004; Shang et al., 2006; Abozaid et al., 2011). Luzio et al. (2016) found that sex ratio was skewed towards male zebrafish when exposed to high temperatures. Ospina-Álvarez & Piferrer (2008) showed that sex ratio under normal rearing temperatures for zebrafish were no different than expected for GSD. It was only when temperatures were increased outside the normal range, that sex ratios became skewed towards males. Since those are extreme temperatures that most individuals will not be exposed to during development, it is likely that temperature is not a primary factor, but it can have a minor impact (Liew et al., 2012).

Tilapia species are known to have GSD with both XY and ZW systems within the group, with no morphological differences between chromosomes (Cnaani et al., 2008). Some species have also been shown to be polygenic, with both major and minor genetic factors contributing to sexual development (Baroiller et al., 1995, 1996; Cnaani et al., 2008). While the species' studied thus far appear to have GSD, in some species, temperature has been shown to influence sex ratios (see Baroiller et al., 2009a, b). Nile tilapia shows sensitivity to temperature, with high temperatures leading to masculinization of XX progeny (Baroiller et al., 1995, 1999). However, not all strains showed sensitivity to temperature (Baroiller et al., 1999). Other tilapia species, *O.*

mossambicus (XY; Wang & Tasi, 2000) and *O. aureus* (ZW; Desprez & Melard, 1998) have also shown high temperature during the critical period increased proportion of males. It appears that tilapia species have a range of sex-determining factors including major genetic components, potential minor genetic components on autosomes, with the possibility of temperature also influencing sex ratios. However, these are not consistent across species and may depend on particular populations, strains, and locality.

The European sea bass has a polygenic sex-determining system (Vandeputte et al., 2007) that can be influenced by temperature (Piferrer et al., 2005). In general, sex ratios may be skewed to males at high temperatures as seen in aquaculture and to females at lower temperatures (see Baroiller et al., 2009a). A recent study on the sea bass investigated the methylation pattern of the aromatase (*cyp19a*) promoter, a gene necessary for the synthesis of androgens to estrogens (Navarro-Martín et al., 2011). They found that the *cyp19a* promoter had twice the number of DNA methylated sites in males compared to females, which would lead to a decrease in expression of the aromatase pathway. On the other hand, females, with fewer methylated sites, would have increased expression, and in theory more estrogen needed for female development (Navarro-Martín et al., 2011). Higher temperature was correlated with increased methylation, and decreased *cyp19a* expression, leading to a male phenotype. XX females were sex-reversed with high water temperature, and results indicated higher methylation. Additionally, temperature effects on the gonadal transcriptome showed up-regulation of testis differentiation markers, down-regulation of ovarian differentiation markers, and an increase in genes connected with epigenetic regulation (Díaz & Piferrer, 2015).

The initial signal for sex determination in fishes appears to be a complex trait, that can involve a single master gene, a single master gene with minor genetic factors or environmental cues, multiple genetic loci, or environmental factors, and the interactions within each ultimately

leads to a male or female phenotype. We have also seen that hermaphroditic species have the ability to change sex in response to environmental cues, further complicating the picture of sex determination among fishes.

Sex Determination in Tetrapods

Tetrapods are a group of jawed vertebrates that include mammals, birds, reptiles, and amphibians, with both GSD and ESD. Mammalian sex determination is probably the most extensively studied in this group. They show high conservation, with all mammals having a male heterogametic XX/XY GSD system. All therian mammals (marsupials and placental mammals) have the basic XX/XY system, where sex is determined by a master gene, *SRY* (Wilhelm et al., 2007). Several studies led to the finding that *SRY* is the testis-determining region in humans and mice, and subsequently other mammals (Sinclair et al., 1990; Koopman et al., 1990; Gubbay et al., 1990; Berta et al., 1990; Koopman et al., 1991). *SRY* is an intronless gene, 887 base pairs long, located on the short arm of the Y chromosome, and encodes a transcription factor, which initiates male development. Monotremes, the other living group of mammals, while also having an XX/XY system, has evolved a different set of sex chromosomes. The platypus male has five sets of XY chromosomes (XYXYXYXYXY), and the female five sets of XX chromosomes (XXXXXXXXXX), and it lacks the *SRY* gene (Foster et al., 1992; Grützner et al., 2004; Rens et al., 2004).

Birds also show extreme conservation, all having a female heterogametic ZZ/ZW GSD system (Ferguson-Smith, 2007; Smith, 2010). More primitive birds appear to have sex chromosomes that are very similar or identical morphologically (Ogawa et al., 1998), while more derived birds have obvious heteromorphic sex chromosomes (Smith, 2010). There are two hypotheses put forward for the mechanism of sex determination in birds: 1) the W chromosome

carries an ovary-determining factor (Nanda et al., 1999; Arlt et al., 2004; Smith, 2007), or 2) dosage of a Z-linked gene (Nanda et al., 1999; Smith & Sinclair, 2004). Smith et al. (2009) found that the Z-linked *DMRT1* gene was required for testis determination in the chicken, giving support to the Z dosage hypothesis. Thus, it would appear that a double dose of *DMRT1* in males leads to testis differentiation and maleness, while a single dose leads to a female phenotype.

Reptiles are less conserved, with both GSD and ESD present within the group. Within ESD, temperature is the only variable known to determine sex, where the temperature during egg incubation determine male versus female sex (Bull, 1980; Janzen & Paukstis, 1991; Valenzuela, 2004). Of reptiles, all crocodylians, tuatara, most turtles, and some lizards have ESD (e.g., Lang & Andrews, 1994; Viets et al., 1994; Deeming, 2004; Nelson et al., 2004; Harlow, 2004; Janzen & Krenz 2004; Yao & Capel, 2005; Bachtrog et al., 2014). All snakes studied to date have a female heterogametic ZZ/ZW system, with distinct sex chromosomes (Matsubara et al., 2006; Ferguson-Smith, 2007). There is both male and female heterogamety present in turtles and lizards with GSD (Modi & Crews, 2005; Ferguson-Smith, 2007). The molecular mechanisms responsible for sex determination in reptiles is still unknown (Rhen & Schroeder, 2010).

The majority of amphibians studied to date have GSD (Nakamura, 2010), with morphologically indistinguishable sex chromosomes (Hayes, 1998; Yoshimoto et al., 2008; Schmid et al., 2010). However, some species do have heteromorphic sex chromosomes (Hayes, 1998). To show the lability of sex determination, the frog, *Rana rugosa*, has both a male XX/XY and female ZZ/ZW sex-determining system. Ogata et al. (2008) found that isolated populations of *R. rugosa* had not only different sex-determining systems, but also that some had homomorphic sex chromosomes, while others showed heteromorphic sex chromosomes. Breeding experiments were used to determine that *Xenopus laevis* has a ZZ/ZW system (Chang & Witschi, 1956). Yoshimoto et al. (2008) suggested that the W-linked gene, *DM-W* is the probable

master sex-determining gene in *X. laevis*. Using transgenic tadpoles, they found that *DM-W* was crucial for ovary development. *DM-W* is a W-linked paralog of *DMRT1*, a gene known to be important in sexual development in vertebrates (Yoshimoto et al., 2008).

Overall, tetrapods have a variety of mechanisms, including male and female heterogametic GSD, with both homomorphic and heteromorphic sex chromosomes, as well as temperature-dependent sex determination (TSD). While some groups are conserved in their mechanism, others showed variation, as well as variation within a species. Overall, there is still a lot not known about the mechanisms of sex determination and differentiation within this group.

Sex Determination in Chondrichthyans

Despite their important basal position and sister group designation in the jawed vertebrate tree of life, the mechanism of sex determination among chondrichthyans, is not yet known. Relatively few cytogenetic studies have been conducted to investigate the number, appearance, and morphology of chromosomes. To date, there have been only about 77 of approximately 1200 species that have been karyotyped to some extent. Overall, chromosome number across species is highly variable, as well as morphology and size, ranging from large metacentric chromosomes down to microchromosomes (Stingo & Rocco, 2001; Heist, 2012). Further details regarding presence or absence of putative sex chromosomes has been given for 21 species (Donahue, 1974; Schwartz & Maddock, 1986, 2002; Kikuno & Ojima, 1987; Asahida et al., 1993; Asahida & Ida, 1995; Maddock & Schwartz, 1996; de Souza Valentim et al., 2006; da Cruz et al., 2011; Aichino et al., 2013; de Souza Valentim et al., 2013). In six species, there is potentially a heteromorphic sex, however, in these studies, the sex was either unknown, and/or the karyotypes were not published (Schwartz & Maddock, 1986, 2002; Maddock & Schwartz, 1996). In one species of ray, there is possible female heterogamety, however, the karyotype was not published (Maddock

& Schwartz, 1996). Three species have shown potential male heterogamety, but again, either the karyotype was not present, or the karyotype was of poor quality to be considered reliable (Donahue, 1974; Asahida et al., 1993; Maddock & Schwartz, 1996). Three other species examined suggested no heteromorphic chromosomes, and in two of these, no karyotypes were published (Maddock & Schwartz, 1996; de Souza Valentim et al., 2006). Lastly, seven species karyotypes showed the presence of putative male heterogamety. *Rhinobatos productus* and *Platyrhinoidis triseriata* showed apparent male heterogamety, while female chromosomes were all homomorphic, indicating an XX/XY GSD system (Maddock & Schwartz, 1996). Asahida & Ida (1995) identified three unpaired chromosomes in a male, and two unpaired in a female *Rhinobatos schlegelii*. Further, one of the unpaired chromosomes in the male, was paired in the female, indicating it may be related to sex. Kikuno & Ojima (1987) also identified three unpaired chromosomes in a male, two unpaired in a female *Rhinobatos hynnicephalus*, but one of the unpaired chromosomes in the male is paired in the female, indicating a potential connection to sex determination. A species of *Potamotrygon* was found to possess an apparent XX/XO sex chromosome system, where the female had a pair of chromosomes in which the male lacked one of the same pair (de Souza Valentim et al., 2013). Aichino et al. (2013) characterized the chromosomes of *Potamotrygon motoro* and identified females with 33 homomorphic pairs of chromosomes, and males with 31 homomorphic pairs, and three heteromorphic chromosomes. They suggested that this species may have a multiple sex chromosome system, $X_1X_1X_2X_2/X_1X_2Y_1$, with male heterogamety. Another study, this time investigating *Potamotrygon falkneri*, similarly identified a multiple sex chromosome system, $X_1X_1X_2X_2/X_1X_2Y_1$, with male heterogamety (da Cruz et al., 2011). The evidence that heteromorphic chromosomes exist in some species strongly suggests a genetic mechanism of sex determination.

Facultative parthenogenesis, asexual reproduction in which an embryo develops from an unfertilized egg, has been documented in both captive and wild populations of cartilaginous fishes (Chapman et al., 2007; Chapman et al., 2008; Feldheim et al., 2010; Robinson et al., 2011; Portnoy et al., 2014; Fields et al., 2015; Harmon et al., 2016; Straube et al., 2016; Dudgeon et al., 2017). In all cases, viable offspring produced by parthenogenesis were female. Molecular microsatellite data was used to verify homozygosity of offspring with mother. There is one case of a documented putative male offspring produced by parthenogenesis, however, while externally the pup showed the formation of male clasper organs, internally, sexual organs were absent (Straube et al., 2016). Additionally, the male offspring did not survive, and no genetic data was collected to help confirm parthenogenesis versus sexual reproduction. Overall, this provides evidence that these species likely have a genetic XX/XY mechanism of sex determination because a female shark (XX) would produce only XX female offspring by parthenogenesis.

Only two studies have attempted to identify potential chromosomal molecular markers that are sex-linked in a ray and shark species (Rocco et al., 2009; Rocco, 2013). Here, they used primers to locate *SRY*-like sequences on the chromosomes of *Torpedo torpedo* (Rocco et al., 2009) and *Atelomycterus marmoratus* (Rocco, 2013). In both species, the *SRY*-like sequence labels hybridized to two pairs of male chromosomes and one pair of female chromosomes. Further, PCR recovered two DNA bands, one at ~ 200 bp, and another at ~ 400 bp for both sexes (Rocco et al., 2009; Rocco, 2013). However, the ~ 400 bp bands showed slightly different lengths (324 bp for male, 380 bp for female), and the male sequences showed similarity with three human genes, *SPATA 16*, *SPATA 18*, and *UTY*, all implicated in human spermatogenesis (Rocco et al., 2009; Rocco, 2013). Overall, this shows potential differences between male and female chromosomes in these two species, indicative of a genetic mechanism of sex determination.

How to Study Sex-Determining Mechanisms

There are three main techniques for identifying a species' genetic sex-determining mechanism, 1) cytogenetics, 2) identification of sex-specific molecular markers, and 3) sex-reversal breeding experiments (Bull, 1983; Charlesworth & Mank, 2010). Cytogenetic techniques can be used to identify chromosomal differences between sexes, which would indicate the heteromorphic and homomorphic sex. Identification of sex-specific molecular markers refers to utilizing a sex-linked marker to verify the heteromorphic sex. Breeding experiments involve artificially manipulating the sex of embryos, then breeding sex reversed individuals to normal individuals, and observing sex ratios of the resulting offspring. Each one has its challenges however. Cytogenetic evidence produces only indirect evidence of heterogamety, while absence of detectable chromosomal differences does not imply heterogamety is absent (Bull, 1983). Thus far, the majority of vertebrates studied have lacked heteromorphic sex chromosomes (Hillis & Green, 1990; Devlin & Nagahama, 2002; Ezaz et al., 2009). Also, breeding experiments can be very difficult to carry out as many species are not easily reared or bred in captivity (Bull, 1983). While identification of sex-specific markers relies on detectable polymorphisms between sexes that are linked with the sex-determining region (Bull, 1983), it may be the most promising approach to identifying sex chromosome systems (Gamble & Zarkower, 2014).

Restriction site-associated DNA sequencing (RAD-seq) is a method that has been used to generate data to discover tens to hundreds of thousands of SNPs or genetic markers in hundreds of individuals in a single experiment (Baird et al., 2008; Davey et al., 2011; Gamble & Zarkower, 2014). It was originally used to identify polymorphisms and genotyping in microarray analysis (Miller et al., 2007), and further modified for next-generation sequencing (Baird et al., 2008). The concept behind RAD-seq is the use of restriction sites randomly distributed throughout the genome, and restriction enzymes that readily cut the genome into fragments at these sites. The

flanking regions of the fragments are then sequenced, allowing for detection of SNPs and gene alleles (Baird et al., 2008; van Tassel et al., 2008; Wiedmann et al., 2008). This method subsamples the whole genome, reducing complexity, and allows for a more cost-effective way to generate data for a large number of samples (Baird et al., 2008). Additionally, RAD-seq can be used with prior genome data, but a reference genome is not always necessary to use this method (Baird et al., 2008; Willing et al., 2011; Andrews et al., 2016).

RAD-seq has been widely used for SNP detection and genotyping of model and non-model organisms (Andrews et al., 2016). It has been applied in a wide range of ecological and evolutionary studies including genetic mapping (e.g. Baird et al., 2008; Baxter et al., 2011), population genetics (Emerson et al., 2010; Hohenlohe et al., 2010; Keller et al., 2012), gene flow and hybridization (Hohenlohe et al., 2011; Keller et al., 2013; Combosch & Vollmer, 2015; Gaither et al., 2015), phylogeography (Emerson et al., 2010), and phylogeny (Rubin et al., 2012; Wagner et al., 2012). RAD-seq also has the ability to identify sex-linked molecular markers, allowing for the identification of a male or female heterogametic sex chromosome system (Baxter et al., 2011; Gamble & Zarkower, 2014; Gamble et al., 2015). It has successfully been used to identify sex-linked polymorphisms and/or map the sex-determining region in the Atlantic halibut (Palaikostas et al., 2013a), the Nile tilapia (Palaikostas et al., 2013b), the salmon louse (Carmichael et al., 2013), *Anolis carolinensis* (Gamble and Zarkower, 2014), medaka (Anderson et al., 2012; Wilson et al., 2014), two species of rockfish (Fowler & Buonaccorsi, 2016), a plant species (Qiu et al., 2016), as well as a species of boa and python (Gamble et al., 2017). This shows how powerful this method can be to identify sex chromosome systems, especially in non-model species, and species that may lack morphologically distinguishable chromosomes.

Objectives and Hypotheses

This research aims to fill a major gap in our understanding of sex determination among vertebrates by investigating the currently unknown primary sex-determining mechanism among chondrichthyan fishes. The first aim is to characterize the chromosomes of somatic cells of male and female chondrichthyan species. Previous work has lent support of a potential genetic mechanism of sex determination through identification of heteromorphic chromosomes in mainly the male sex versus homomorphic chromosomes in mainly the female sex. Thus, this experiment is designed to use karyotyping and chromosomal analysis to determine number and morphology of chromosomes and infer whether heteromorphic chromosomes are present in representative male and female chondrichthyans. It is hypothesized that males will be represented by at least one pair of heteromorphic chromosomes, while females will be the homomorphic sex, indicating an XX/XY sex chromosome system and GSD.

No studies to date have attempted to identify genes or genetic markers that may be responsible for sex determination in these fishes, nor confirmed the hypothesized GSD mechanism. Thus, the second aim is to screen the genome for sex-linked molecular markers in representative chondrichthyan species using RAD-seq that would indicate a genetic mechanism of sex determination. Further, candidate sex-linked markers will be validated using PCR to isolate putative sex-determining regions revealed by the RAD-seq screen. It is hypothesized that there will be at least one sex-linked molecular marker present in the male sex of the study species, indicating an XX/XY sex chromosome system and genetic mechanism of sex determination.

Materials and Methods

Chromosome Analysis

Sample Collection

Tissue samples (blood, gill, gonad, spleen, rectal gland) were collected from shark species either through anesthetized animals at the South Carolina Aquarium, or by gillnet and/or hook and line through the South Carolina Department of Natural Resources off of Charleston, South Carolina. Tissues were extracted using sterile scalpel and forceps and placed in a sterile tube with culture media. Blood was collected in a vacutainer containing either lithium heparin or sodium heparin. See Tables 4.2-4.8 for details of each sample collection and their respective protocols.

Chromosome Preparation

Chromosome preparation followed the short-term culture method of Ida et al. (1978), Asahida and Ida (1990), and Foresti et al. (1993), followed by the air-drying method and Giemsa staining. Each experiment has its own unique methodology that is modified from original methods, and each is presented in appendices 4.1-4.8. In general, 1-3 $\mu\text{g/ml}$ of colchicine was added to minced tissue or blood mixed with culture media for 2-24 hours, at room temperature. Cell suspension was then centrifuged, supernatant discarded, and pellet re-suspended. A hypotonic 0.075 M KCl solution was added to cell suspension for 60-120 minutes. Five drops of 4°C Carnoy's fixative, a methanol-acetic acid solution at 3:1, was added to tube and gently stirred. After 5 minutes, an additional ~ 7 ml of 4°C fixative was added to tube and gently stirred. The cell suspension was then centrifuged, supernatant removed, pellet re-suspended, and repeated two more times. Pellet was then re-suspended in fresh fixative and mixed until homogeneous, with a

slightly cloudy suspension. Three drops of cell suspension were dropped onto a microscope slide, supported inside a water bath at 60°C and allowed to air dry. Once dried, the slide was stored for one day. Several slides were made per specimen.

Giemsa Staining & C-banding

Air dried slide preps were stained with Giemsa following Sumner (1971) but slightly modified for sharks and rays following Stingo et al. (1995) for C-banding. For standard Giemsa staining, slides were rinsed with distilled water, followed by immersion in 5% Giemsa stain for 10 minutes, then rinsing with distilled water two times. For C-banding, slide preps were incubated in 0.2 N HCl for 30 minutes, then immersed in a saturated solution of Ba(OH)₂ for 5 minutes at 40°C. Then slide preps were incubated in 2X SSC (salt, sodium citrate) for 10 minutes at 60°C. Slides were then rinsed with distilled water and stained with 5% Giemsa diluted with sodium phosphate buffer (pH 6.8) for 10 minutes, rinsed and blotted dry.

Chromosome Visualization

Chromosome preparation slides were visualized using a light microscope at 100X – 400X.

Cell Culture

Whole blood was obtained via caudal venipuncture from unanesthetized animals using a sterile lithium or sodium heparin vacutainer tube to prevent coagulation. Tubes were mixed gently through several inversions. Blood was kept at room temperature or on ice, for no more than 24 hours post collection.

For whole blood cell culture, in a sterile cell culture hood, the blood tube was assessed for clotting. Whole blood was mixed with an elasmobranch-modified RPMI cell culture media,

transferred to a T25 flask, and placed in an incubator at approximately 25°C for three to four days. In one experiment, the mitogen concanavalin A was added to the cell culture before being placed in the incubator. Previous work had shown that nurse shark, *Ginglymstoma cirratum*, leukocytes did not respond to normal or elevated levels of phytohaemagglutinin (PHA), a mitogen used in cell culture for humans and mice (Lopez et al., 1974). However, they did show that leukocytes responded to quite high concentrations (500-1000 µg/ml) of the mitogen concanavalin A, that lead to increased radioisotope incorporation and higher cell counts (Lopez et al., 1974). Concentrations of concanavalin A in this experiment (75 µg/ml and 250 µu/ml) were much lower than those shown to increase cell density (500-1000 µg/ml), due to the limited amount of concanavalin A available.

For white blood cell culture, peripheral blood leukocytes (PBL) were isolated from whole blood through slow speed centrifugation to pellet erythrocytes while PBL remains suspended above. Centrifugation at 50 x g for 15-20 minutes, followed by checking tube for separation. Additional centrifugation for 10 minutes was undertaken if necessary to pellet erythrocytes from PBL layer. In a cell culture hood, the PBL layer was carefully aspirated and added to a new sterile tube. Elasmobranch-modified RPMI culture media was added to PBL, and transferred to a T25 flask. Flasks were viewed under a light microscope for presence of cells. Flasks were placed in an incubator at approximately 25°C for three to four days. Details of each experiment may slightly deviate from the above methodological description. Detailed protocols are presented in Appendix 4.1-4.7.

Cell viability was monitored by mixing an aliquot of cell culture suspension with 0.4% trypan blue and applying mixture to a hemocytometer. Live and dead cells were visualized based on the presence or absence of trypan blue uptake.

Cell Culture & Chromosome Preparation at MUSC - Cytogenetics

Whole blood was obtained from two shark samples in a vacutainer containing sodium heparin. Tubes were mixed gently through several inversions. Blood was kept at room temperature or on ice, for no more than 24 hours post collection.

Blood cells were counted using a hemocytometer. Whole blood was mixed with a 2% acetic acid diluent and loaded onto the hemocytometer. Live cells were counted in four squares in each of the four quadrants, added together, and divided by five to get the number of cells/ml. Cell density of whole blood was too high to count, so blood was diluted (10:1, 20:1, 30:1, 40:1), till an accurate count could be achieved. An average cell count per square of < 160 was the target.

A total of six cell cultures were set up for each sample. Cultures A, B, C contained 0.5 ml of whole mixed blood, no dilution, with 9.5 ml of RMPI media used for normal human cell culture. Cultures D, E, F contained diluted blood and media to a volume of 10 ml. For sample ID 27, *Carcharhinus isodon*, a 40:1 dilution was optimal, with 0.75 ml of 40:1 diluted blood mixed with 9.25 ml media. For sample ID 28, *Carcharhinus plumbeus*, a 30:1 dilution of whole blood was optimal, with 0.5 ml diluted blood mixed with 9.5 ml media. Different mitogens were added to the cell cultures to try and stimulate cells to divide. Cultures A-D were treated with 200 µl of PHA, culture E with 200 µl of pokeweed, and culture F with 200 µl of IL-2. Culture A was placed in an incubator at 37°C and 5% CO₂, while the remaining cultures were incubated at room temperature (~ 24°C) on the counter, with the lid tight. All cultures were left for three days. After incubation, cell cultures were harvested and prepped for chromosome analysis. A detailed protocol is presented in Appendix 4.6, and follows the MUSC Department of Cytogenetics protocol for peripheral blood harvest and chromosome preparation.

Concanavalin A Experiment

Whole blood was obtained from two shark samples in separate vacutainers containing sodium heparin. Tubes were mixed gently through several inversions. Blood was kept at room temperature, for no more than 24 hours post collection.

Blood cells were counted using a hemocytometer. Whole blood was mixed with a 2% acetic acid diluent and loaded onto the hemocytometer. Live cells were counted in four squares in each of the four quadrants, added together, and divided by five to get the number of cells/ml. Cell density of whole blood was too high to count, so blood was diluted 30:1 with elasmobranch-modified RPMI, where an accurate count could be achieved. An average cell count per square of < 160 was the target.

An aliquot of whole blood was then used to retrieve PBL. Blood was centrifuged at 50 x g for 15 minutes. Plasma, buffy coat and top layer of erythrocytes were aspirated and placed into a new tube. New tube was centrifuged at 50 x g for 15 minutes. PBL layer was aspirated, leaving erythrocytes behind, and placed into a new tube. PBL cells were counted using a hemocytometer as detailed above. Cell density was too high, so PBL cells were diluted 10:1.

The diluted whole blood and PBL cell counts were used to determine the amount of the respective cell suspensions to mix with elasmobranch-modified RMPI to seed a 96-well plate at ~6250 cells/well. Cells were then treated with concanavalin A in differing concentrations. Undiluted concanavalin A at 10,000 µg/ml was diluted to 2000 µg/ml, 1250 µg/ml, 1000 µg/ml, 750 µg/ml, 500 µg/ml, and 250 µg/ml, and added to the appropriate wells. See Appendix 4.8 and 4.9 for 96-well plate design. Each treatment was replicated three times. A media only blank was used, which included only the elasmobranch-modified media, no cells, no concanavalin A, with BrdU incorporation. Controls included: 1) cells (whole blood, PBL), with no concanavalin A treatment, and no BrdU incorporation, and 2) media only (no cells) with concanavalin A

treatment. The plates were checked under a light microscope for presence of cells. Plates were placed in an incubator at ~ 25°C and 5% CO₂. After two days of incubation, plates were removed from incubator and 1X BrdU solution add to appropriate wells. Plates were placed back into incubator for 7-18 hours. Plates were then removed from the incubator and the BrdU cell proliferation assay kit (BioVision) protocol was followed for BrdU detection and measurement. After TMB substrate was added, absorbance at 650 nm was measured to monitor color development up to 30 minutes. The stop solution was added, and absorbance read at 450 nm on a BioTek Synergy HT microplate reader. See Appendix 4.7 for further details.

3RAD-seq

Library Preparation

A modified method of RAD-seq (3RAD) was used to prepare DNA libraries to identify potential sex-linked markers in a shark species. Tissue was collected from six males and six female sandbar sharks, *Carcharhinus plumbeus*, off the southeastern USA coast of South Carolina (Table 4.2). Total genomic DNA was extracted using the EZNA® Tissue DNA Kit (Omega Bio-Tek, Norcross, GA) and stored at -20°C. The 3RAD method is similar to double-digest RAD-seq, but uses three restriction enzymes, two which cut the DNA at specific recognition sites, and one enzyme that cuts dimerization of adapters (Graham et al., 2015; Hoffberg et al., 2016). Only those fragments that have two different adapters will be carried through the library preparation and sequenced. 3RAD library preparation follows the detailed

Table 4.2 *Carcharhinus plumbeus* samples used in 3RAD (restriction site-associated DNA) sequencing.

LAB ID	Species	Sex	Locality
14810	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Bulls Bay
15098	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Bulls Bay
15254	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Morris Island
15255	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Morgan Island
15310	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Morgan Island
15372	<i>Carcharhinus plumbeus</i>	Female	Atlantic Ocean, South Carolina, Morgan Island
14809	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Bulls Bay
15080	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Bulls Bay
15109	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Bulls Bay
15233	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Bulls Bay
15312	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Morgan Island
15392	<i>Carcharhinus plumbeus</i>	Male	Atlantic Ocean, South Carolina, Bulls Bay

protocols of Graham et al. (2015) and Hoffberg et al. (2016). A subset of samples was used to test restriction enzymes, by running a digestion on DNA and examining the resulting fragmentation via gel electrophoresis. The restriction enzymes, *EcoRI* and *NheI* showed good fragmentation and smear on the gel, with no defined bands. The 3RAD library preparation and sequencing took place at the University of Georgia. Briefly, a restriction enzyme digestion was performed using *EcoRI*, *NheI*, and *XbaI* on genomic DNA. Indexed i5 and i7 adapters for *EcoRI* and *NheI* were ligated to the ends of fragmented DNA. The reaction was cleaned up using Speedbeads (Rohland & Reich, 2012) mixed with a NaCl-PEG solution. A 1:1 ratio of Speedbeads to DNA was used to remove previous reaction reagents and unligated adapters. A PCR was used to complete adapter and index sequences. Samples were normalized to approximately the same DNA concentration and pooled together. The pooled samples were cleaned up using a 1:1 Speedbead to DNA ratio. The cleaned up pooled sample was quantified on a Qubit. A PippinPrep (Sage Science, Inc.) was used to size-select fragments (~ 550 bp), followed by PCR to enrich fragments, a clean-up step, and quantification. Samples were sequenced on an Illumina NextSeq v2 300 cycle kit at the University of Georgia, to yield 2 x 150 bp reads.

RAD Marker Identification

The program Stacks v1.35 (Catchen et al., 2011; Catchen et al., 2013) was used to initially process sequence data. The *process_radtags* script was used to clean raw sequence data, by checking for proper indexes and cut sites, and checking quality. Any reads with uncalled bases were removed. Reads with low quality scores were discarded based on a sliding window (15% of read length) along read, checking average quality score. If the score dropped below 90%, the read was removed. Barcode and cut sites were rescued within 1 bp of their expected sequence,

otherwise, they were discarded. The Stacks pipeline, *denovo_map* was used to analyze data. First, the *ustacks* program takes data of an individual and aligns reads into exactly matching stacks; then within the individual, stacks are compared, and loci are built de novo, and SNPs are called. The program *cstacks* then creates a catalog of consensus loci across all samples based on sequence similarity. The program *sstacks* searches the stacks and loci built for each individual in *ustacks* against the catalog of loci from *cstacks*, matching samples against catalog. The *denovo_map* pipeline was run with these requirements: highly repetitive stacks were removed in *ustacks*; the minimum number of identical raw reads required to make a stack for an individual at a locus (*m*) was 3; the maximum distance between stacks (*M*) in order to merge them into a locus for an individual was set at 2; the number of mismatches allowed between putative loci when building the catalog (*n*) in *cstacks* was set to 3. Each resulting locus was scanned for presence/absence in males and females, as well as strict allelic variation between sexes.

Additionally, raw data was sent to Dr. Tony Gamble at Marquette University, where the data was analyzed using a custom bioinformatic pipeline (Gamble & Zarkower, 2014; Gamble et al., 2015; Gamble et al., 2017). Raw reads were demultiplexed, filtered, and trimmed using the Stacks *program process_radtags*. The program RADtools v.1.2.4 (Baxter et al., 2011) was used to generate putative RAD markers for each individual and putative loci across all individuals from each restriction site, individually. Program settings for RAD marker generation included a cluster distance of 10, minimum quality score of 20, read threshold (minimum number of reads needed to form a RAD marker) of 5. For putative loci generation, settings included a tag count threshold of 4, and maximum number of mismatches set at 2. To be considered a sex-specific RAD marker, the minimum number of individuals needed is one less than the total number of individuals of that sex, while the opposite sex is absent. RADtools outputs the presence and absence of each RAD locus and allele for each sample. Species were analyzed separately. Next,

a custom python script (Gamble et al., 2015) was used to screen putative sex-specific RAD markers identified by RADtools against the raw reads of the opposite sex. If a RAD marker had one or more matches in the raw reads of the opposite sex, the marker was excluded from further analysis. This will remove false positive markers which lack the required number of reads to create the original RAD marker. Next, the candidate sex-specific RAD markers that pass this filter had their forward and reverse reads extracted for further analysis.

RAD Marker Validation

Primers were designed for each of the putative sex-specific markers identified by the *denovo_map* program. PCR was carried out in a 25 μ l volume reaction using TaKaRa Ex Taq and 1:10 diluted DNA (~ 2.5 ng/ μ l) of the original six male and six female *C. plumbeus* samples. The reaction was denatured at 94°C for 2 minutes, subjected to 30 cycles at 94°C for 30 seconds, 48-52°C for 30 seconds, and 72°C for 1 minute, followed by 72°C for 5 minutes. A 3 μ l sample of the PCR product was run on a 1% agarose gel.

Single Digest RAD-Seq

Taxon Sampling

White muscle tissue was extracted from multiple individuals of a chimaera, *Callorhynchus callorynchus*, a shark, *Carcharodon carcharias*, and a ray, *Maculabatis randalli*, and stored in 95% ethanol. Individuals were sexed at the time of collection. There was a total of six male and eight female *C. callorynchus*, eight male and eight female *C. carcharias*, and seven male and seven female *M. randalli* (Table 4.3).

Table 4.3 Chondrichthyan male and female samples used for single digest RAD-seq library preparation and sequencing. See Gamble et al. (2015) for P1 and P2 adapter sequences and barcodes.

Lab ID	Species	Sex	Locality	P1 Adapter	P2 Adapter
17406	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-A1	P2-longI4
17407	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-A2	P2-longI4
17408	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-A3	P2-longI4
17409	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-A4	P2-longI4
17411	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-B1	P2-longI4
17418	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-Ado1	P2-longI4
17419	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-Ado2	P2-longI4
17420	<i>Callorhynchus callorynchus</i>	Female	Argentina	P1-Ado3	P2-longI4
17405	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Ado4	P2-longI3
17410	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Ado5	P2-longI3
17412	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Ado6	P2-longI3
17413	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Wil1	P2-longI3
17414	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Wil2	P2-longI3
17416	<i>Callorhynchus callorynchus</i>	Male	Argentina	P1-Wil3	P2-longI3
14282	<i>Maculabatis randalli</i>	Female	Iran	P1-Wil1	P2-longI11
14297	<i>Maculabatis randalli</i>	Female	Iran	P1-Wil2	P2-longI11
14329	<i>Maculabatis randalli</i>	Female	Iran	P1-Wil3	P2-longI11
14410	<i>Maculabatis randalli</i>	Female	Iran	P1-Wil4	P2-longI11
14423	<i>Maculabatis randalli</i>	Female	Iran	P1-Wil5	P2-longI11
14442	<i>Maculabatis randalli</i>	Female	Iran	P1-A1	P2-longI11
14446	<i>Maculabatis randalli</i>	Female	Iran	P1-A2	P2-longI11
14165	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado1	P2-longI14
14426	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado2	P2-longI14
14428	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado3	P2-longI14
14430	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado4	P2-longI14
14434	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado5	P2-longI14
14443	<i>Maculabatis randalli</i>	Male	Iran	P1-Ado6	P2-longI14
14445	<i>Maculabatis randalli</i>	Male	Iran	P1-B1	P2-longI14
10688	<i>Carcharodon carcharias</i>	Female	Australia	P1-Wil4	P2-longI5

17581	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-Wil5	P2-longI5
17589	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-A1	P2-longI5
17598	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-A2	P2-longI5
17602	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-A3	P2-longI5
17603	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-A4	P2-longI5
17608	<i>Carcharodon carcharias</i>	Female	Guadalupe Island	P1-Ado5	P2-longI5
17767	<i>Carcharodon carcharias</i>	Female	Long Island, USA	P1-Ado6	P2-longI5
10687	<i>Carcharodon carcharias</i>	Male	Australia	P1-B1	P2-longI6
17582	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Ado1	P2-longI6
17590	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Ado2	P2-longI6
17592	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Ado3	P2-longI6
17597	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Ado4	P2-longI6
17599	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Wil1	P2-longI6
17605	<i>Carcharodon carcharias</i>	Male	Guadalupe Island	P1-Wil2	P2-longI6
17772	<i>Carcharodon carcharias</i>	Male	Long Island, USA	P1-Wil3	P2-longI6

Library Preparation

Sex-specific molecular markers were investigated using the single digest RAD-seq method of Etter et al. (2011), and modified by Gamble et al. (2015). A detailed protocol that was used for library preparation of chondrichthyan samples is presented in Appendix 4.10. Briefly, genomic DNA was extracted from muscle tissue with RNase A treatment, quantified, and checked for quality on a gel. A starting DNA concentration of at least 25 ng/μl was used for all samples. Genomic DNA was digested using a high fidelity *SbfI* restriction enzyme. P1 adapters with indexes (Table 4.3) were ligated onto restriction cut sites of digested DNA using T4 DNA ligase. Samples were pooled, in equimolar ratios, making male and female libraries for each species. Pooled libraries were sheared to approximately 500 bp fragments. Sheared libraries were cleaned up using a MinElute PCR purification kit, followed by size selection and agarose gel extraction for fragments spanning 250-700 bp. Libraries were blunt-end repaired and an adenine was added to the 3' end of DNA fragments, to prepare for P2 adapter ligation. A P2 adapter, each with a unique Illumina barcode, was ligated to each library (Table 4.3). Libraries were PCR amplified using 14 cycles with KAPA HiFi HotStart Ready Mix. The PCR product was cleaned up using magnetic beads and a PEG/NaCl buffer, followed by another round of size selection using agarose gel extraction. Libraries were concentrated using the magnetic beads once more, and quantified. Libraries were sent to Duke University for sequencing on an Illumina HiSeq 4000, using a 150 bp paired-end sequencing approach (386,483,506 clusters passed filtering; % bases with \geq Q30 = 85.84%).

RAD Marker Identification

Raw fastq read files were demultiplexed by Illumina barcode using the fastq-multx command line tool in the ea-utils package (Aronesty, 2011). These demultiplexed files were then

analyzed using a custom bioinformatic pipeline (Gamble & Zarkower, 2014; Gamble et al., 2015; Gamble et al., 2017). The data was first further demultiplexed into individual samples using their respective in-line barcodes in the Stacks program *process_radtags*, along with filtering and trimming. The pipeline protocol details are outlined above under the 3RAD methodology, RAD marker identification section. Briefly, RADtools was used to generate putative RAD markers for each individual, and RAD loci across all individuals. Each species was analyzed separately. The output identified presence and absence of each RAD locus for each sample. The custom python script (Gamble et al., 2015) takes the putative sex-specific RAD loci identified by RADtools and screens them against the raw reads of the opposite sex. If a putative RAD locus has an exact identity to a raw read present in the opposite sex, the marker was excluded. Forward and reverse reads were extracted for those candidate sex-specific RAD markers that passed this validation.

Sex-Specific RAD Marker Validation

Further validation is required of the candidate sex-specific RAD markers identified by RAD-seq. All candidate sex-specific RAD markers that passed the initial validation were assembled into contigs using their paired-end reads in Geneious v.7.1.9 (Biomatters Ltd, Auckland, NZ). First, forward and reverse reads were set as pairs. The paired-end reads were assembled de novo with the following settings for the high sensitivity/medium method: gaps allowed, maximum number gaps per read at 20%, maximum mismatches per read at 40%, maximum gap size of 5, maximum number of ambiguities at 16, and a minimum overlap identify of 80%.

PCR was used to validate the sex specificity of the candidate sex-specific RAD markers in *C. callorynchus* only. This species was chosen as it had the least number of candidate markers. Future work will investigate the candidate sex-specific RAD markers of the other two species.

Two sets of primers were designed for each candidate sex-specific RAD marker (Table 4.4), using IDT (Integrated DNA Technologies, Inc.) PrimerQuest and OligoAnalyzer3.1. Additionally, a positive internal control was designed based on a 103 bp region of the mitochondrial *NADH2* gene to monitor amplification performance (Table 4.4). The positive control was validated in the original eight females and six males, as well as an additional two males. A no DNA negative control was included. The Phusion high fidelity (HF) DNA PCR kit (New England Biolabs) was used to carry out the PCR reaction. Briefly, the reaction was carried out in a 25 μ l reaction following kit protocol and included, 1X Phusion HF buffer, 200 μ M dNTPs, 0.25 μ M of forward and reverse primers, 1 unit of Phusion polymerase, and nuclease-free water up to 25 μ l. Approximately 50 ng of DNA was used in the reaction for each sample. The PCR conditions were 98°C for 30 seconds, followed by 30 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 15 seconds, then an extension at 72°C for 10 minutes, and a hold at 4°C. The PCR reaction was run on a 1.0% agarose gel to confirm amplification of target sequence.

Next, a PCR was conducted to identify the best pair of forward and reverse primers for each RAD marker to be used in subsequent experiments. One male sample was chosen to validate the primers. Each RAD marker had two forward and two reverse primers designed. Each possible pair for the appropriate marker was tested (i.e., F1R1, F1R2, F2R1, F2R2). A negative control was used for each of the four pairs of primers. Also, a primer control was included, which included only one of the primers, along with DNA and reagents (e.g., Forward 1 only, DNA, reagents). The same Phusion HF PCR kit was used, following the kit protocol. For this experiment, primer concentration was increased to 0.5 μ M per primer. An amount of DNA (~ 50 ng) was used, as recommended by the PCR kit. The same PCR conditions as stated for the

Table 4.4 Primers designed for candidate sex-specific RAD contigs in *Callorhynchus callorhynchus*. PC=positive control

RAD Marker	Primer Name	Forward Sequence	Reverse Sequence
1889	CC_M1889_1	CTCTTCCGCCCTCAAAGAAGAACTAATC	GGATGTCCTCCCAATGTGAAATG
1889	CC_M1889_2	GTACGGTAGCACACTTCCATTAG	CTAGATTGGGTCTGGGAAACAC
1895	CC_M1895_1	AGATGTGGTTACAGACTGGTAGG	CAAGCAGTGTGTGTTCTTATTGG
1895	CC_M1895_2	GCTGTGCCAACACACCCCTTAG	CTCGTCCAGGCATCTGTTCTAC
PC	Chimaera_ND2	TCTCCTTAGCTATTATCTCCACACT	CATCCTAGGTGAGCAATTGAAGAG

positive control experiment was used here. The PCR results were visualized on a 1.5% agarose gel to confirm the primers amplified only the correct molecular weight region. The primer pair with the best amplification as evidence by a tight, clean band on the gel, was chosen for subsequent experiments for each RAD marker.

A PCR was conducted to then test for amplification of the RAD marker and positive control region in a co-amplification PCR. The same PCR reaction protocol was used here, except both the RAD marker primers (0.5 μ M) and positive control primers (0.25 μ M) were included in the same reaction. Primers for RAD marker 1889 included CC_1889_1_Forward and CC_1889_2_Reverse, which amplify a 360 bp region, and for RAD marker 1895, CC_1895_1_Forward and CC_1895_1_Reverse, which amplifies a 519 bp region. A 1.0% agarose gel was used to confirm the presence and co-amplification of positive control and RAD marker.

The last set of experiments were conducted to validate presence of the RAD markers in the eight male specimens and determine whether they are present or absent in the eight female genomes. The two RAD markers were run in separate experiments. A PCR was carried out using following the kit protocol, including both the RAD marker primers and positive control primers, as outlined above, as well as a negative control. A 1.0% agarose gel was used to identify presence or absence of the respective RAD markers in males and females, or significant size differences that would indicate two alleles on sex chromosomes in the heterogametic sex.

Additionally, validated RAD markers were examined using BLAST (Altschul et al., 1990) against the whole genome shotgun sequences of a male *Callorhinchus milii* in the NCBI database. The NCBI open reading frame (ORF) finder was used to search each major contig for potential protein-coding regions, followed by BLASTP of each potential ORF protein sequence.

Results and Discussion

Cell Culture & Chromosome Analysis

A total of 30 shark tissue samples were collected for either cell culture and/or chromosome analysis experiments. Tissue samples included gills, gonad, spleen, rectal gland, and whole blood. The first two shark tissue samples (lab ID 1 & 2; gill tissue), were used for chromosome analysis. No cells were present on prepared chromosome slides (Table 4.5). No cell pellet was observed during the chromosome preparation protocol. The gill tissue may not have been minced/ground up properly to release cells into the media for further processing. The next chromosome preparation used three shark samples, of which, four different tissue types were extracted (lab ID 3-14; Table 4.5). Normal appearing cells were present on all chromosome slides, but no cells were observed in metaphase (Table 4.5).

The next experiment used whole blood from two shark samples (lab ID 15 & 16; Table 4.6). Peripheral blood leukocytes were extracted from the whole blood and cultured with no mitogens. Cells were present in both cell cultures, along with cellular debris. No cells were present on the prepared chromosome slides (Table 4.6). An additional eight shark samples (lab ID 17-24) were collected for cell culture and chromosome analysis. Two cell cultures used whole blood, and six cultures used extracted PBLs (Table 4.6). During extraction of the PBL layer, it was difficult to define the buffy coat layer, and some erythrocytes ended up being aspirated and included in the cell cultures. Cell cultures also had some apparent blood coagulation as seen by obvious blood clots, which may have had an effect on the cell cultures. Two types of additives were used for anti-coagulation, lithium and sodium heparin, but neither appeared to be better than the other. Cell cultures were checked for presence of cells, and in all but one (lab ID 17), cells

Table 4.5 Shark tissue samples collected for chromosome preparation protocols only. Detailed protocols are found in the Appendix.

Lab ID	Species	Sex	Tissue	Giemsa/C-banding	Visualization	Protocol
1	<i>Sphyrna tiburo</i>	F	Gill Arch	No	No cells present	Appendix 4.1
2	<i>Sphyrna tiburo</i>	F	Gill Arch	No	No cells present	Appendix 4.1
3	<i>Carcharhinus limbatus</i>	F	Gonad	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
4	<i>Carcharhinus limbatus</i>	F	Gill Arch	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
5	<i>Carcharhinus limbatus</i>	F	Spleen	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
6	<i>Carcharhinus limbatus</i>	F	Rectal gland	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
7	<i>Carcharhinus isodon</i>	M	Gonad	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
8	<i>Carcharhinus isodon</i>	M	Gill Arch	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
9	<i>Carcharhinus isodon</i>	M	Spleen	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
10	<i>Carcharhinus isodon</i>	M	Rectal gland	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
11	<i>Carcharhinus isodon</i>	M	Gonad	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
12	<i>Carcharhinus isodon</i>	M	Gill Arch	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
13	<i>Carcharhinus isodon</i>	M	Spleen	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2
14	<i>Carcharhinus isodon</i>	M	Rectal gland	2 Giemsa/2 C-banding	Cells present, no metaphases	Appendix 4.2

Table 4.6 Shark blood samples collected for cell culture and chromosome preparation. PBL = peripheral blood leukocytes. Cell viability was checked for some of the samples using a hemocytometer. Detailed protocols are found in the Appendix.

Lab ID	Species	Sex	Tissue	Giemsa Staining	Visualization	Protocol	Tube Additive	Cell Viability
15	<i>Scyliorhinus retifer</i>	F	PBL	No	No cells present	Appendix 4.3	Lithium heparin	-
16	<i>Scyliorhinus retifer</i>	M	PBL	No	No cells present	Appendix 4.3	Lithium heparin	-
17	<i>Sphyrna tiburo</i>	F	PBL	Giemsa	No cells present, some debris	Appendix 4.4	Sodium heparin	No cells present
18	<i>Sphyrna tiburo</i>	F	PBL	Giemsa	Some cells, lots of debris, no metaphases	Appendix 4.4	Sodium heparin	Lots of live cells
19	<i>Sphyrna tiburo</i>	F	PBL	Giemsa	Cells present, little debris, no metaphases	Appendix 4.4	Lithium heparin	Lots of live cells
20	<i>Sphyrna tiburo</i>	F	Whole Blood	Giemsa	Cells present, some debris, no metaphases	Appendix 4.4	Lithium heparin	Some live cells
21	<i>Sphyrna tiburo</i>	F	Whole Blood	Giemsa	Cells present, some debris, no metaphases	Appendix 4.4	Sodium heparin	Some live cells
22	<i>Sphyrna tiburo</i>	F	PBL	Giemsa	Lots of cells, no metaphases	Appendix 4.4	Lithium heparin	Lots of live cells
23	<i>Rhizoprionodon terraenovae</i>	M	PBL	Giemsa	Lots of cells, some debris, no metaphases	Appendix 4.4	Sodium heparin	Lots of live cells
24	<i>Sphyrna tiburo</i>	M	PBL	Giemsa	Few cells, some debris, no metaphases	Appendix 4.4	Lithium heparin	Lots of live cells

were alive and abundant. All chromosome prepared slides, except lab ID 17, had presence of cells, but none were in metaphase (Table 4.6).

Two additional shark whole blood samples were collected to test if the addition of a mitogen, concanavalin A, would induce cell division. Three different concentrations (0, 75, 250 $\mu\text{g/ml}$) were used in one sample, and two concentrations (0, 75 $\mu\text{g/ml}$) used in the other (not enough blood to make a third culture). Cell cultures were checked for cell viability, and all had abundant live cells, too dense to be counted. After one day in incubation, cell cultures were again checked for cell viability. Sample lab ID 25 had abundant, live cells in all three cultures (Table 4.7). Sample lab ID 26 had few live cells with lots of dead cells present in the no concanavalin A, however, checking the culture flask showed lots of cells present. The aliquoted sample for cell viability may not be representative of the cell culture. Additional aliquots should have been taken to confirm the original findings. The 75 $\mu\text{g/ml}$ concanavalin A culture showed few cells, but had more alive than the no concanavalin culture (Table 4.7). After two days in incubation, cell cultures were again checked for cell viability. The three cell cultures of sample lab ID 25 showed abundant, live cells throughout. The two cultures of sample lab ID 26 were much darker in appearance, which made it difficult to distinguish cell abundance and viability. Cell cultures were checked again for viability before chromosome preparation. Cells were present and alive, with some dead cells in the three cultures of lab ID 25. Very few cells were present and alive in the two cultures of lab ID 26. During chromosome preparation both cultures of lab ID 26 were extremely dark and thick, and would not spin down to separate cell suspension, so the cultures were discarded and chromosome preparation stopped. The three cultures of lab ID 25 were successfully prepared for chromosome analysis, however, after visualizing the slides, while cells were present, no metaphase spreads were found (Table 4.7).

Table 4.7 Shark whole blood samples collected for cell culture with differing concentrations of the mitogen concanavalin-A, followed by chromosome preparation. Cell viability was checked on hemocytometer. Detailed protocol is found in the Appendix.

Lab ID	Species	Sex	Tissue	Visualization	Protocol	Cell Viability	Concanavalin-A Conc.
25	<i>Sphyrna tiburo</i>	F	Whole Blood	Cells present, no metaphases Did not complete chromosome prep	Appendix 4.5	Lots of live cells	0 µg/ml, 75 µg/ml, 250 µg/ml
26	<i>Carcharhinus isodon</i>	F	Whole Blood		Appendix 4.5	Lots of live cells	0 µg/ml, 75 µg/ml

Table 4.8 Whole blood sample from a male *Carcharhinus isodon* (lab ID 27) collected for cell culture experiment and chromosome preparation at the Medical University of South Carolina, Department of Cytogenetics. Six cell cultures were set up. Blood refers to the amount of whole or diluted blood used for culture. Detailed protocol is found in the Appendix.

Cell Culture	Blood	Media	Mitogen	Culture type	Culture Length	Visualization	Protocol
A	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Incubator 37 °C	3 days	Cells present, no metaphases	Appendix 4.6
B	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
C	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
D	0.75 ml 40:1 diluted blood	9.25 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
E	0.75 ml 40:1 diluted blood	9.25 ml RPMI	200 µl pokeweed	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
F	0.75 ml 40:1 diluted blood	9.25 ml RPMI	200 µl IL-2	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6

Table 4.9 Whole blood sample from a female *Carcharhinus plumbicus* (lab ID 28) collected for cell culture experiment and chromosome preparation at the Medical University of South Carolina, Department of Cytogenetics. Six cell cultures were set up. Blood refers to the amount of whole or diluted blood used for culture. Detailed protocol is found in the Appendix.

Cell Culture	Blood	Media	Mitogen	Culture type	Culture Length	Visualization	Protocol
A	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Incubator 37 °C	3 days	Cells present, no metaphases	Appendix 4.6
B	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
C	0.5 ml whole blood	9.5 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
D	0.5 ml 30:1 diluted blood	9.5 ml RPMI	200 µl PHA	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
E	0.5 ml 30:1 diluted blood	9.5 ml RPMI	200 µl pokeweed	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6
F	0.5 ml 30:1 diluted blood	9.5 ml RPMI	200 µl IL-2	Counter ~ 24 °C	3 days	Cells present, no metaphases	Appendix 4.6

Two shark whole blood samples were collected for an experiment to test different mitogens and cell densities, utilizing protocols that are successful at culturing and karyotyping human tissues (Table 4.8, 4.9). For shark sample, lab ID 27, whole blood cells density was found to be incredibly high compared to humans, and thus, the blood was diluted 40:1 (83 cells/square average), which equaled approximately 16.6×10^6 cells/ml. Shark sample, lab ID 28, also showed high cell density, with an average cell count per square of hemocytometer of 226 at 20:1 and 175 at 30:1. The 175 cells/square average equals approximately 35×10^6 cells/ml. These diluted blood cell densities are similar to the higher end of cell densities used for human cell culture. The whole blood cultures, which contain too many cells to count, may be consuming media resources too quickly, not allowing cells to grow. Both whole undiluted blood and diluted blood cultures were set up (Table 4.8, 4.9). Cells remained alive for three days. After chromosome preparation, slides were viewed under a microscope, and in all slides, cells were present. However, no cells were viewed in metaphase, which is necessary for further chromosome analysis.

A BrdU cell proliferation assay was used to test if the mitogen, concanavalin A, at differing concentrations, stimulates shark blood cells to divide. The experiment was conducted twice, using different blood samples, and included both whole blood and PBL cells. Concanavalin A concentrations were based on a previous study that showed a concentration between 500-1000 $\mu\text{g/ml}$ induced proliferation of blood cells in a shark species (Lopez et al., 1974). Thus, concentrations ranged from 0-2000 $\mu\text{g/ml}$. The results of the BrdU assay for one replicate are presented in Table 4.10. The media only + BrdU, and blood cells only, with no treatment and no BrdU, showed a baseline absorbance between 0.118-0.155. The whole blood and PBL absorbance with no treatment and BrdU incorporation show values similar to blank and controls, indicating that the BrdU is not having an effect. The whole blood cells with

concanavalin A treatment showed an increase in absorbance with concanavalin A treatment, compared to no treatment (Table 4.10). While the lowest value was seen at the lowest concentration, all values at the other concentrations were very similar, indicating no dose-response. The PBL with treatment also showed an increase in absorbance with concanavalin A treatment, compared to no treatment (Table 4.10). There was also no obvious dose response with increasing treatment concentration. In a few cases, the three replicates for each treatment showed extensive variability. The media control with concanavalin A treatment was expected to be similar to baseline values. However, the addition of concanavalin A caused an increase in absorbance for all concentrations, which were similar to values seen in both the whole blood and PBL treated cells (Table 4.10). Thus, it appears that the concanavalin A is having an effect on the absorbance, and the increase in absorbance seen in the whole blood and PBL treated cells cannot be attributed to an increase in cell proliferation.

The second replicate for this experiment showed similar results (Table 4.11). Blank and controls indicated that the presence of BrdU and cells were not having an effect on absorbance, but that the media was likely having a small effect. Whole blood treated cells showed absorbance values very similar to the untreated cells (Table 4.11). The PBL treated cells also showed absorbance values near the untreated cells (Table 4.11). The absorbance of the untreated PBL cells was quite high (0.437), with the 3 replicates showing a lot of variability, compared to the PBL control cells with no BrdU treatment (0.279), which also had an outlier replicate. Thus, values measured for the treated cells, ranging from 0.172 (2000 $\mu\text{g/ml}$) to 0.529 (1000 $\mu\text{g/ml}$), could not be distinguished from no treatment. The media control with concanavalin A treatment showed an increase in absorbance across all the treatment concentrations (Table 4.11). This would indicate that an increase in absorbance, seen without the presence of cells, is due to an

Table 4.10 BrdU cell proliferation assay for *Carcharhinus isodon* blood sample treated with differing concentrations of concanavalin A. PBL = peripheral blood leukocytes. ConA Conc. = concanavalin A concentration. 450 abs. = measured absorbance of each well at 450 nm. Count = number of replicates. Std Dev = standard deviation. CV = coefficient of variation.

Sample Type	ConA Conc. (µg/ml)	450 abs.	Count	Mean	Std Dev	CV (%)
Media Only	0	0.102	3	0.118	0.014	11.773
	0	0.125				
	0	0.127				
Whole Blood, No BrdU	0	0.109	3	0.121	0.017	13.974
	0	0.113				
	0	0.14				
PBL, No BrdU	0	0.125	3	0.155	0.033	21.553
	0	0.191				
	0	0.149				
Whole Blood	2000	0.331	3	0.498	0.207	41.684
	2000	0.432				
	2000	0.73				
Whole Blood	1250	0.52	3	0.602	0.128	21.278
	1250	0.75				
	1250	0.537				
Whole Blood	1000	0.514	3	0.438	0.078	17.714
	1000	0.359				
	1000	0.44				
Whole Blood	750	0.53	3	0.487	0.131	26.944
	750	0.592				
	750	0.34				
Whole Blood	500	0.915	3	0.513	0.348	67.866
	500	0.315				
	500	0.309				
Whole Blood	250	0.276	3	0.237	0.04	16.671
	250	0.238				
	250	0.197				
Whole Blood	0	0.119	3	0.122	0.014	11.269
	0	0.11				
	0	0.137				
PBL	2000	1.953	3	1.18	0.738	62.502
	2000	0.484				
	2000	1.103				

PBL	1250	0.56	3	0.48	0.082	17.099
	1250	0.484				
	1250	0.396				
PBL	1000	0.339	3	0.401	0.056	13.951
	1000	0.448				
	1000	0.415				
PBL	750	0.311	3	0.367	0.05	13.55
	750	0.384				
	750	0.406				
PBL	500	0.33	3	0.308	0.033	10.797
	500	0.27				
	500	0.325				
PBL	375	0.266	3	0.387	0.191	49.204
	375	0.289				
	375	0.607				
PBL	0	0.273	3	0.194	0.071	36.401
	0	0.137				
	0	0.172				
Media, No Cells	2000	0.842	3	1.127	0.733	65.034
	2000	1.959				
	2000	0.579				
Media, No Cells	1250	0.81	3	0.607	0.181	29.797
	1250	0.463				
	1250	0.548				
Media, No Cells	1000	0.441	3	0.361	0.094	25.95
	1000	0.258				
	1000	0.385				
Media, No Cells	750	0.391	3	0.387	0.189	48.845
	750	0.196				
	750	0.574				
Media, No Cells	500	0.57	3	0.489	0.086	17.676
	500	0.5				
	500	0.398				
Media, No Cells	250	0.384	3	0.318	0.057	18.018
	250	0.289				
	250	0.281				

Table 4.11 BrdU cell proliferation assay for *Rhizoprionodon terraenovae* blood sample treated with differing concentrations of concanavalin A. PBL = peripheral blood leukocytes. ConA Conc. = concanavalin A concentration. 450 abs. = measured absorbance of each well at 450 nm. Count = number of replicates. Std Dev = standard deviation. CV = coefficient of variation.

Sample Type	ConA Conc. (µg/ml)	450 abs.	Count	Mean	Std Dev	CV (%)
Media Only	0	0.136	3	0.185	0.094	50.849
	0	0.126				
	0	0.294				
Whole Blood, No BrdU	0	0.163	3	0.172	0.061	35.802
	0	0.115				
	0	0.237				
PBL, No BrdU	0	0.111	3	0.279	0.293	104.998
	0	0.109				
	0	0.618				
Whole Blood	2000	0.174	3	0.2	0.04	19.819
	2000	0.181				
	2000	0.246				
Whole Blood	1250	0.21	3	0.206	0.019	9.375
	1250	0.223				
	1250	0.185				
Whole Blood	1000	0.209	3	0.288	0.073	25.477
	1000	0.302				
	1000	0.354				
Whole Blood	750	0.219	3	0.48	0.245	51.141
	750	0.514				
	750	0.706				
Whole Blood	500	0.216	3	0.296	0.07	23.777
	500	0.349				
	500	0.322				
Whole Blood	250	0.167	3	0.2	0.057	28.772
	250	0.166				
	250	0.266				
Whole Blood	0	0.132	3	0.288	0.136	47.097
	0	0.356				
	0	0.377				
PBL	2000	0.175	3	0.172	0.046	26.849
	2000	0.216				

	2000	0.124				
PBL	1250	0.184	3	0.187	0.038	20.395
	1250	0.226				
	1250	0.15				
PBL	1000	0.254	3	0.529	0.245	46.227
	1000	0.611				
	1000	0.722				
PBL	750	0.506	3	0.344	0.146	42.491
	750	0.302				
	750	0.223				
PBL	500	0.977	3	0.48	0.439	91.32
	500	0.318				
	500	0.146				
PBL	250	0.162	3	0.379	0.225	59.395
	250	0.363				
	250	0.611				
PBL	0	0.756	3	0.437	0.29	66.392
	0	0.364				
	0	0.19				
Media, No Cells	2000	0.421	3	0.276	0.138	49.961
	2000	0.262				
	2000	0.146				
Media, No Cells	1250	0.251	3	0.287	0.066	23.077
	1250	0.363				
	1250	0.246				
Media, No Cells	1000	0.513	3	0.515	0.096	18.536
	1000	0.612				
	1000	0.421				
Media, No Cells	750	0.363	3	0.346	0.063	18.174
	750	0.276				
	750	0.398				
Media, No Cells	500	0.574	3	0.607	0.041	6.678
	500	0.594				
	500	0.652				
Media, No Cells	250	0.556	3	0.49	0.15	30.535
	250	0.596				
	250	0.319				

effect of the concanavalin A and media, and any increase seen in treated cells cannot be attributed to cell proliferation.

Overall, in no instance were cells present in metaphase that would allow visualization of chromosomes for karyological study. There have been several previous studies that have successfully prepared chromosome slides which contained metaphases that were used for further analyses. The majority of these studies used *in vivo* methods, where animals were held in captivity, and treated by injection with a solution of colchicine while alive, and then after a period of treatment, were sacrificed, tissues removed, and further processed (e.g., Donahue, 1974; Asahida et al., 1987; Asahida et al., 1988; Asahida & Ida, 1989; Stingo et al., 1995). Tissues that were used included those with high cell turnover, testis, spleen, gills, and kidney, tissues that were tried in this study, but using a different method. A few studies have used an *in vitro* method, whereby tissues were removed from the animal, minced, and treated with colchicine (Asahida & Ida, 1990; da Cruz et al., 2011). One recent study that resulted in successful karyotypes, included both direct injection of colchicine to animals, as well as lymphocyte culture and chromosome preparation (Aichino et al., 2013). Maddock & Schwartz (1996) also used blood samples of elasmobranchs in short-term culture, where colcemid was added to the blood and incubated for a period of 6-12 hours, followed by further processing, and in long-term culture, where blood was incubated for 5-7 days before further processing. One other successful method for observing metaphase cells in elasmobranchs was the induction of mitosis by injecting live animals with a yeast solution to stimulate cell division (de Souza Valentim et al., 2006; de Souza Valentim et al., 2013). It was not feasible in this study to keep animals in captivity to then use these direct techniques of injection, which appear to be rather successful in a variety of species. The fact that several studies have been successful using *in vitro* methods and culturing of blood, shows that these methods can work. Similar *in vitro* methods were used in this study; however, it is likely

that more troubleshooting and optimizing of the protocol for elasmobranch cell culture and chromosome preparation would be needed to successfully prepare metaphase chromosome slides. Resources and time were limited to further carry out optimization of cell culture and chromosome analysis.

There are several issues that could have occurred during cell culture and chromosome preparation, and resulting optimizations that could be used to remedy problems. In cases where no cells were present on slides after chromosome prep, there was likely an issue during the preparation that caused the loss of cells, or during the dropping of cells on to the slides. There are several factors that can affect cells adhering to the slide and spreading out, including temperature, humidity, and dropping technique, all of which could be altered. In the majority of cases, cells were present on the slides, but no cells were present in metaphase. For tissues like gill, spleen, testis, that were used, one technique for future studies would be to make sure the tissue is digested properly, potentially using agents like collagenase or trypsin, to release cells from the tissue, but not break open the cells. This may have been a major issue when using these types of tissue here, that not enough cells were released from the tissue chunks, or that many of the cells were damaged during mincing.

It would seem that the main issue was that no or few cells were in the process of cell division, and thus, no metaphase cells. Cell culture was used in the possibility that blood cells would divide during the incubation, and metaphase cells would be present for chromosome preparation. There are several factors that could be altered during cell culture. Cells can be very sensitive to the conditions in which they are incubated. The media needs to provide the optimal environment for the cells to live and grow. A modified media was used here in most experiments, which is altered for elasmobranch cellular environment. Future studies could try to adjust the media components to see if these have an effect on cell viability (e.g., type of media, salt, urea,

serum, antibiotics). Other conditions that can have an effect are the temperature, humidity, and gases during incubation. Since most elasmobranchs have a body temperature the same as their environment, cells were incubated near the temperature of the water in which they were collected. However, any of these conditions may have an effect on cell growth, and would benefit from being tested in future work. The timing of incubation is also a very important factor. Cells should be incubated long enough that the cells have time to divide, but not too long to where the environment becomes toxic and there are no more nutrients. However, there are no established protocols for elasmobranch cell culture of blood cells. Also, while there is evidence that cells do divide in the peripheral blood, there is not much work in this area. Thus, there is no information on the doubling time of these types of cells to establish a good incubation time. Future work aimed at better understanding elasmobranch blood cell division, and timing of incubation are warranted. In order for cells to divide, they need the proper signals, and the blood cells may not be receiving these signals to induce mitosis. In order to induce mitosis, generally, a mitogen is added to the cell culture to stimulate the cells to divide. One study in particular looked at the mitogens PHA and concanavalin A in a shark species, which showed that concanavalin A induced proliferation at high concentrations (Lopez et al., 1974). Thus, this mitogen was used in experiments, as well as other mitogens often used in human cell cultures, PHA, pokeweed, and IL-2. However, none of these mitogens appeared to induce mitosis in blood cells. Either these mitogens do not provide the proper signals for cell division in these species, or the concentrations were not optimal. Future studies should explore these mitogens, as well as others, and use different concentrations. Another potential problem involves the density of cells used in cell culture. If too many cells are placed in the culture, they may be too crowded and/or use up all the nutrients too fast, not allowing cells to grow and divide. All of these factors could also be highly species-dependent, and that should be taken into account when optimizing methods.

Other potential issues that could have arisen involve the methods of chromosome preparation. Major steps include colchicine treatment, hypotonic solution treatment, fixation, and slide dropping. The concentration and timing of colchicine treatment, the timing of hypotonic solution treatment, and the proper fixation and washing of cells can all have an effect on cells. These steps may need to be optimized to retrieve optimal chromosome spreads.

The BrdU assay showed high variability and no difference between controls and treated cells. It appeared that the treatment solution, concanavalin A, was having an effect on outcome, and not cell proliferation. Thus, it could not be concluded whether the mitogen induced cell proliferation. Potential issues could be human error with the plate preparation and subsequent BrdU protocol, as well as potential cell death during the process. This assay needs to be optimized in future studies using this mitogen, and other potential mitogens, to determine which, if any, induces proliferation and at what concentration. Optimization of cell seeding, treatment time, media, incubation conditions, and BrdU incorporation time are needed. Additionally, a positive control needs to be included to make sure that the assay is working to validate the results.

3RAD

The female 3RAD library had a total of 4,028,606 reads, and the male library had 3,810,076 reads. Female individual raw reads ranged from 548,322 to 873,982, and males ranged from 478,972 to 751,090 (Table 4.12). RAD-tags (filtered reads) for female individuals ranged from 532,533 to 827,839 (mean=644,522), and for male individuals ranged from 456,142 to 726,200 (mean=608,593) (Table 4.12). Females had an average of 50,869 RAD loci, and males had an average of 48,351 RAD loci (Table 4.12). RAD loci had an average of 4x coverage (Table 4.12). Overall, there were a total of 180,908 unique RAD loci among all twelve individuals.

Unique RAD loci were screened for the presence or absence in each sex. There were five RAD loci recovered in all six males, and 50 loci recovered in five of the six males, not found in any of the females (Table 4.13). There were nine RAD loci recovered in all six females, and 77 loci recovered in five of the six females, not found in any of the males (Table 4.13). Additionally, RAD loci found in all 12 individuals were screened for potential allelic differences between sexes, of which none showed sex-specific alleles. RAD loci found in 11 of the 12 individuals also showed no sex-specific alleles. Thus, nine putative sex-specific RAD markers were identified in females and five in males.

Primers were only designed for each of the putative sex-specific RAD markers (9 female, 5 male) identified in all individuals of a sex through the Stacks pipeline results (Table 4.14). All putative sex-specific markers appeared to fail the PCR validation. However, in the majority of PCR experiments, there was a lot of non-specific material amplified, as well as weak bands, and no amplification in some of the individuals. This indicates that the PCR conditions were not optimal, and in order to accurately conclude whether these markers are sex-specific, optimization of PCR conditions and potentially primers would need to occur. For the five putative male-specific markers, there appeared to be a band at approximately the right molecular weight present in both males and females (Figure 4.1-4.5). There were also non-specific bands amplified for all male specific primers, in both sexes. The presence of non-specific material is not ideal, and it would be advised to optimize the PCR conditions (e.g., DNA concentration, number of cycles, annealing temperature) and potentially design new primers for the markers in order to provide more accurate evidence that these markers are present in both males and females. For female RAD marker 67066, a non-specific smear was present in both males and females, and a low molecular weight band in both sexes, that is likely the target band of ~ 140 bp (Figure 4.6). Female-specific markers 62654 (Figure 4.7) and 63104 (Figure 4.8) showed weak amplification

Table 4.12 Summary of 3RAD sequencing data analysis of male and female *Carcharhinus plumbeus* using exclusively the Stacks program. # RAD tags = number of filtered reads. # SNPs = number of single nucleotide polymorphisms recovered.

Lab ID	Sex	# Raw reads	# RAD tags	# RAD Loci	Mean RAD loci coverage	# Polymorphic loci	# SNPs
14810	Female	548322	532553	45733	4.14	2006	3013
15098	Female	674692	649049	47839	4.15	2158	3354
15254	Female	567400	538045	40720	3.99	1543	2335
15255	Female	793592	763334	65857	4.40	3528	5175
15310	Female	570618	556312	41731	4.07	1808	2749
15372	Female	873982	827839	63336	4.10	2890	4445
14809	Male	657546	629190	52343	4.20	2437	3645
15080	Male	734468	701463	59143	4.19	2863	4470
15109	Male	751090	726200	68406	4.46	3634	5366
15233	Male	642924	616267	42625	4.14	2065	3208
15312	Male	478972	456142	31352	3.91	1126	1743
15392	Male	545076	522295	36239	4.04	1593	2380

Table 4.13 Summary of 3RAD putative sex-specific loci identified from the Stacks program.

RAD Loci ID	Sex-Specificity	# Individuals w/ Loci	Mean Read Coverage
9998	Female	6	4.50
12075	Female	6	4.17
54677	Female	6	5.17
58030	Female	6	4.33
62654	Female	6	6.33
63104	Female	6	4.83
66178	Female	6	4.00
67066	Female	6	3.33
56104	Female	6	4.83
9873	Male	6	4.33
515	Male	6	6.33
12965	Male	6	4.50
48396	Male	6	4.17
17488	Male	6	3.67

Table 4.14 Primers designed for male and female-specific RAD loci identified by the Stacks pipeline for the 3RAD sequencing data.

Primer ID	Forward Sequence	Reverse Sequence
Male_48396	AAT TCC ATA GTT TGG CTG TTC C	CAA ACT CAT CAC CTC TAA AC
Male_17488	CTA GCG CTT CAG TTG TTA AC	ACA TGT CGG TAA TTA AGG AAC C
Male_12965	CTC TGG AAT TAA GAA TCA ACC G	CAT GTA GGC CAA ACC AGG
Male_9873	CTA CAT GTA TTT TAC TTT CCC CAG	ACA AGG AAT TAT AGG TGT CC
Male_515	TCC AGA ACT TGA AGC TCG G	CCA TCT CTA ATT ATC TTC AGT TGC
Female_67066	AAT TCA GCA TGG CCA ATCC	AGC GTC AGT GAC CCG AG
Female_66178	CAC CAA AAG CTA AAA GAC TTG	AAC AGA CAC GAT TCC ATC
Female_63104	CAA ATC AGG ATG GAA ATT CTC	GGT CCA AAT TGT TGT CCC
Female_62654	CTA GCT GCC TGA AAG ATT C	CCC AAA TTA TAT TCC ACC TGC
Female_58030	AAT TCA ATG CAA TCA GCC TG	AGG CAA AGA AAA TTG GAA GC
Female_56104	AAT TCC CTG TCC AGT GAA G	TCA TGG CTG ATC TTA TTG TG
Female_54677	CTG TCC AAT GGA TTT CCC	GGG CAT ATT TAT AAA GCA GG
Female_12075	AAT TCA TCG GAC ACA GGA G	GTT ACA TAA ACT TGC CGT C
Female_9998	CTC TGT CCC AAA AGG TGC	CAT GGC TGA TCG ATG CC

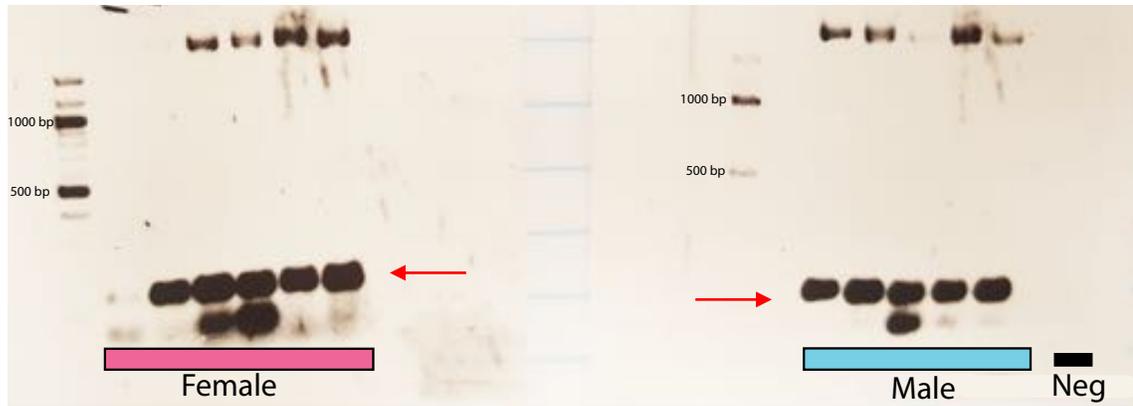


Figure 4.1 Male RAD marker 48396 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side of each gel. Red arrows indicate PCR product band for 3RAD marker of interest.

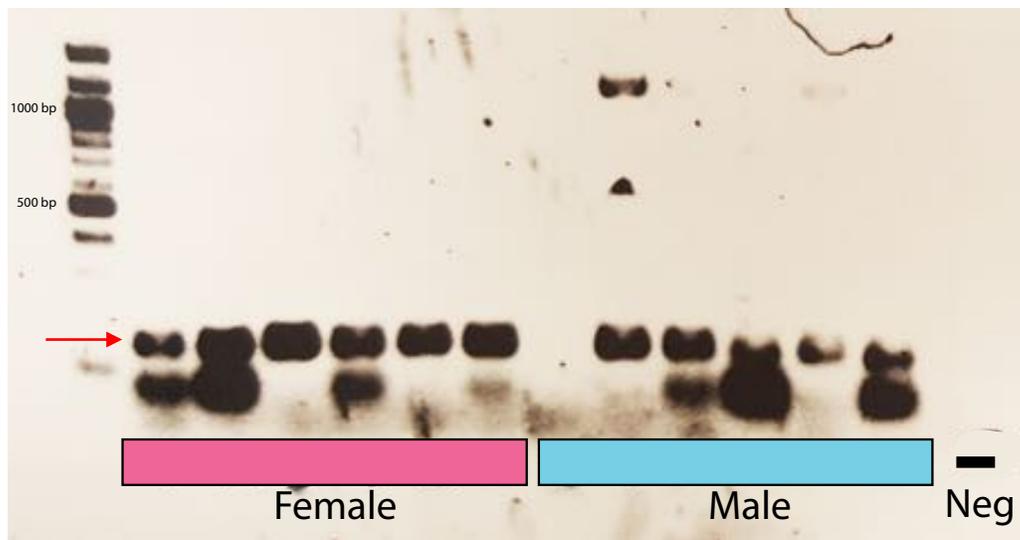


Figure 4.2 Male RAD marker 17488 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel. Red arrow indicates PCR product band for 3RAD marker of interest.

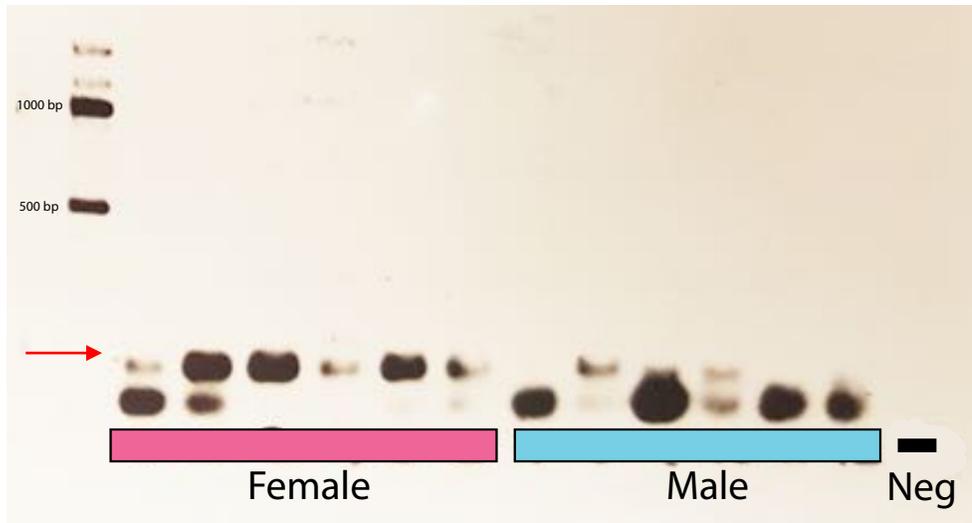


Figure 4.3 Male RAD marker 9873 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel. Red arrow indicates PCR product band for 3RAD marker of interest.

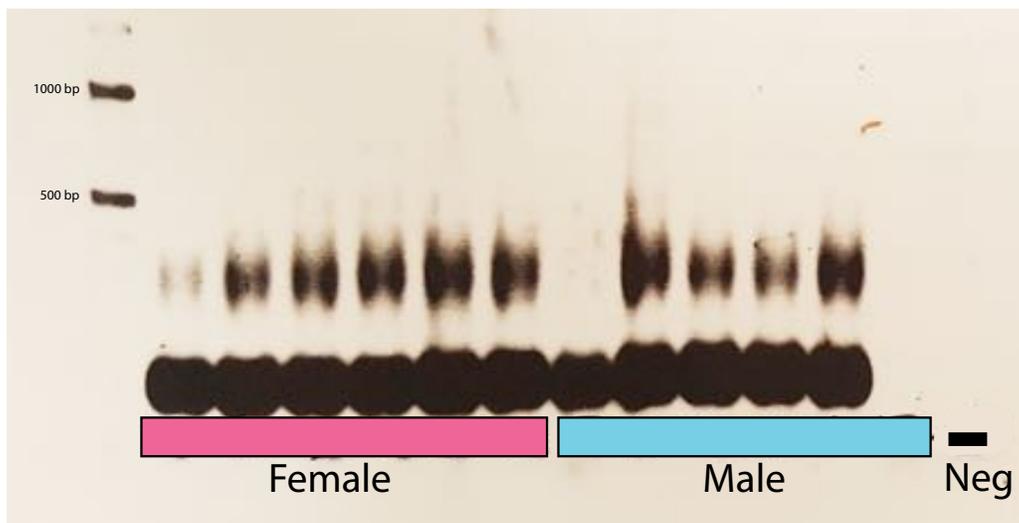


Figure 4.4 Male RAD marker 12965 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel.

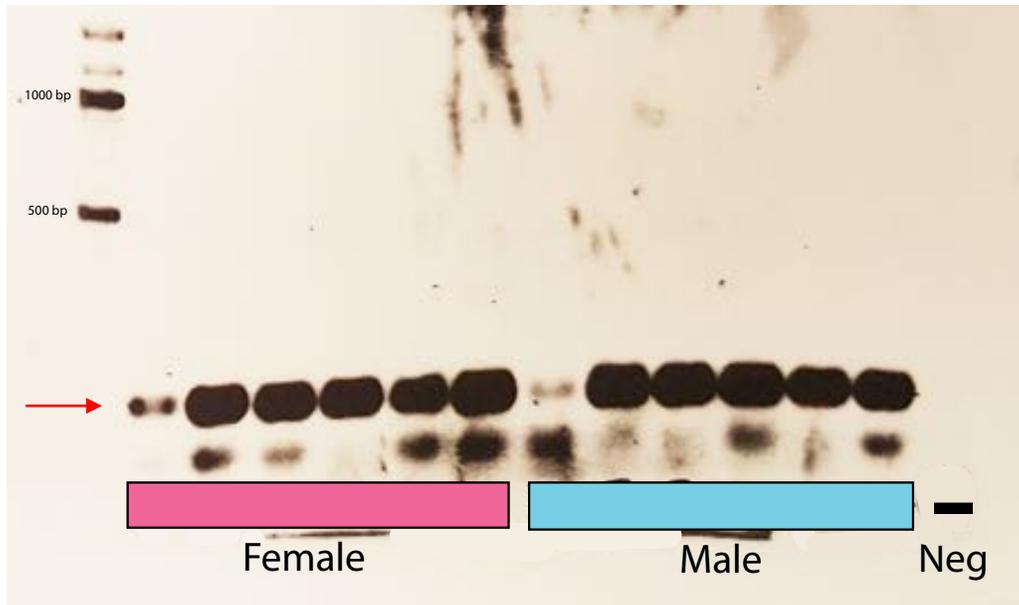


Figure 4.5 Male RAD marker 515 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel. Red arrow indicates PCR product band for 3RAD marker of interest.

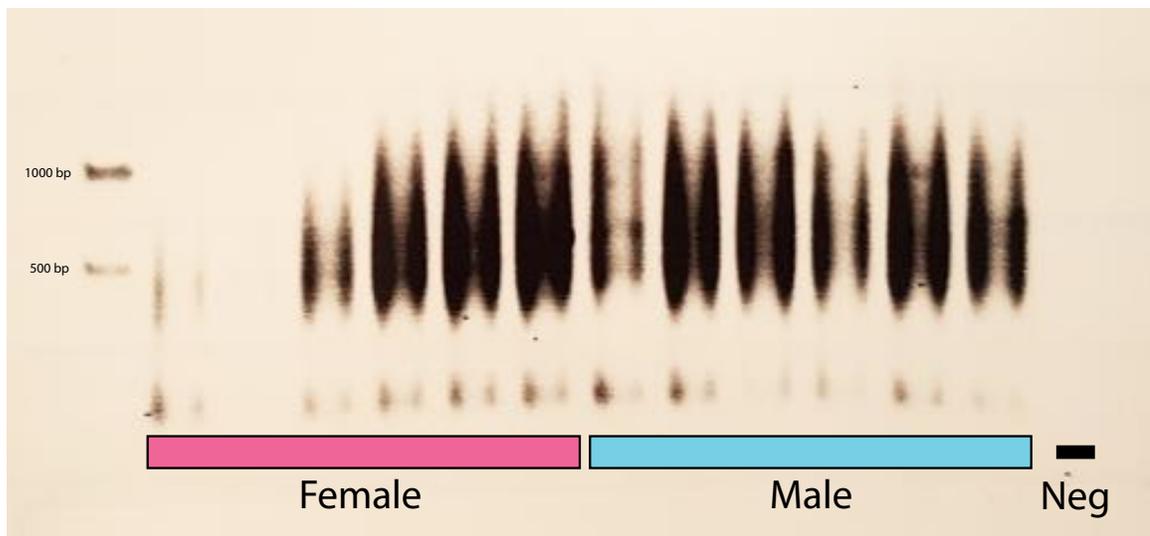


Figure 4.6 Female RAD marker 67066 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel.

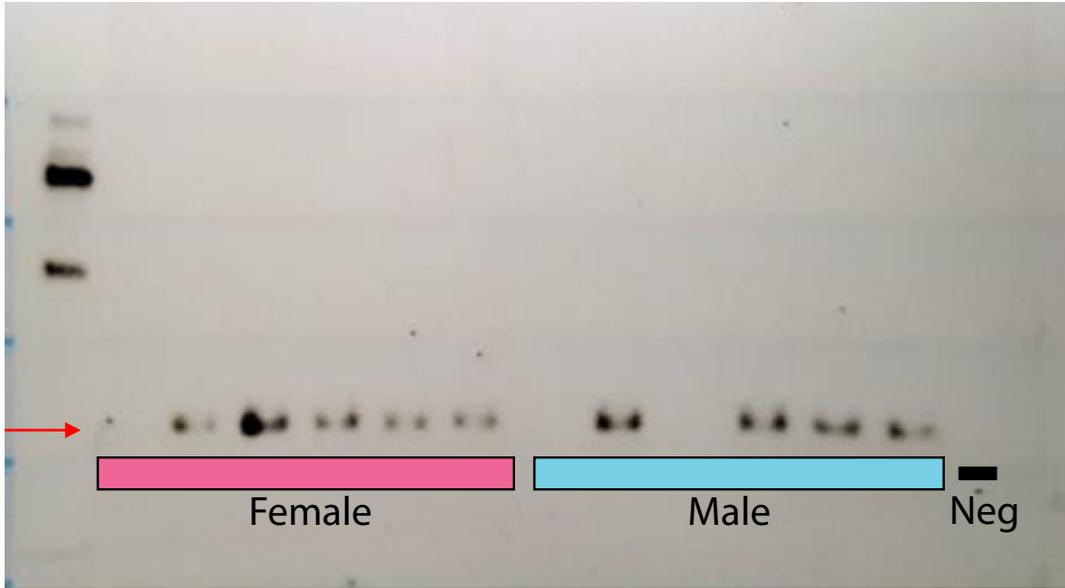


Figure 4.7 Female RAD marker 62654 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel (unmarked). Red arrow indicates PCR product band for 3RAD marker of interest.

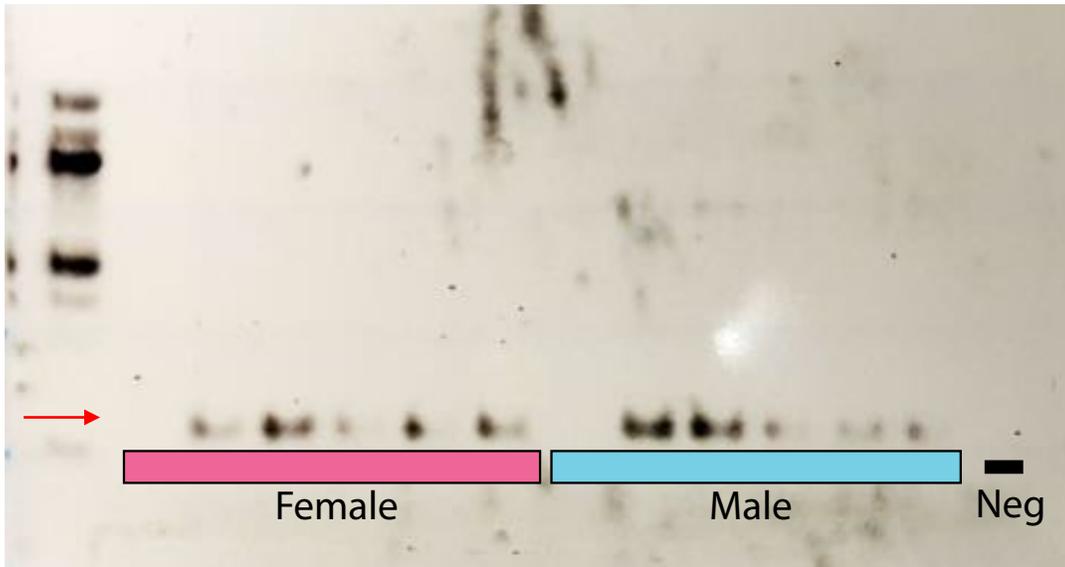


Figure 4.8 Female RAD marker 63104 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel (unmarked). Red arrow indicates PCR product band for 3RAD marker of interest.

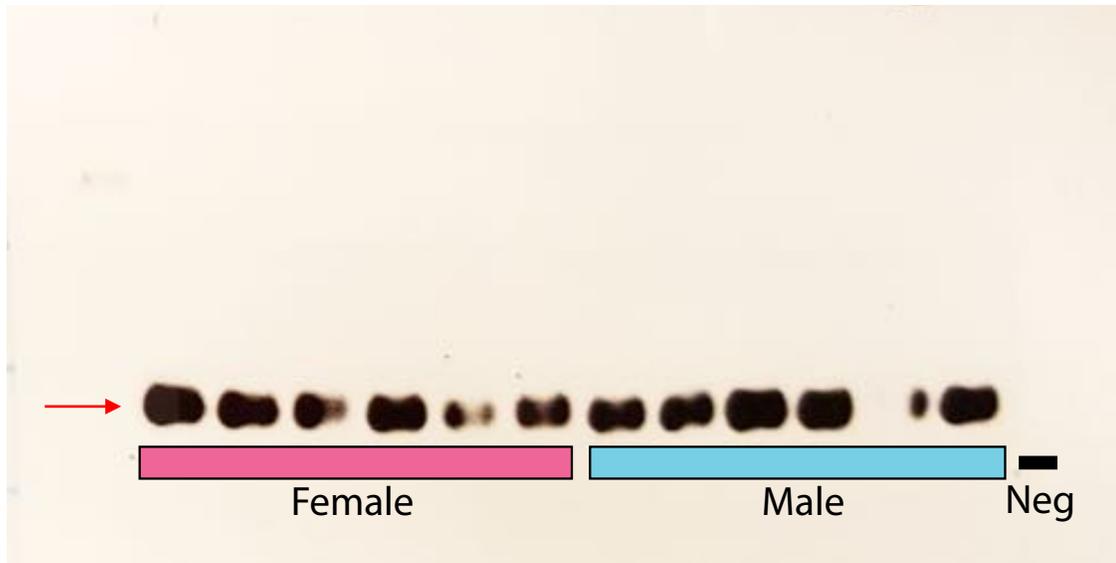


Figure 4.9 Female RAD marker 56104 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel (unmarked). Red arrow indicates PCR product band for 3RAD marker of interest.

of a low molecular weight band at approximately 140 bp, in both males and females. Some individuals of both males and females did not amplify, indicating the need for PCR optimization. Also, there is lack of high quality resolution of the DNA ladder that makes interpretation of the exact size of the DNA band in the images impossible. Female marker 56104 showed a strong band in both sexes (Figure 4.9), and a single low molecular weight fragment was amplified in female marker 54677 in both sexes, with one female and one male not amplifying (Figure 4.10), indicating that neither are sex-specific. The female-specific marker 9998 showed several non-specific bands, and what may be target amplification in both sexes (Figure 4.11). However, this validation is deemed inconclusive, and would require further PCR optimization and validation. Female RAD marker 66178 was amplified in both males and females (Figure 4.12), denoting that the marker is not sex-specific. The RAD marker 58030 had several non-specific bands present in both sexes, but no clear low target amplification (Figure 4.13). A new set of primers would need to be designed to explore this marker in more detail. Last, female marker 12075 had both non-specific bands, as well as a low molecular weight band around the target sequence size for both sexes (Figure 4.14), implying that this marker is also not sex-specific. Overall, many of the putative sex-linked RAD loci showed no sex-specificity through PCR validation. With a small sample size, it is possible that just by chance some RAD loci would be identified as sex-specific. Also, it may be that these identified RAD markers are false positives, not sequenced at a deep enough coverage to become a stack or loci. It would have been beneficial to assemble the paired-end reads of the putative RAD markers into contigs, and then design the primers based on these contigs. It may have provided better primer design locations, and a longer sequence to amplify, and thus, confirm if markers are sex-specific. Additionally, a sex-specific marker may lie within the RAD loci identified that showed presence in five of the six individuals of one sex, absent in

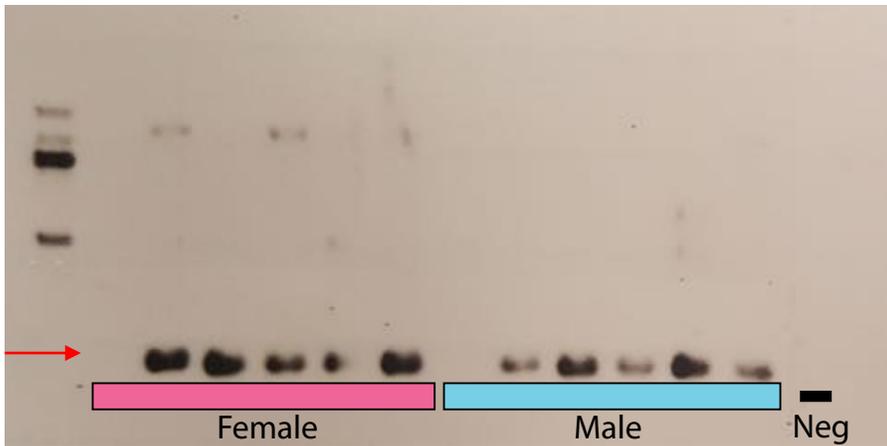


Figure 4.10 Female RAD marker 54677 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel (unmarked). Red arrow indicates PCR product band for 3RAD marker of interest.

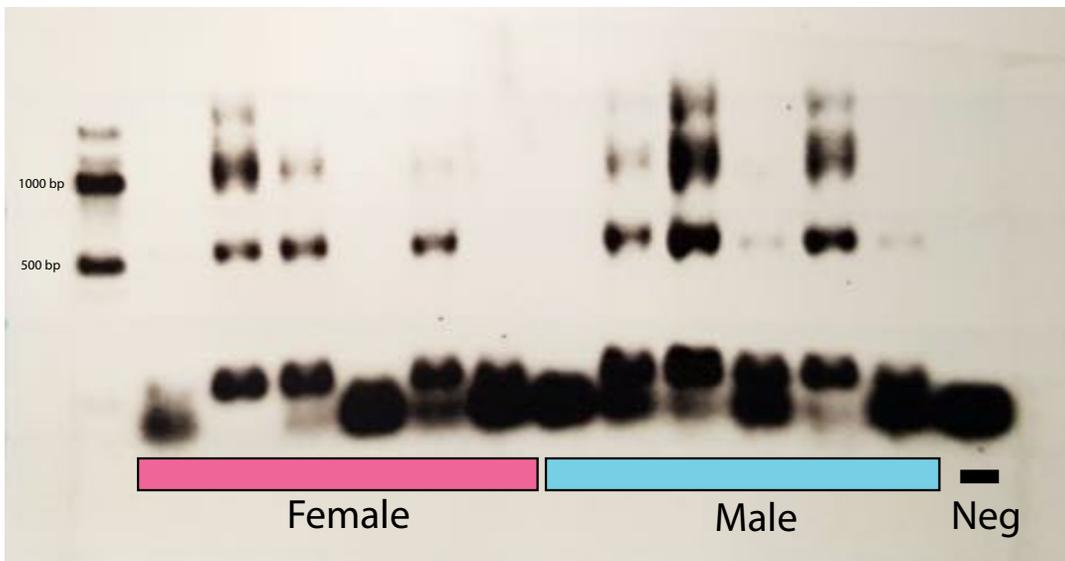


Figure 4.11 Female RAD marker 9998 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel.

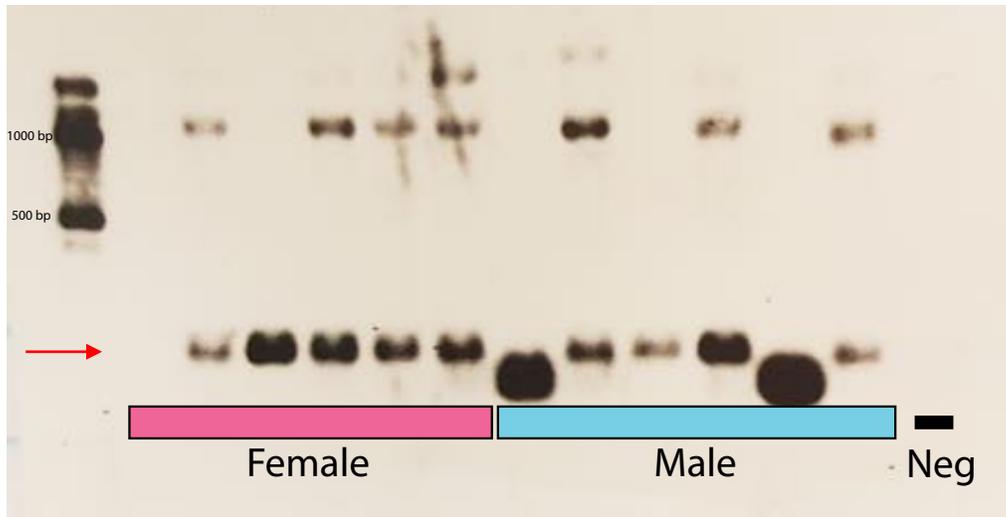


Figure 4.12 Female RAD marker 66178 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel. Red arrow indicates PCR product band for 3RAD marker of interest.

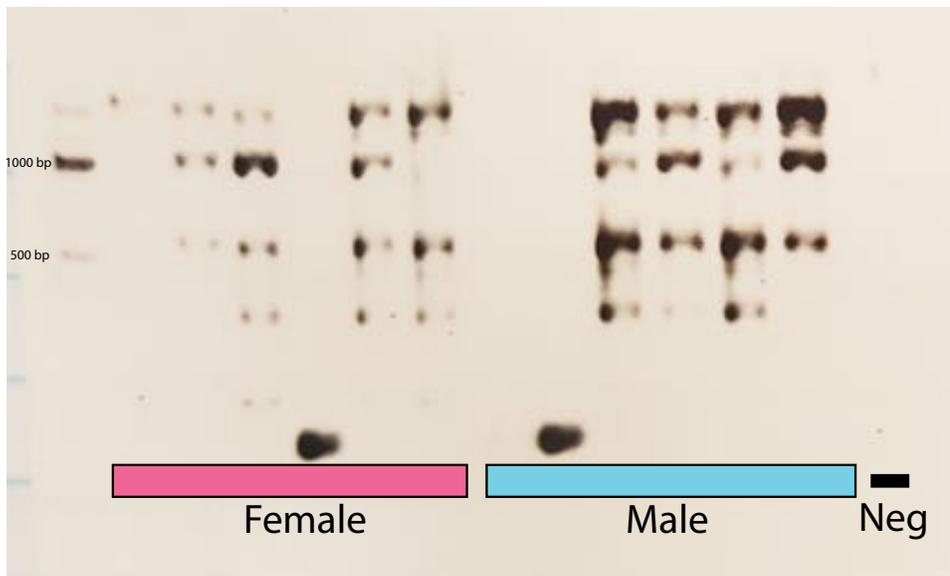


Figure 4.13 Female RAD marker 58030 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel.

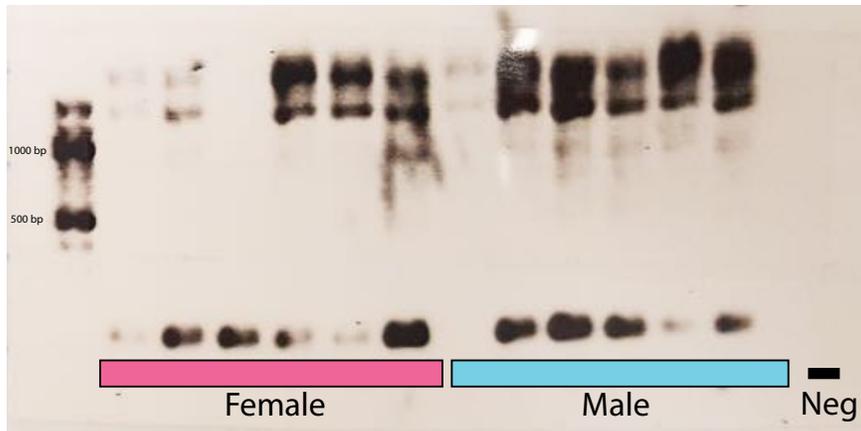


Figure 4.14 Female RAD marker 12075 PCR amplification in male and female *Carcharhinus plumbeus*. Female and male lanes are marked. Neg lane represents the negative control. 1000 bp ladder is present on the left side gel.

the other. However, time and resources did not allow for further validation of these potential RAD markers.

The 3RAD sequencing data analyzed via a custom bioinformatics pipeline produced different results than the Stacks pipeline. Paired-end reads were run through the pipeline individually, resulting in approximately 23,000 RAD loci for the *EcoRI* restriction site, and 25,000 RAD loci for the *NheI* restriction site. The slight difference in the number of RAD loci is likely due to the over splitting of some RAD loci with multiple SNPs. One putative male-specific marker was identified for both the *EcoRI* and *NheI* restriction site. No female-specific markers were identified. However, the male-specific marker failed the validation test. That is, the putative sex-specific reads associated with the marker were identified in the raw reads of the female library. They likely were not sequenced at a sufficient read depth in each of the females to assemble into a RAD locus. Thus, no candidate sex-specific markers were identified in *C. plumbeus* using the 3RAD method and custom bioinformatics pipeline analysis.

The Stacks pipeline results indicated nine female and five male-specific RAD markers, while the custom bioinformatics pipeline identified one male-specific RAD marker. There was no overlap in RAD marker identification between pipelines. The putative sex-specific RAD markers identified by the Stacks pipeline failed PCR validation, with presence in males and females. Also, the one male-specific marker failed the custom pipeline validation. The custom pipeline has been successfully used in several studies to identify and validate sex-specific markers (Gamble & Zarkower, 2014; Gamble et al., 2015; Gamble et al., 2017). Thus, based on this 3RAD analysis of *C. plumbeus*, no sex-specific markers were recovered.

There are several explanations as to why no sex-specific molecular markers may have been identified using the 3RAD protocol. First, the mechanism of sex determination is genetic, however, this particular protocol did not result in a sex-linked marker. The 3RAD library

consists of fragments that have two different restriction cut sites, within a particular size range, and this can greatly limit the number of RAD loci. It could be that the sex-determining region does not contain, or is not near, one of these restriction site regions. Another potential issue is the type of restriction enzyme. It may be that the sex-determining region contains a different restriction enzyme than used here. One potential reason that the pipelines indicated sex-specific RAD marker, but validation failed, is that the restriction site may be sex-specific, however, the flanking region is conserved in both sexes. Thus, the RAD loci would be amplified in both sexes. Further validation would be needed, by amplifying and then digesting the DNA, where the heteromorphic sex would produce two bands, and the homomorphic sex, only one band. Allelic differences in a region, and not presence/absence of a marker, could also be a potential reason why no sex-linked RAD markers were validated. In this case, the custom pipeline would not pick this up, as it only identified presence/absence of RAD loci between sexes. Another genetic mechanism that would not be picked up by this method is the presence of no sex-specific gene, but instead copy number variation of a sex-related gene on sex chromosomes. There would need to be differentiation in sex chromosomes, so that the homomorphic sex chromosomes would have the presence of two copies, and the heteromorphic sex chromosomes only one copy. It would be a dose-dependent concentration of the gene that would initiate sex differentiation. One example of dosage-dependent GSD is in the chicken, where it is speculated that dosage of a Z-linked gene, *DMRT1*, is responsible for sex determination (Nanda et al., 1999; Smith & Sinclair, 2004). If this were the case, a closer look at the karyotype would be warranted, to try and identify the heteromorphic chromosomes, and then further studies could be conducted to explore these sex chromosomes in more detail to identify the gene responsible. Lastly, another alternative explanation is that sex may be determined by environmental cues in this species, in which there would be no consistent genetic difference between sexes that could be identified. Future work

would benefit from utilizing different restriction enzymes, and using single-digest RAD-seq, where more RAD-tags can be sequenced.

Single-digest RAD-Seq

There was a total of 542,058,680 raw reads after filtering for ambiguous barcodes. The six libraries had similar number of raw reads and RAD-tags (filtered reads), except *M. randalli* males had a much greater total number of reads (Table 4.15). The *C. callorhynchus* female library had 87.3 million raw reads and 83.6 million RAD-tags, and males had 82.2 million raw reads and 80.1 million RAD-tags. The *C. carcharias* female library has 89.3 million raw reads and 88 million RAD-tags, and males had 92.5 million raw reads and 89.9 million RAD-tags.

Maculabatis randalli females had a total of 78 million raw reads and 76.8 million RAD-tags, while males had significantly more, with 112.4 million raw reads and 109.5 million RAD-tags.

Candidate sex-specific RAD loci were identified in each of the three species of chondrichthyans (Table 4.16). A total of 12 male-specific and two female-specific RAD loci were identified in *C. callorhynchus* (Table 4.17). These sex-specific loci were then screened against the raw reads of the opposite sex. If a match was found in the reads of the other sex, the RAD loci was excluded. Two candidate male-specific RAD loci passed the screening, while the two female-specific markers were excluded. *Carcharodon carcharias* had 19 putative male-specific RAD loci, and two female-specific RAD loci (Table 4.18). After screening against the reads of the opposite sex, only four male-specific RAD loci remained as candidate sex-linked markers, while no female markers were validated. The ray, *M. randalli*, had 108 putative male-specific RAD loci, and 10 female-specific RAD loci, but after validation, 71 male-specific markers and 6 female-specific markers remained as candidate sex-specific markers.

Table 4.15 Summary of sequenced individual RAD libraries on an Illumina HiSeq4000.

Lab ID	Species	Sex	# Raw reads	# Filtered reads
17406	<i>Callorhinchus callorynchus</i>	Female	11053562	10731097
17407	<i>Callorhinchus callorynchus</i>	Female	11104900	10912802
17408	<i>Callorhinchus callorynchus</i>	Female	11222868	11044126
17409	<i>Callorhinchus callorynchus</i>	Female	10898730	10721836
17411	<i>Callorhinchus callorynchus</i>	Female	9486794	9234427
17418	<i>Callorhinchus callorynchus</i>	Female	13010260	11655719
17419	<i>Callorhinchus callorynchus</i>	Female	11278872	10649223
17420	<i>Callorhinchus callorynchus</i>	Female	9326232	8688378
17405	<i>Callorhinchus callorynchus</i>	Male	13712244	13253081
17410	<i>Callorhinchus callorynchus</i>	Male	12333718	11978356
17412	<i>Callorhinchus callorynchus</i>	Male	12120016	11950398
17413	<i>Callorhinchus callorynchus</i>	Male	15099128	14825575
17414	<i>Callorhinchus callorynchus</i>	Male	15932236	15262611
17416	<i>Callorhinchus callorynchus</i>	Male	13100346	12859338
10688	<i>Carcharodon carcharias</i>	Female	9256998	9136576
17581	<i>Carcharodon carcharias</i>	Female	8105296	7977993
17589	<i>Carcharodon carcharias</i>	Female	10440542	10312866
17598	<i>Carcharodon carcharias</i>	Female	13069248	12937580
17602	<i>Carcharodon carcharias</i>	Female	15304978	15132761
17603	<i>Carcharodon carcharias</i>	Female	13724734	13397466
17608	<i>Carcharodon carcharias</i>	Female	10298544	10104470
17767	<i>Carcharodon carcharias</i>	Female	9149500	9079699
10687	<i>Carcharodon carcharias</i>	Male	8992904	8812418
17582	<i>Carcharodon carcharias</i>	Male	14301688	13356450
17590	<i>Carcharodon carcharias</i>	Male	10573448	10382637
17592	<i>Carcharodon carcharias</i>	Male	12246038	11967756
17597	<i>Carcharodon carcharias</i>	Male	11157090	10863493
17599	<i>Carcharodon carcharias</i>	Male	11880546	11685972
17605	<i>Carcharodon carcharias</i>	Male	12912382	12551289
17772	<i>Carcharodon carcharias</i>	Male	10441790	10319226
14282	<i>Maculabatis randalli</i>	Female	11383632	11267992
14297	<i>Maculabatis randalli</i>	Female	11468434	11345440
14329	<i>Maculabatis randalli</i>	Female	11950284	11804545
14410	<i>Maculabatis randalli</i>	Female	10580980	10491417
14423	<i>Maculabatis randalli</i>	Female	8475678	8429598
14442	<i>Maculabatis randalli</i>	Female	12617298	12106215
14446	<i>Maculabatis randalli</i>	Female	11582628	11380081
14165	<i>Maculabatis randalli</i>	Male	17529752	17228106
14426	<i>Maculabatis randalli</i>	Male	16606848	16177095
14428	<i>Maculabatis randalli</i>	Male	15621606	14910735
14430	<i>Maculabatis randalli</i>	Male	18971576	18175920
14434	<i>Maculabatis randalli</i>	Male	15761032	15410697
14443	<i>Maculabatis randalli</i>	Male	13481182	13366645
14445	<i>Maculabatis randalli</i>	Male	14492118	14301963

Table 4.16 Summary of RAD-Seq library analyses identifying candidate sex-specific molecular markers in *Callorhynchus callorhynchus*, *Carcharodon carcharias*, and *Maculabatis randalli*.

Species	Samples	Total # RAD Loci	Male-Specific RAD Markers	Female-Specific RAD Markers	# Candidate Male-Specific RAD Markers	# Candidate Female-Specific RAD Markers	Approx. Genome Size (Gb)
<i>C. callorhynchus</i>	6 males, 8 females	55966	12	2	2	0	1
<i>C. carcharias</i>	8 males, 8 females	196486	19	2	4	0	6.3 ¹
<i>M. randalli</i>	7 males, 7 females	169345	108	10	71	6	3.8 ²

¹ Schwartz & Maddock, 1986

² Based on average genome size of Dasyatidae species in Schwartz & Maddock, 2002

Table 4.17 Putative sex-specific RAD markers in *Callorhynchus callorhynchus* with the number of reads per individual. Highlighted RAD markers represent those markers validated by the bioinformatics pipeline, and represent candidate sex-specific markers.

RAD Marker	Males										Females																				
	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID	ID											
1877	17405	17410	17412	17413	17414	17416	17406	17407	17408	17409	17411	17418	17419	17420	156	231	0	397	126	99	0	0	0	0	0	0	0	0			
1886	131	122	104	0	135	123	0	0	0	0	0	0	0	0	131	122	104	0	135	123	0	0	0	0	0	0	0	0			
1887	85	108	99	93	123	71	0	0	0	0	0	0	0	0	85	108	99	93	123	71	0	0	0	0	0	0	0	0	0		
1888	65	73	68	80	89	92	0	0	0	0	0	0	0	0	65	73	68	80	89	92	0	0	0	0	0	0	0	0	0		
1889	86	104	91	134	118	88	0	86	104	91	134	118	88	0																	
1890	92	74	107	116	110	100	0	0	0	0	0	0	0	0	92	74	107	116	110	100	0	0	0	0	0	0	0	0	0		
1891	89	98	81	130	129	95	0	0	0	0	0	0	0	0	89	98	81	130	129	95	0	0	0	0	0	0	0	0	0	0	
1892	86	107	106	114	108	92	0	0	0	0	0	0	0	0	86	107	106	114	108	92	0	0	0	0	0	0	0	0	0	0	
1893	71	100	83	106	100	97	0	0	0	0	0	0	0	0	71	100	83	106	100	97	0	0	0	0	0	0	0	0	0	0	
1894	107	94	115	128	102	88	0	0	0	0	0	0	0	0	107	94	115	128	102	88	0	0	0	0	0	0	0	0	0	0	
1895	90	65	83	129	108	98	0	90	65	83	129	108	98	0																	
1896	193	157	206	181	213	222	0	0	0	0	0	0	0	0	193	157	206	181	213	222	0	0	0	0	0	0	0	0	0	0	0
22994	0	0	0	0	0	0	4	10	4	0	6	5	7	6	0	0	0	0	0	0	4	10	4	0	6	5	7	6	6		
23501	0	0	0	0	0	0	4	5	5	5	0	9	5	6	0	0	0	0	0	0	4	5	5	0	9	5	5	6	6		

Table 4.18 Putative sex-specific RAD markers in *Carcharodon carcharias* with the number of reads per individual. Highlighted RAD markers represent those markers validated by the bioinformatics pipeline, and represent candidate sex-specific markers.

RAD Marker	Males										Females									
	ID 10687	ID 17582	ID 17590	ID 17592	ID 17597	ID 17599	ID 17605	ID 17772	ID 10688	ID 17581	ID 17589	ID 17598	ID 17602	ID 17603	ID 17608	ID 17767				
8597	6	6	0	5	6	8	6	4	0											
8598	6	7	0	4	4	4	7	9	0	0	0	0	0	0	0	0				
8599	6	4	0	5	4	8	5	5	0											
8713	12	5	4	8	0	4	7	4	0	0	0	0	0	0	0	0				
8714	10	9	5	4	0	5	4	5	0	0	0	0	0	0	0	0				
8715	4	5	5	6	0	5	6	5	0	0	0	0	0	0	0	0				
8716	4	11	6	4	0	5	5	8	0	0	0	0	0	0	0	0				
8717	4	7	5	4	0	4	6	4	0	0	0	0	0	0	0	0				
8718	9	5	8	8	0	8	6	7	0	0	0	0	0	0	0	0				
8719	6	6	5	4	0	8	5	7	0	0	0	0	0	0	0	0				
8720	4	5	7	4	0	5	5	5	0	0	0	0	0	0	0	0				
8722	9	8	5	5	5	0	6	6	0	0	0	0	0	0	0	0				
8723	12	4	4	11	4	0	6	4	0	0	0	0	0	0	0	0				
8724	5	7	7	6	4	0	6	8	0											
8725	4	8	4	4	4	9	0	5	0											
8726	9	8	5	6	7	6	4	0	0	0	0	0	0	0	0	0				
8727	8	8	4	6	7	6	4	0	0	0	0	0	0	0	0	0				
8728	6	8	5	7	14	10	6	0	0	0	0	0	0	0	0	0				
8729	7	18	5	7	6	11	5	4	0	0	0	0	0	0	0	0				
25440	0	0	0	0	0	0	0	0	0	9	6	7	6	8	8	8				
34842	0	0	0	0	0	0	0	0	4	8	6	4	4	6	0	4				

The shark and ray species had significantly more total RAD loci than the elephantfish. This is likely a function of their genome size. The elephantfish has a genome size of approximately 1 Gb, while the white shark genome is about 6.3 Gb (Schwartz & Maddock, 1986). No genome information is available for the ray, *M. randalli*, however, other species in the family Dasyatidae with genome size information averages about 3.8 Gb. A larger genome would likely translate into many more restriction sites, and thus, more RAD loci. For example, an eight-base restriction site, like *SbfI*, would by chance occur randomly roughly every 65,536 bp (assuming equal base frequencies). Thus, with a genome of 1 Gb, it is expected that there would be approximately 15,258 *SbfI* restriction sites. This translates into 30,516 RAD loci, since each restriction site is flanked by two RAD-tags. The actual number of RAD loci found for the elephantfish was higher at 55,966. For the white shark genome at 6.3 Gb, approximately 96,130 restriction sites would be present by chance, making 192,260 RAD loci, very similar to the total number of RAD loci recovered.

Maculabatis randalli males had significantly more putative sex-specific RAD markers than the other two species, which corresponded to many more candidate sex-specific RAD markers. Male *M. randalli* had significantly more filtered reads than females (Table 4.15). While this did not translate into more total RAD loci compared to the other species (Table 4.16), this could be a potential explanation as to why there were more male-specific RAD markers. It is possible that the majority of these male-specific RAD markers are also present in females, but they just were not sequenced due to a bias during the library preparation or sequencing. Another possible explanation for more male-specific RAD markers, is that males may contain a large non-recombining region on their sex chromosome(s), which would correspond to many more sex-specific markers, compared to a small non-recombining region (Gamble et al., 2017). An alternative explanation could be an artifact of sample size. With small samples sizes, it is likely

that more putative sex-specific markers will be identified between the sexes, and that the majority of these will be false positives (Gamble et al., 2017).

The two *C. callorynchus* candidate male-specific RAD markers (marker 1889 & 1895) were assembled into contigs. Marker 1889 produced three contigs, while marker 1895 produced one contig (Table 4.19). Male-specific marker 1889, contig 1 was produced from the majority of the reads, for a total of 619 bp and a depth of coverage of 288 (Table 4.20). Contig 2 and 3 for marker 1889 both contained only two reads, each 151 bp in length. The male-specific marker 1895 contig of forward and reverse reads produced a locus of 563 bp and a mean coverage of 291 (Table 4.21).

The positive control of a 103 bp region of the mitochondrial *NADH2* gene was validated in the eight male and eight female *C. callorynchus* (Figure 4.15). Each potential primer pair for the respective RAD markers showed one strong band at the appropriate molecular weight (Figure 4.16 & 4.17). For RAD marker 1889, the forward 1 and reverse 2 were chosen, which amplify a 360 bp region (Figure 4.16). A larger amplicon was chosen for RAD marker 1895, so that it could be easily distinguished from marker 1889. It amplified a 519 bp region using forward 1 and reverse 1 primers (Figure 4.17). A test PCR showed co-amplification of both the RAD marker and positive control regions (Figure 4.18), indicating that the PCR conditions were optimal and there was no inhibition of amplification.

Candidate male-specific RAD markers for *C. callorynchus* were further validated to be associated with phenotypic sex through a PCR assay. RAD marker 1889 was amplified in all eight males, including two males that were not used in RAD-seq library generation (Figure 4.19). While this RAD marker was slightly amplified in some of the females, the bands were weak compared to the males (Figure 4.19). Thus, RAD marker 1889 is confirmed as a male-specific genetic marker in this species. The likely reason behind some weak amplification in females is

the fact that the X and Y (or Z and W) chromosomes were at one time homologous autosomes. The Y chromosome is therefore a 'degraded' version of the X chromosome. Consequently, some limited homology is not unexpected between the X and the sex-determining region of the Y chromosome (Gamble, Pers. Comm.). It is just likely that the sex-determining region of the Y chromosome has evolved over time to be dissimilar from its' X counterpart, for its function in sex determination. One way to test this hypothesis would be to perform another PCR, but increase the annealing temperature. It would be expected that if there is some limited homology, that as the annealing temperature increases, less amplification would occur due to less specificity between the primers and the female region, while the male amplification would remain strong. On the other hand, if it is a case of contamination by a male sample, increasing the temperature would not lead to a decrease of amplification. Additionally, a qPCR could be used to quantify the product in males and females. This could then be followed by additional qPCRs with increased temperature to better monitor the amount of product to determine if weak amplification in some females was the result of potential contamination or weak homology in females for this marker. RAD marker 1895 was amplified in all males, and there was no amplification in the eight females (Figure 4.20). Thus, RAD marker 1895 is also validated as a male-specific genetic marker in *C. callorynchus*. As seen in Figure 4.20, one male sample showed weaker amplification (ID 17417). A replicated PCR also resulted in weak amplification for that sample. This sample was not included in the RAD-seq library preparation due to poor quality DNA. This could have contributed to the weaker amplification, however, this sample showed strong amplification in RAD marker 1889. Another potential explanation is that the sample shows divergence in this male-specific marker. A small effective population size can lead to the fixation of alleles more

Table 4.19 Summary of *Callorhinchus callorhynchus* candidate male-specific marker contigs.

RAD Marker 1889	# Reads	Length (bp)	Coverage
Contig 1	1199	619	288
Contig 2	2	151	2
Contig 3	2	151	2
RAD Marker 1895			
Contig 1	1099	563	291

Table 4.20 RAD marker 1889 Contig 1 sequence.

```
TGCAGGATGGGAGGAAAGAGTTTGGACACAATTGAAACCATGACCTTATAAGCTG
AAGCACATTTCGCAGCTGGTTATATTTAATACTGCAAGCTTAAAATCACACAAC
TGCAGATTAATAAAAAATCCTCTTCCGCCTCAAAGAACTAATCATGGGATGTTCC
AAGGGTACGGTAGCACACTTTCCATTAGTTATCTGCATCGCCAATCGGTACATGGC
CATTTCGGCTGGGAGGACTACGAGTGAGAAATCACACCAGGGACTAAACTTTGT
GAACTCTGTCCAGATAACAAGCACTTTCTAGCAGCAGACAGCACAGCCAGGGTGAA
TTTTCTTTGGGCCCAATTAGCACTGAGACCAACTAAAGTTTGACTGCCTTGGCTGA
GATGATCAAATACAACAACAAAATTACATTTGTGTACCACATTTACATTGGGAGG
ACATCCCAAGATCCTGAACAATCAGTGTTCAGACCCAATCTAGGGGTCAGGAT
CGGGGAGGGTGGAGGCCAGGGAGGGTAGGGGCAGGAGGGTGACTGAAGGAGCAG
TTGAAGCAGAGGGTATTGAGATGGGTCTTGAAGGAAGAAAGGGATGGGGCGAGG
CAGAGAGACT
```

Table 4.21 RAD marker 1895 Contig 1 sequence.

```
TGCAGGGTTGATGAGGGGAGAACGGGAAGAGATGTGGTTACAGACTGGTAGGGCT
GACTGCTGTGCCAACACACCCTTAGCCTAATTAATAATAATAATAATAATCCT
TGCATTTGATATAGCGCTTTTCACGTCGGGAGGACGTCCCAAAGCGTTTCACAGAA
CACACTGTAAAGTGAATTAAGTGTATATTTGTGGGCGAACGCGGCAGCCAATTGCG
CACAGCAATGTCCACAAACAGTCGTAGATTGAAAATGACCAATTTATTATTGTTT
TTGGTTTTGGAGGATTATTATTAGGCTCTGTGTGGAGTAAGGAACTCGCCCGAGG
TTTGGCCCTGACCCCCGGGATTGAACTCAGGTCTCTCACTTGCAAAGCGAGTGC
TCTAACCCTGAGCTACAGGACTCCACAAATTGTAGAACAGATGCCTGGACGAGC
GATCCCCTCATGATCTCTGTTATAAACAACATGCTCAAGTCCATTCAATTATACCTC
ATGTAATAACAGTGCCCAAATAAGAACACACACTGCTTGTTAGGCCCTCACTCTG
CTCA
```

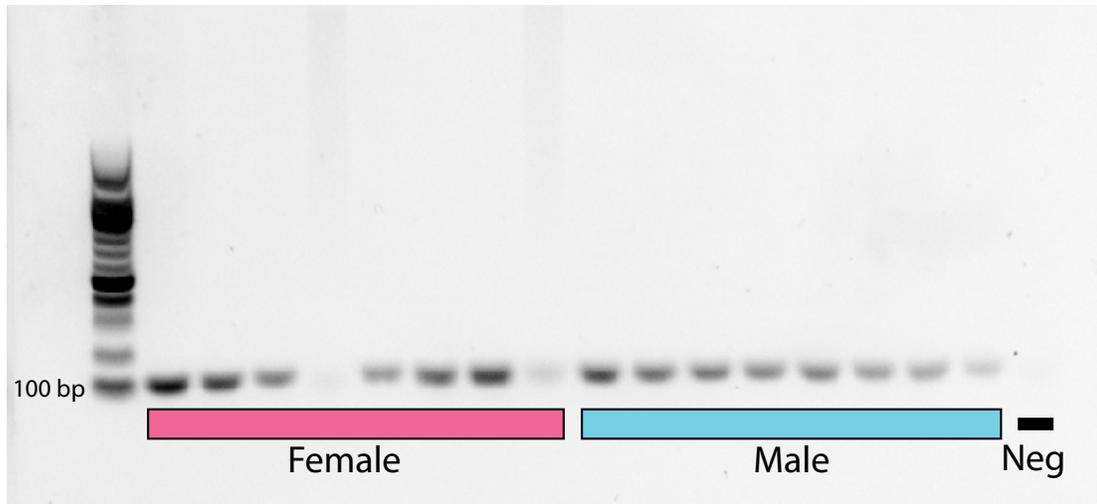


Figure 4.15 Positive control PCR amplification of a 103 bp region of the mitochondrial *NADH2* gene in male and female *Callorhynchus callorhynchus*. Female and male lanes are marked. Neg lane represents the negative control. 100 bp ladder is present on the left side gel.

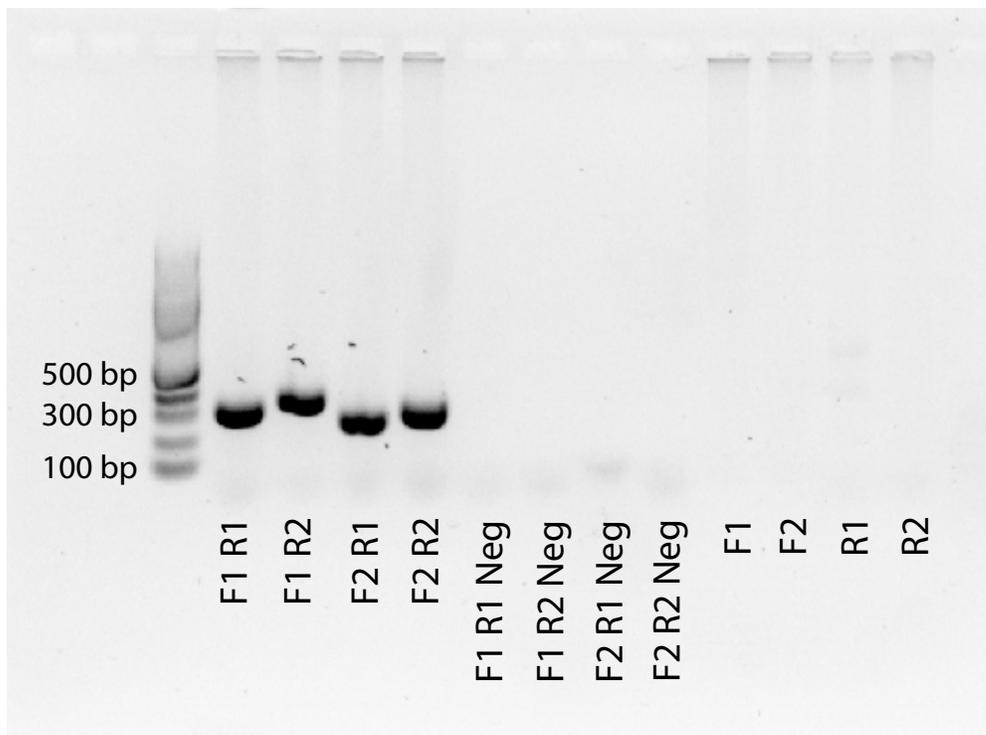


Figure 4.16 Candidate male-specific RAD marker 1889 primer validation PCR in one male (GN 17417) *Callorhynchus callorhynchus*. F1=CC_M1889_1_Forward. R1=CC_M1889_1_Reverse. F2=CC_M1889_2_Forward. R2=CC_M1889_2_Reverse. Neg lane represents the negative controls with respective primer pairs. 100 bp ladder is present on the left side gel.

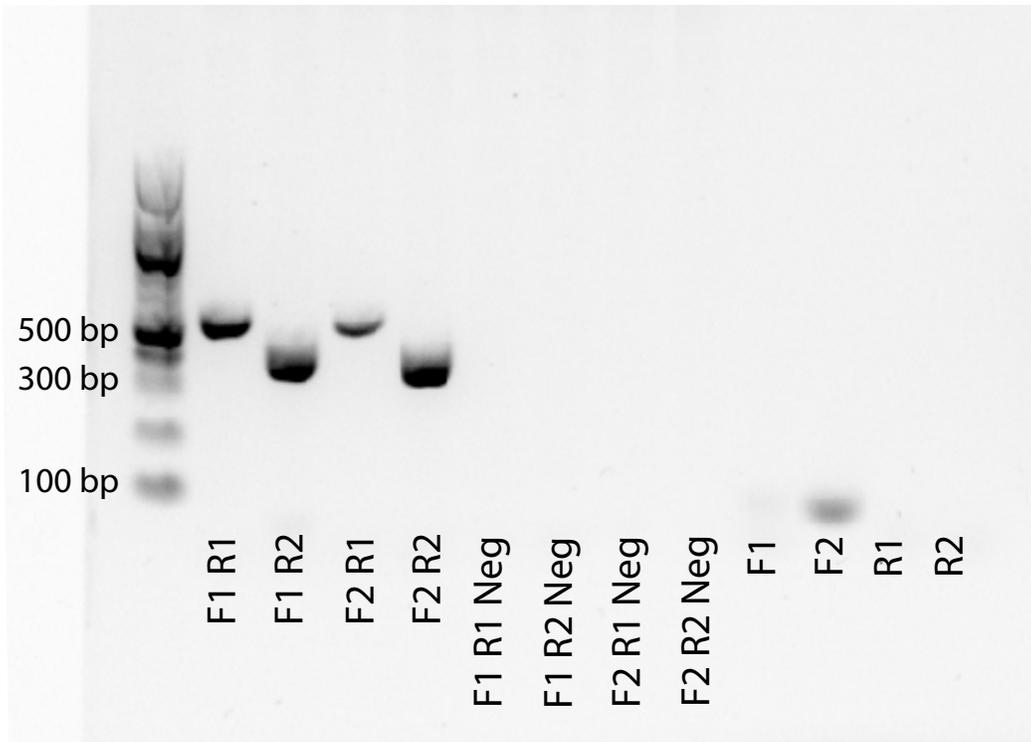


Figure 4.17 Candidate male-specific RAD marker 1895 primer validation PCR in one male (GN 17417) *Callorhynchus callorhynchus*. F1=CC_M1895_1_Forward. R1=CC_M1895_1_Reverse. F2=CC_M1895_2_Forward. R2=CC_M1895_2_Reverse. Neg lane represents the negative controls with respective primer pairs. 100 bp ladder is present on the left side gel.

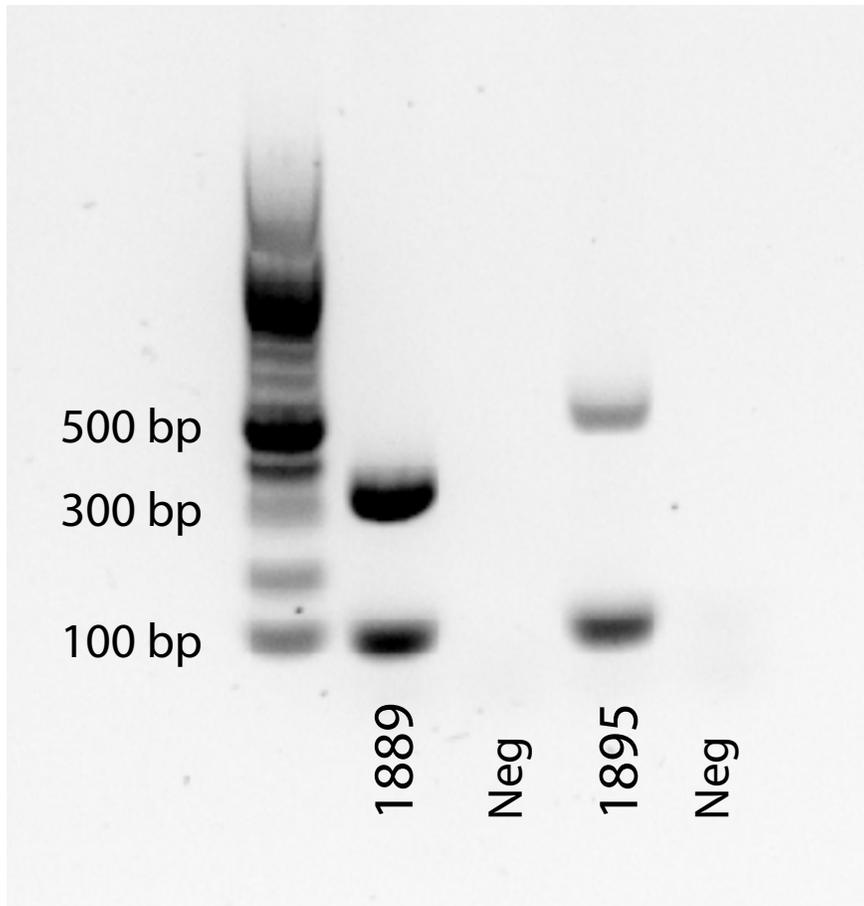


Figure 4.18 Candidate male-specific RAD marker primer and positive control validation PCR in one male (GN 17417) *Callorhynchus callorynchus*. RAD marker 1889 primers: CC_M1889_1_Forward and CC_M1889_2_Reverse. RAD marker 1895 primers: CC_M1895_1_Forward and CC_M1895_1_Reverse. Neg lane represents the negative controls with respective primer pairs. 100 bp ladder is present on the left side gel.

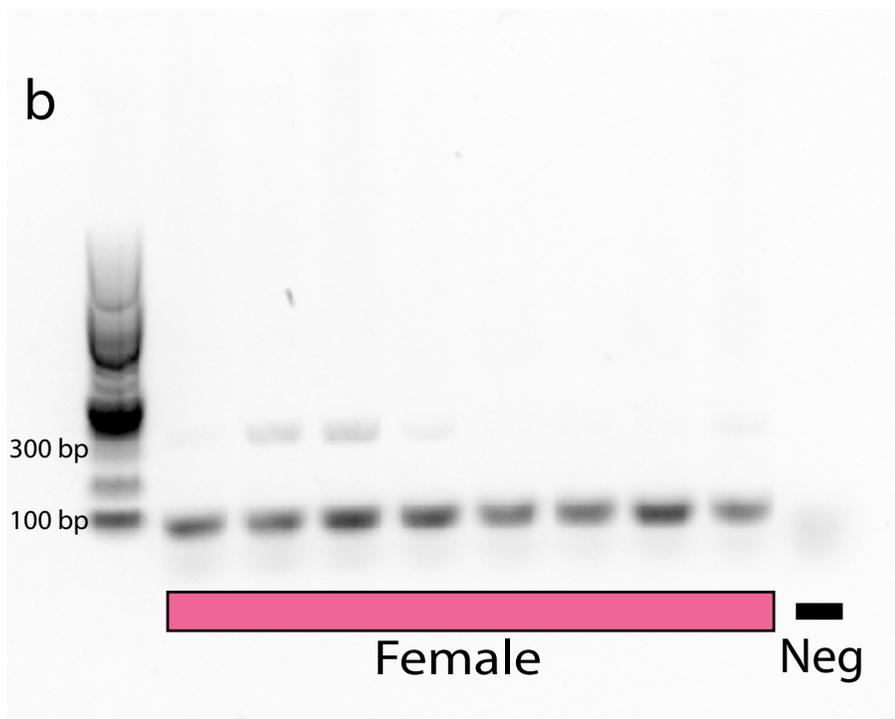
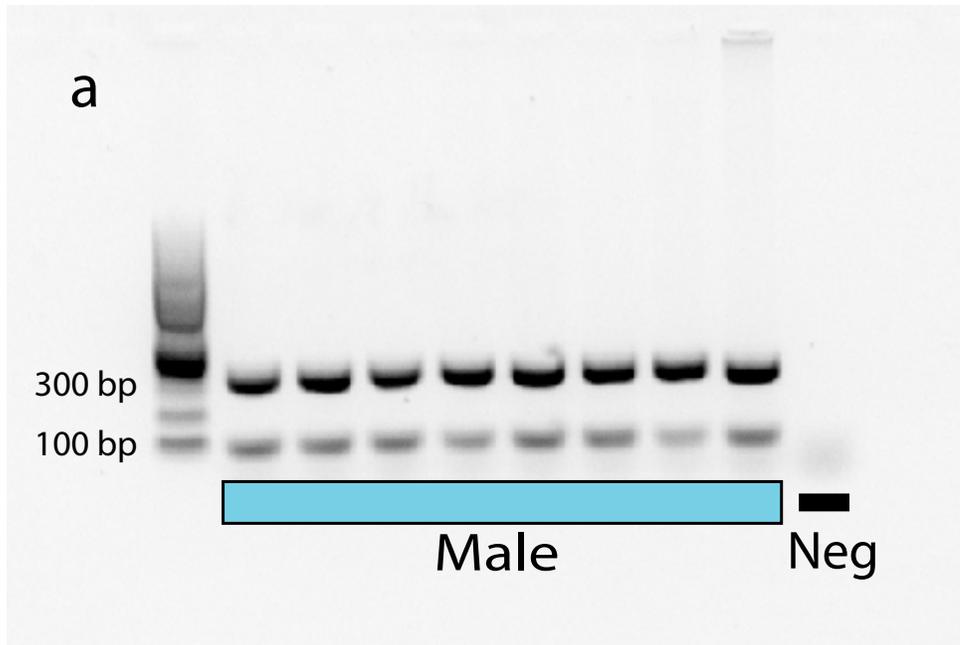


Figure 4.19 Candidate male-specific RAD marker 1889 PCR validation in a) eight male and b) eight female *Callorhinchus callorhynchus*. 103 bp positive control amplified in all samples. Neg lane represents the negative control. 100 bp ladder is present on the left side gel.

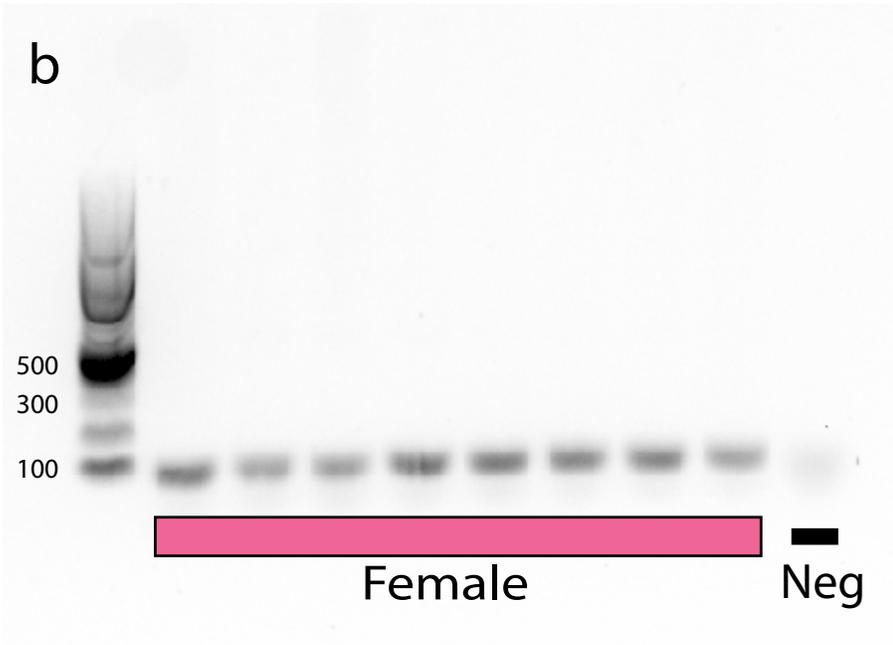
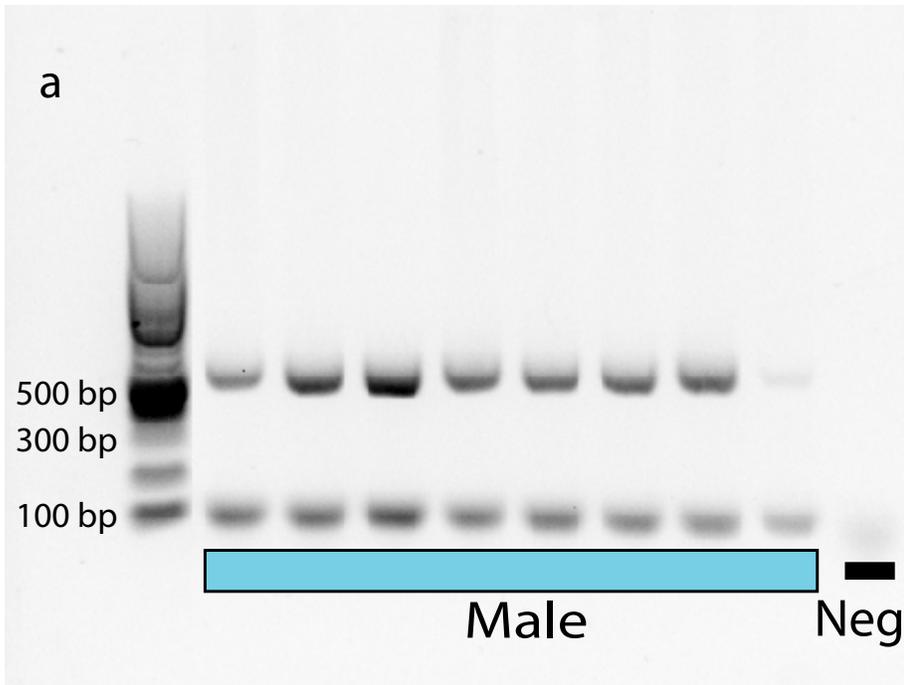


Figure 4.20 Candidate male-specific RAD marker 1895 PCR validation in a) eight male and b) eight female *Callorhynchus callorhynchus*. 103 bp positive control amplified in all samples. Neg lane represents the negative control. 100 bp ladder is present on the left side gel.

rapidly, and thus, divergence between or within a population. Overall, two male-specific genetic markers were identified in *C. callorynchus*, not present in female samples, consistent with male heterogamety and an XX/XY sex chromosome system.

The *C. callorynchus* male-specific RAD marker 1889 contig 1 was BLAST against the *C. milii* genome assembly 6.1.3. There was 100% coverage and 99% identity to a 619 bp region of scaffold 2177 (GPS_003800317.1). There was a total of two mismatches between the two sequences. The unplaced genomic scaffold 2177 consists of 10238 bp, of which RAD marker 1889 starts at 1745 bp to 2363 bp. The only annotations available for this scaffold are a gap between 1392 to 1696 bp, and a putative gene at 2464 to 3351 bp, with a CDS from 2464-2580 with a protein product induced myeloid leukemia cell differentiation protein Mcl-1 homolog, based on automated computational analysis and transcript data. The RAD marker 1889 was analyzed for potential ORFs, of which four were found. They were found at: 226-80 bp (147 bp/48 amino acids total), 161-271 bp (111 bp/ 36 amino acids total), 156-251 bp (96 bp/31 amino acids total), and 449-366 bp (84 bp/27 amino acids total). A BLAST for the protein sequences did not recover any significant matches. A search of the scaffold 2177 found 64 potential ORFs.

The *C. callorynchus* male-specific RAD marker 1895 contig 1 was also BLAST against the *C. milii* genome assembly of a male specimen. There was 100% coverage of the 563 bp region to scaffold 1466 (GPS_003799606.1), with nine gaps and two mismatches. Scaffold 1466 contains 14,834 bp, and RAD marker 1895 begins at 7160 bp to 7724 bp (with the 9 gaps). There are several annotated gaps, and two putative genes, one at the complement of 366-3975 bp, and the other at complement of 5219-13113 bp, both with a predicted protein product identified as histone-lysine N-methyltransferase ASH1L-like, based on automated gene prediction. The ORF finder identified two potential ORFs within RAD marker 1895. One ORF ranged from 490-86 bp for a total of 405 bp and 134 amino acids. The second potential ORF ranged from 116 to at least

3 bp for 114 bp and 37 amino acids. Neither protein search resulted in a match using BLASTP. An overall search of scaffold 1466 identified 103 potential ORFs. It is possible that these male-specific contigs are non-coding sequences with regulatory functions related to the sex-determining pathway, and thus, would not result in protein matches. Next steps would be to check contigs for domains and motifs (e.g., structural and regulatory), upstream regulatory sequences, repeats, nuclear localization signals, and signal peptides.

The four male-specific candidate RAD markers in *C. carcharias* were assembled into contigs from their respective forward and reverse reads. Male RAD marker 8597 assembled into two contigs, contig 1 contained only the forward reads from the restriction site for a total of 110 bp. Contig 2 contained only the reverse reads for 580 bp. Similarly, male RAD marker 8599 had two contigs, contig 1 was only the forward reads with 110 bp, and contig 2 was only the reverse reads for 559 bp. Male-specific RAD marker 8724 also had two contigs, contig 1 contained forward reads from the restriction site at 110 bp, and contig 2 was compiled from the reverse reads for a total of 523 bp. Last, male RAD marker 8725 assembled into one contig containing both forward and reverse reads for a total of 334 bp. The lack of overlap in the first three RAD markers between forward and reverse reads indicates that these fragments were larger than the sequenced 300 bp (150 bp paired-end). Additionally, the number of reads were also quite low for each of these markers, which likely contributed to the lack of overlap between reads (Table 4.18). Each contig was aligned against the two *C. callorhynchus* candidate sex-specific markers, and no similarity was observed. The contigs were BLAST against the *C. milii* genome, with no significant matches. Given that only male-specific RAD markers were recovered, this is consistent with a male heterogametic XX/XY sex-determining mechanism in this species. These two species could have evolved different sex-specific regions, yet both have a male heterogametic system. It would be interesting to test the *C. callorhynchus* male-specific RAD marker primers in

males and females of *C. carcharodon*, to see if this region is conserved, and was either not sequenced, or there may be divergence in the restriction site in *C. carcharodon*. On the other hand, it may be present in both sexes, and thus, not sex-specific in *C. carcharodon*. A recent study by Gamble et al. (2015) showed that in gecko species, there are many transitions among male and female heterogamety, with different sex-specific markers identified by RAD-seq. Also, the sex-determining gene, *dmy*, of medaka, was not found to be present in other species of *Oryzias* studied (Kondo et al., 2003; Kondo et al., 2004; Tanaka et al., 2007; Myosho et al., 2012). Thus, it would not be uncommon for species as divergent as *C. callorhynchus* and *C. carcharodon* to have evolved different sex chromosomes.

Identification of male-specific genetic markers in *C. callorhynchus* is consistent with a male heterogametic XX/XY sex chromosome. The identification of candidate male-specific markers in *C. carcharodon*, and an excess of male-specific markers compared to female markers in *M. randalli*, also are consistent with a genetic mechanism and male heterogamety. However, the candidate RAD markers for the shark and ray species will need to be validated in future work to verify they are sex-specific. The validated sex-specific markers in *C. callorhynchus* both mapped to two different scaffolds in the *C. milii* genome assembly. Either these scaffolds are both present on the Y chromosome, or potentially, the scaffolds may lie on two different sex chromosomes. The *C. milii* genome assembly will be invaluable for future studies to identify the complete sex-determining region and sex-determining gene. This is the first study to identify and validate sex-specific markers in a chondrichthyan species, as well as identify candidate sex-specific markers in a shark and ray species. Previous work has centered around cytogenetic evidence, which can be difficult to obtain in this group of fishes. Also, while potential male and female heterogamety has been observed using karyotypes in some chondrichthyan species, other species have shown homomorphic chromosomes. In these cases, when cytogenetic techniques are not feasible or

homomorphic chromosomes exist, the use of a method like RAD-seq to identify sex-linked markers is vital to establishing the sex-determining mechanism. Furthermore, validated sex-specific markers and primers can be used in a PCR-based assay to sex individuals with unknown sex information.

Conclusions

Cytogenetic approaches were unsuccessful in producing metaphase chromosome preparations for karyotyping and subsequent fluorescent *in situ* hybridization (FISH). While the methods used in these experiments did not produce results, other studies have provided karyotypes of chondrichthyans species. Cytogenetics can be an invaluable tool for exploring potential sex chromosomes and sex-determining mechanisms in animals. Not only can it provide evidence of heteromorphic sex chromosomes and thus, the heteromorphic sex, but it can also be used in experiments to probe the genome for particular genetic markers, such as those that are putatively sex-linked. Using FISH (Heiskanen et al., 1996) can provide physical mapping of a particular genomic sequence to determine if it is located on a sex chromosome. However, cytogenetics does have its limitations. First, for non-model organisms, producing quality karyotypes can be difficult due to the lack of established protocols. Secondly, karyotypes only produce indirect evidence of a genetic mechanism of sex determination. Last, if no heteromorphic chromosomes are visualized in a karyotype, this does not mean that sex is not determined genetically. In this instance, when species have homomorphic chromosomes, cytogenetics will fail to show which sex is truly 'heteromorphic', in those cases where the species has a genetic sex-determining mechanism. Other methods would need to be used to study how sex is determined, like searching for sex-linked genetic markers, or using breeding experiments.

Several species of elasmobranchs have putatively been found to have homomorphic chromosomes, but some species have also been shown to have heteromorphic, putative sex chromosomes. Thus, while difficult to produce here, cytogenetic techniques should still be pursued. It would provide further evidence of heterogamety, if found, as well as an important approach for identifying genomic marker locations, especially with regards to a sex-specific gene. Future works should aim to establish a common protocol for chromosome preparation, that does not necessarily rely on injecting specimens, holding them in captivity for a period of time, and sacrificing. Also, it would be valuable to determine an approach to stimulating cell proliferation in cell culture samples to provide adequate cells in metaphase.

A double digest approach to RAD-seq (3RAD) was initially used to investigate sex-specific markers in a shark species. However, those results revealed no sex-linked markers, as validation both through a bioinformatic pipeline and PCR assay failed. There were significantly fewer raw reads and RAD loci per individual using the 3RAD preparation compared to the single-digest RAD method. A limitation of ddRAD is that it restricts the number of fragments due to the fact that two enzymes are used and only fragments with both restriction sites are carried through the protocol, coupled with size selection. Thus, this likely had an impact on the resulting number of RAD loci identified, and lack of sex-specific markers. The sex-determining region would have to include a region that is flanked by both restriction sites, and is within the size selection range. While ddRAD has its advantages, like the ability to be highly customizable with different enzymes, fragment sizes, and general ease of the library preparation, the use for identification of sex-specific markers, especially in species that may have quite small non-recombining regions, may not be ideal. There are some additional alternative explanations given that no sex-specific markers were identified in this species that should be noted. First, the sex-determining mechanism is genetic, however, the two restriction enzymes used are not found within the sex-

determining region, in which other enzymes would need to be evaluated. Second, the restriction site may be sex-specific, however, the flanking region, which is ultimately used for PCR validation is conserved, and thus would amplify in both males and females. Here, a PCR would need to be carried out followed by restriction enzyme digestion, which would identify two fragments in the heterogametic sex, and one fragment in the homogametic sex. Third, it may be that both sexes carry the sex-determining region, however, the homogametic sex has a higher copy number, and sex is determined based on dosage. Last, on the other end of the spectrum, sex may be determined by the environment, however, evidence of heterogamety in some species through karyotypes and parthenogenesis data point towards GSD. But, the influence of environment cannot be ruled out at this point.

Single-digest RAD-seq was used to subsample the genomes of three species, a chimaera, shark, and ray, to identify candidate sex-specific markers that would indicate a genetic mechanism of sex determination, as well as the sex chromosome system. Validated male-specific genetic markers were identified in *C. callorhynchus*, which is consistent with GSD, and an XX/XY sex chromosome system. While the shark and ray species did not have their candidate sex-specific markers validated, the excess of male-specific markers compared to female-specific markers suggests that these two species also have GSD, and likely an XX/XY sex chromosome system. The two male-specific RAD markers in *C. callorhynchus* were found to map to two separate scaffolds in the *C. milii* genome. No similarity was found between the male-specific RAD marker contigs of *C. callorhynchus* and *C. carcharodon*, and there were no matches in the *C. milii* genome for any of the four male-specific RAD markers of *C. carcharodon*. This suggests that these two species may have evolved different sex chromosomes, and thus, different sex-determining regions and genes. However, future studies will need to validate the white shark sex-specific markers, and explore these further.

RAD-seq was chosen as the method to identify sex-specific markers in the genome, as it provides a way to subsample the genome, reducing a very complex genome into a subset of fragments. Hundreds of thousands of genetic markers can be identified using RAD-seq, in lots of individuals, in a short amount of time (Baird et al., 2008; Davey et al., 2011; Etter et al., 2011). It is relatively inexpensive and provides deeper coverage compared to whole genome sequencing (Rokas & Abbot 2009; Hohenlohe et al. 2010; Emerson et al. 2010; Etter et al., 2011; Andrews et al., 2016). While whole genome sequencing of males and females would produce clear evidence of sex-specific genomic differences, it is very time and resource expensive. RAD-seq is also advantageous due to the many restriction enzymes available. Thus, if one enzyme does not work, others can be tried. Here, *SbfI* was chosen as a starting point because it cuts relatively infrequently, with the expectation that it would provide a good balance in terms of the number of RAD loci and coverage of those loci, for the number of individuals included in the library preparation. But in the case that *SbfI* did not produce any sex-specific RAD markers, a more frequent cutting restriction enzyme could have been used, which would provide many more RAD loci and thus, potential sex-specific markers. Another great advantage is the fact that this method can be used for both model and non-model organisms, and no previous genomic data is necessary (Baird et al., 2008; Hohenlohe et al., 2012; Emerson et al., 2010; Hess et al., 2013). Single-digest RAD-seq uses random shearing, and when coupled with paired-end sequencing can be useful in assembling reads de novo into long contigs. Longer contigs provide more sequence for primer design and searching against sequence databases. Another method that can be used to identify potential sex-linked genetic markers is amplified fragment length polymorphism (AFLP). RAD-seq is advantageous over AFLP as the putative sex-linked amplified markers would need further amplification, cloning, and sequencing in order to identify the markers sequence and create PCR

primers to validate the markers, whereas with RAD-seq, primers can easily be created from RAD markers (Griffiths & Orr, 1999; Gamble & Zarkower, 2014; Gamble, 2016S).

As with any DNA library preparation and next-generation sequencing method comes some limitations that must be understood. RAD-seq only represents a small portion of the genome, and even if sex-linked markers exist, they may not be found using restriction enzymes. However, the expectation is that at least one type of restriction site lies within a sex-specific region, which would be sequenced by the method, but may take lots of trial and error with restriction enzymes. The library preparation protocol can be technically challenging, and may need to be altered on a per species basis. Also, this method requires the use of high quality DNA, which is not always available for many species (Bergey et al., 2013). Read lengths tend to be short, for example, 150 bp reads, and even with paired-end reads, comes out to 300 bp. However, with RAD-seq, shearing of fragments can help to provide an array of fragment sizes from the same restriction site to be sequenced, and these can be assembled to make longer contigs. A major limitation of the ddRAD method is the requirement for two restriction sites and size selection, which may limit the number of RAD loci, and relies on the digestion of both enzymes. In the end, next-generation sequencing produces an abundance of data, which requires the proper resources to store and manage, as well as bioinformatics to analyze the data, which is not always available. A major limitation of RAD-seq, which is more so of a problem for population genetics, is allele dropout, where a polymorphism is present in the restriction site, which is not cut by the enzyme and subsequently sequenced. Here, if one sex has a polymorphism in the restriction site, and the other does not, it is sex-specific, and would show presence/absence using the bioinformatics pipeline. However, downstream validation would depend on whether the flanking region of the restriction site is conserved, and if so, would not be validated through PCR. Also, if there are sex-specific alleles present, validation by PCR would require them to be distinct in size.

Alleles that are similar in size, yet differ in SNPs or short indels, would require additional sequencing of PCR products to validate sex-specificity. Some general limitations of PCR and sequencing that can affect the resulting RAD alleles include: GC bias, where fragments with higher GC content may be amplified less, resulting in less coverage, which may not be enough for bioinformatics pipelines, such as the one used here, to form a loci or allele (Davey et al., 2013); restriction fragment size bias, where there is preferential amplification of shorter fragments, which can affect coverage (Davey et al., 2013).

The results from this research provide a molecular framework for further exploration of the sex-determining mechanisms and regions among chondrichthyan fishes. Future work should first aim to validate the sex-specific markers in *C. carcharodon* and *M. randalli* identified by RAD-seq. This would provide additional evidence that these two species, along with *C. callorhynchus*, all species highly divergent from one another, have an XX/XY sex chromosome system. Secondly, the validated markers in the respective species can then be used in a PCR-based assay to assess the presence or absence of the sex-specific markers across chondrichthyan species. Since these are sex-specific markers, they likely reside in the sex-specific region of the Y chromosome. The *C. callorhynchus* male-specific markers were mapped to two scaffolds in the *C. milii* genome. These scaffolds can be explored in more detail to identify the sex chromosomes, the full sex-determining region, and ultimately the sex-determining gene. The availability of sequence data provided by the scaffolds can be used, via primer walking, to explore the presence and absence in males and females to define the sex-determining region. Once the *C. carcharodon* sex-linked markers are validated by PCR, its' draft genome can also be used to map the markers and identify the sex chromosome linkage group. These results can then be used to explore similarities and differences between the sex chromosomes of *C. carcharodon*, *C. callorhynchus*, and other known vertebrates. Another important future goal should be to produce high quality

karyotypes of these species, and then use FISH to visualize the location of these sex-specific markers on the sex chromosomes. The eventual objective would be to define and characterize the master sex-determining genes in these chondrichthyan species and providing definitive evidence that the gene is both necessary and sufficient for the determination of sex. To do this, one must first identify expression of the gene in males and females throughout development, before, during, and after sex differentiation; additionally, it should be shown that overexpression of the gene, in the case of a male-determining gene, would induce testis formation in genetic females, and inactivation of the gene in genetic males produces XX females.

CHAPTER 5

DISCUSSION

The first goal of this dissertation was to estimate the evolutionary history of holocephalan fishes by providing a comprehensive sampling of taxa, and using two independent data sets, to explore congruence, incongruence, and provide confidence in the estimated phylogeny. Holocephalans occupy a critical phylogenetic position in the jawed vertebrate tree of life as they, collectively with elasmobranchs, are the sister group to bony vertebrates. They are commonly assumed to be the extant clade, along with elasmobranchs, that approximates the ancestral jawed vertebrate condition (Janvier, 1996; Neyt et al., 2000; Venkatesh et al., 2001; Chiu et al., 2002, 2004; Tanaka et al., 2002; Amores et al., 2004; Kikagawa et al., 2004; Robinson-Rechavi et al., 2004; Mulley et al., 2009). Thus, the basal holocephalan clade has the potential to be instructive in evaluating hypotheses about early jawed vertebrate evolution and the origin of jawed vertebrate diversity in a comparative framework.

Phylogenetic trees can be powerful tools as they can be used to explore patterns of change over the course of evolution for genes, proteins, traits, behaviors, etc. of interest. In this regard, we can infer when and where a particular change likely occurred, and then the process can be studied in more detail to determine what changes occurred, giving insight into the evolutionary process. They also provide the vital information about ancestor-descendant relationships, necessary for comparative studies (Garamszegi & Gonzalez-Voyer, 2014). It is more informative when the phylogenetic framework encompasses the majority of the diversity within a group. One particular lineage may not be a good representative for the group, as genetic changes have been accruing independently in extant taxa for a very long time. However, the more lineages that are included, the more information that is available to accurately reconstruct these ancestor-descendant relationships, and create hypotheses about the ancestral state of internodes and the group as a whole. Also, the conclusions drawn from comparative evolutionary studies is dependent on the phylogeny used, and thus, it makes sense to use the most comprehensive

phylogeny in order to produce more accurate results. Nevertheless, the major problem remains that holocephalans have been poorly studied, and there are no comprehensive phylogenetic estimates on the inter-relationships for this group.

The nuclear multi-locus data produced mainly congruent results across the different data sets and analyses. All resulting tree topologies recovered the same set of clades and species within those clades. Partitioning of the data did not have a major effect on the results, and ML and BI methods produce similar tree topologies, as expected. The major differences among topologies were at shallower nodes and species-level relationships. The data sets that contained more exons, produced more highly supported and resolved topologies. Nucleotide characters produced better resolution compared to amino acid characters.

The mitochondrial data resulted in largely congruent tree topologies across analyses. Like the nuclear data, shallower nodes and species-level relationships were the major differences among topologies. Partitioning had no major effect. Character data (i.e., nucleotide, amino acid) effected resulting topologies, potentially due to model mis-specification and/or less information in the amino acid character set. The method of analysis, ML or BI, did not have an effect on results, producing identical or very similar topologies.

Both nuclear and mitochondrial data sets produced similar overall tree topologies and relationships among lineages. They both contained the same set of major clades and species within those clades. Major differences were evident in the species-level relationships, and placement of a few of the clades within Chimaeridae. The family level relationships were identical to previous work (Didier, 1995; Ward et al., 2008; Inoue et al., 2010; de la Cruz-Aguero et al., 2012; Licht et al., 2012). It was recovered that the two genera within Chimaeridae, *Chimaera* and *Hydrolagus*, are not monophyletic. Previous molecular work has speculated that these two genera are paraphyletic, and these results support their findings (Ward et al. 2008; de la

Cruz-Aguero et al. 2012; Licht et al. 2012). The anal fin character seems to be a plastic trait within the group, and should not be used as a means to separate species into separate genera. It is stated here that all members of Chimaeridae be placed into one genus, *Chimaera*, based on these findings. Additionally, two genera within Rhinochimaeridae, *Harriotta* and *Neoharriotta*, were also found to not be monophyletic in all analyses. In fact, it appears that at least two new genera may need to be described within this family. Also, the morphological characters that were used to place species in their respective genera within Rhinochimaeridae needs to be re-evaluated based on the molecular findings. An extensive morphological and molecular examination of these lineages is warranted to fully resolve the species boundaries. Other taxonomic suggestions based on the phylogenetic results indicate likely new species, as well as cases where described species lineages likely need to be synonymized into a single species (e.g., *Callorhinchus* species). These findings have major implications for understanding the diversity of this group of fishes, as well as conservation and management of species.

Nuclear and mitochondrial data also produced largely congruent divergence time estimations for the major splits, and similar to previous estimated credible ranges. However, differences were evident within the families, where there was less resolution at shallower nodes and species-level relationships, which was shown in the phylogenetic analyses. The divergence time trees can be useful for following patterns and interpreting the timing of trait evolution in future work. Overall, the fact that two independent data sets produce similar results for the evolutionary relationships among lineages and divergence time dating support that these are likely true relationships. From a historical biogeographical perspective using the phylogeny of holocephalans, it seems likely that the ancestor of extant lineages was present in the southern hemisphere – Southern Ocean region, with subsequent speciation events resulting in outward and upward migration to the Atlantic, Indian, and Pacific Oceans.

Results from both nuclear and mitochondrial loci for phylogenetic analysis have produced a comprehensive estimate of the evolutionary relationships among diverse holocephalans lineages. Future work should aim to find molecular markers to better resolve more recent divergences among species, such as markers that show more variation. Also, incongruences and low resolution of some relationships (i.e., shallow nodes) will need to be investigated in future studies as there is likely ILS (incomplete lineage sorting) between nuclear and mitochondrial data as well as between loci. More importantly, the tree topology estimates can be used in future studies to test evolutionary hypotheses both within holocephalans and among chondrichthyans. Additionally, the Bayesian inference phylogenies could be useful for future studies, as the data gained here could be used as prior information in phylogenetic reconstruction. Also, the phylogeny can be used to reconstruct the ancestral state of a particular trait of interest within holocephalans. The estimated ancestral state for a trait can be used comparatively to explore the evolution of the trait across jawed vertebrates, using a phylogeny, which would then include other jawed vertebrate taxa of interest. Information can be extracted about how the trait has transitioned across taxa, and identify what changes have occurred, with the hopes of gaining a better understanding of how the trait functions.

First, the phylogenies estimated can be used to better define species boundaries and diversity within this group of chimaeroid fishes. Given the taxonomic uncertainty of some of these lineages, this should be a major next step from this research. We must first know what species are present before we can study them. The results here provide a us with the needed information to start exploring these lineages and their relationships with one another in more detail in order to more accurately define what are species versus populations. A few applications of phylogenies beyond identifying species relationships are to explore trait evolution to better understand adaptation, innovation, and evolutionary constraint; the process of diversification to

identify patterns in biodiversity and how species are assembled and retained over time; and gene evolution to explore their origins, functions, constraints, and modifications over time and across species. Thus, second, the estimated phylogeny can be used in future studies to map traits or genes of interest, and explore character evolution. One approach could be to explore characters such as the anal fin, which was found here to be quite labile within the group, by mapping presence and absence on the tree in order to identify patterns, and explore its evolutionary history and constraint. Another approach would be to explore specialized traits across the families (e.g., snout morphology, anal fin, tail morphology, tooth morphology) to help study how these traits have evolved. It would be interesting to identify the genes responsible for the development of these unique features, and how they may function in other vertebrate species. Additionally, other traits like body size and habitat (e.g., depth, distribution) can also be investigated along the phylogeny. The tree can also be used to study the rate of diversification (speciation and extinction) within the group. More broadly, these data can be used along with data from elasmobranchs, to investigate patterns of evolution across chondrichthyans. This would provide a better understanding of overall chondrichthyan evolution, and a means to better approximate the ancestral condition of chondrichthyans. Again, character trait evolution can be explored, like reproductive mode, habitat, body size across the phylogeny. Diversification within this group as a whole would be interesting to investigate, and how different factors may be inter-related in explaining this process. The phylogeny could be taken a step further, even more broadly, and incorporated into comparative studies with bony vertebrates. As mentioned, these fish have the potential to be informative in these types of studies exploring vertebrate evolution. For example, one major area of research is in developmental biology, and understanding the origin and modifications of this process and genes involved across organisms to identify how such vastly different and complex species arise. Thus, the chimaeroid phylogeny can be used to explore

developmental patterns and genes, and use this information to reconstruct the ancestral condition for this process in chimaeroids. That information can then be used comparatively with other bony vertebrate groups to explore particular pathways and genes, and identify origins, modifications, and evolutionary constraints with respect to development.

The second goal of this dissertation was to identify the sex-determining mechanism in chondrichthyan fishes by examining chromosomes and screening a subset of the genome of a holocephalan species, *Callorhinchus callorynchus*, for sex-linked molecular markers that would validate the presence of a genetic mechanism of sex determination. The expectation is that this information can then be used to identify the sex-determining mechanism and ultimately the sex-determining gene(s) across all holocephalans lineages. This developmental process data can then be mapped onto the estimated phylogenetic tree for holocephalans to explore the evolution of sex determination within the group, and to estimate the ancestral state of the mechanism of sex determination in jawed vertebrates. However, while it has been speculated based on karyotypes (Donahue, 1974; Schwartz & Maddock, 1986, 2002; Kikuno & Ojima, 1987; Asahida et al., 1993; Asahida & Ida, 1995; Maddock & Schwartz, 1996; de Souza Valentim et al., 2006; da Cruz et al., 2011; Aichino et al., 2013; de Souza Valentim et al., 2013) and parthenogenetic data (Chapman et al., 2007; Chapman et al., 2008; Feldheim et al., 2010; Robinson et al., 2011; Portnoy et al., 2014; Fields et al., 2015; Harmon et al., 2016; Straube et al., 2016; Dudgeon et al., 2017), that sex is determined by a genetic mechanism in chondrichthyan fishes, it has yet to be explored and validated using genetic data.

It was first hypothesized that males would be represented by at least one pair of heteromorphic chromosomes, and females would be the homomorphic sex, indicating an XX/XY sex chromosome system and confirm GSD. However, experiments aimed at preparing chromosome spreads via direct methods and cell culture methods for karyotyping and analysis

were unsuccessful. Consequently, it could not be determined through visualization of chromosomes if one sex contained heteromorphic chromosomes.

Next, it was hypothesized that there would be at least one sex-linked molecular marker present in the male sex of the study species, indicating an XX/XY sex chromosome system and GSD. Results from RAD-seq of three chondrichthyan species, a holocephalan, *C. callorhynchus*, a shark, *C. carcharias*, and a ray, *M. randalli*, are consistent with the presence of a genetic mechanism of sex determination. This is the first study to explore the genome of chondrichthyans to identify sex-linked markers. Two male-specific markers were identified in *C. callorhynchus*, and both were validated using PCR to be present in males and absent in females. This provided evidence that this species has male heterogamety, an XX/XY sex chromosome system. While the shark and ray species did not undergo PCR validation, the results for both also indicate an XX/XY sex chromosome system based on the identification of candidate male-specific RAD markers. These results greatly contribute to both the field of chondrichthyan research, and also developmental biology and sex determination in general, as it provides the first genetic evidence of sex-linked markers, which is consistent with the pattern of GSD and XX/XY sex chromosomes in species of chondrichthyan fishes. Additionally, the two male-specific RAD markers in *C. callorhynchus* were mapped to two separate scaffolds in the male genome of the closely related *C. milii*, identifying both as linked to the sex chromosomes. Interestingly, the four contigs of the male-specific markers identified in *C. carcharias* showed no similarities to the male-specific contigs for *C. callorhynchus*, nor could they be successfully mapped to the *C. milii* genome. This points to these two species having different sex chromosomes and sex-determining genes. Thus, these two species likely co-opted different autosomes for sex chromosomes. While we have seen that mammals are conserved in their use of the sex-determining gene, *SRY*, we have also seen an array of sex-determining systems and genes in teleost fishes (see Chapter 4).

Therefore, it would not be rare to find species that have been evolving independent for at least the last 400 million years to have evolved different sex-determining genes and chromosomes. With additional evidence from other chondrichthyan species, we will be able to gain a better understanding of sex-determining mechanisms, sex chromosome origins, and the forces driving their evolution.

This dissertation research on sex determination in chondrichthyan fishes provides an initial molecular framework for further exploration of sex-determining mechanisms and sex chromosomes in these fishes. A primary future goal is to validate the candidate sex-specific RAD markers identified in *C. carcharias* and *M. randalli* using the same PCR based approach as in *C. callorhynchus*. It is expected that these two species will also be confirmed as having male-specific genetic markers, and an XX/XY sex chromosome system, indicating that this system is widespread across highly divergent taxa. A second goal is to use the *C. milii* and *C. carcharias* genomes to determine the sex-determining locus, and explore similarities and differences between species. It has been shown here that the two male-specific markers of *C. callorhynchus* mapped to two scaffolds in the *C. milii* genome. Since a sex-linked marker should be present on the sex chromosome, future work can utilize the sequence data to ‘walk’ along the scaffolds and determine the full extent of the sex-determining region of the Y chromosome. Once this region has been defined for both species, potential genes and their function can be predicted. The next step to identifying and characterizing the sex-determining gene, is to use RT-PCR to quantify gene expression of potential genes throughout development and into adulthood in males and females in the appropriate tissues. It is expected that the master sex-determining gene would be expressed exclusively in males at a point right before sexual differentiation, and may be present throughout differentiation. Genes that are not the sex-determining gene are expected to show expression at other periods of time during development or after, but not during the critical period

before sexual differentiation. Subsequently, when a gene has been identified as the candidate sex-determining gene based on expression profiles, it needs to be characterized as both necessary and sufficient for male development. This is problematic in chondrichthyan fishes, as keeping many species in captivity is difficult to impossible, as would be *in vivo* manipulation of embryos. However, hypothetically, two sets of experiments would need to be carried out *in vivo*. First, the gene would need to be overexpressed in genetic females, and secondly, knocked out or inactivated in genetic males using targeted genome modification (e.g., CRISPER-Cas9). It is expected that overexpression in genetic females would induce testis differentiation, while inactivation in genetic males would produce XX females. This would provide the necessary evidence that this is the sex-determining gene.

Another future direction based on this research would be to explore the presence of the identified sex-specific RAD markers across the chondrichthyan tree of life. The validated *C. callorhynchus* primers can be used in a PCR assay to test for presence/absence in the remaining holocephalans lineages in males and females. Additionally, these primers can be used on other elasmobranch species. Once the *C. carcharias* and *M. randalli* RAD markers have been validated, their primers can then be used to explore conservation across the tree as well. It is expected that closely related species, such as all holocephalans, may share the same sex-determining region, and thus, show presence of the sex-linked markers in males and absence in females. This would indicate conservation of the sex-determining region and the same sex chromosomes within the Chimaeriformes, therefore, the common ancestor of this group likely had the same mechanism and sex chromosomes. If the sex-specific RAD markers are not present across holocephalans, it would indicate that this process is quite labile within the group. It has been shown that sex determination can be an incredibly diverse process both broadly across vertebrates, as well as within a group, with numerous transitions and closely related species

having different mechanisms (Marín & Baker, 1998; Janzen & Phillips, 2006; Marshall Graves, 2012).

The phylogenetic trees estimated in this dissertation can eventually be used here to map the sex-determining mechanisms, sex chromosomes, and sex-determining genes and networks across holocephalans, to explore sex chromosome origins, potential transitions in mechanisms, genes, chromosomes, and estimate the ancestral state. If different sex-determining loci are found among holocephalans lineages, this can be used to identify potential evolutionary forces that contribute to transitions among sex-determining mechanisms. I would not expect there to be major differences in the sex-determining mechanisms and sex chromosomes among holocephalans given that many of the lineages were found to show little genetic variation, particularly within families. The phylogenetic reconstruction can be taken more broadly to include other chondrichthyan species, as well as other jawed vertebrates. An enhanced representation of sex determination diversity across vertebrates requires more than your typical model organisms, and with the addition of information from non-model organisms like holocephalans and elasmobranchs, an improved understanding of the evolution of vertebrate sex determination will follow. Thus, comparative analyses will be fundamental to gain a better perspective on the history of vertebrate genomes and evolution with regards to this central developmental process.

A diverse phylogenetic framework with respect to vertebrate sex determination can be used to explore unanswered questions in the field. It is well known that sex determination mechanisms can evolve fairly rapidly (Marín & Baker, 1998). What factors (i.e., genetic, environmental) and forces play a role in driving turnover and evolution of sex-determining mechanisms and sex chromosomes? How are novel genes assimilated into the sex-determining pathway? Many different genes have been co-opted into the role of master sex-determining gene.

For example, some have evolved from genes known to be involved in the sex-determining pathway (i.e., *DMRT1*, *SOX3*), while others are not known to play a role in this process (i.e., *sdY*). So, how have unrelated genes been recruited for this function, and how are they chosen among different lineages? Therefore, it is necessary to understand the sex-determining genes and sex chromosomes among different lineages, in order to explore how and why different genes are initiated as the master regulator of sex determination. Last, lability of sex determination in some groups (i.e., teleost fishes), and conservation in others (i.e., mammals) begs the question as to why some groups appear to have evolutionary constraint on this process, while others show more plasticity. Overall, this research provides new and essential information where there was previously a large gap in knowledge. First, we now have a better understanding of the evolutionary history among holocephalans fishes, with an updated phylogenetic framework for their relationships, which also brought about important findings with regards to taxonomic diversity. Last, we have identified and validated two male-specific molecular markers in a holocephalan species, which is consistent with an XX/XY sex-determining genetic mechanism. Sex-linked markers were also identified for a shark and ray species, not yet validated, that is also consistent with a sex chromosome genetic mechanism. The coupling of this new knowledge has the ability to spark future research aimed at better understanding the evolution of sex determination.

GLOSSARY

Among-site rate variation—variation in the substitution rate of nucleotide bases among different sites in a DNA sequence

Base frequencies—the frequency of each nucleotide base at each site in a DNA sequence

Basal—found near the base, bottom, root of phylogenetic tree

Branch length—estimated number of substitutions per site

Consensus tree—a summary of the agreement between two trees

Derived—characters or traits present in modern lineage, not present in ancestors

Gamma parameter (Γ)—gamma distribution used to model substitution rate heterogeneity over alignment sites; the form of the distribution is controlled by alpha, the shape parameter.

General-time reversible (GTR) substitution model—a time reversible model of molecular evolution; parameters include 4 equilibrium base frequencies and a rate matrix; 6 substitution transition rate parameters (e.g., $T \rightarrow C$, $T \rightarrow A$, $T \rightarrow G$, etc); it is time reversible because the substitution rate matrix is symmetric (e.g., $T \rightarrow C = C \rightarrow T$); nucleotide bases can occur at different frequencies

Gnathostome—vertebrate organism with true jaws

Heterotachy—variation in lineages-specific evolutionary rates over time; substitution rates can change through time in a gene

HIVB model—uses the empirical HIV between-patient substitution model matrix (Nickle et al., 2007), derived from multiple HIV-1 alignments and viral genes to model between-host amino acid evolution

Homoplasy—a character state/site shared by two or more lineages that is not present in their common ancestor, and is the result of multiple nucleotide substitutions at that site

Incomplete lineage sorting—the presence of polymorphisms in a gene of an ancestor, followed by differential retention of those alleles in the descendent lineages of the ancestor; failure of alleles to coalesce within a lineage

JTT model—uses the JTT matrix (Jones et al., 1992) to estimate amino acid substitutions along a phylogeny; corrects for multiple substitutions

LG model—uses the LG matrix (Le & Gascuel, 2008) to estimate amino acid substitutions along a phylogeny; incorporates variation of evolutionary rates across sites

Lineage—an organism connected through a line of common descent from ancestor

Molecular clock—in a phylogenetic tree, the constancy of the evolutionary rate among lineages; mutations in DNA sequence occur at a relatively constant rate, the genetic distance between two DNA sequences for the same gene increases linearly with time

Monophyletic/Clade—a group of organisms that have descended from a common ancestor

MTMAM model—uses the empirical MTMAM model matrix (Yang et al., 1998), which was derived from proteins of the mitochondrial genome of mammals, to estimate amino acid substitutions along a phylogeny

Node/internode/internal node—on a phylogenetic tree, represents common ancestor of descendants

Paraphyletic—a group of organisms that have descended from a common ancestor, but does not include all the descendant/extant lineages

Polytomy—internal node of a phylogenetic tree that has more than two descendants

Primitive—close to or approximating an early ancestor; modern lineage having characteristics or traits that are the same or similar to the ancestral type; little evolved from ancestor

Proportion of invariant sites (I)—used to model rate heterogeneity over sites; fraction of sites assumed to never vary

Substitution model—a set of assumptions that describes the process of nucleotide substitutions during evolution

Terminal node—on a phylogenetic tree, represents the extant lineage/species/taxa

Transition/Transversion Ratio—ratio of the number of transitions to the number of transversions for a pair of DNA sequences

Ultrametric tree—tree where all the branch lengths from the root to the tips are equal

Voucher—a representative of a particular organism that has been collected, expertly identified, and in many cases preserved and deposited in a permanent collection (i.e., museum) for authentication and future research

APPENDIX 4.1 CHROMOSOME PREPARATION REAGENTS AND DETAILED PROTOCOL #1

Reagents

- Modified DMEM Culture Medium (DMEM media, 4.5 g/L NaCl, 20 g/L Urea, 30 g/L Hepes, 49.3 ml/L 7.5% sodium bicarbonate)
- 0.075 M KCl
- 1% colchicine
- Methanol-Acetic Acid (3:1, prepared fresh)
- 5% Giemsa

Sample Collection

1. Remove small pieces of tissue from animal with sterile scalpel and forceps.
2. Place tissue in sterile tube with modified culture media to a volume of 10 ml.
3. Place tube on ice or maintain at room temperature until back in the lab. Sample should be prepared as soon as possible.
4. Mince tissue with culture media, and return to tube.

Chromosome Preparation

5. Add ~ 3 µg/ml colchicine to tissue and culture medium (3 µl of 1% colchicine in 10 ml volume), gently shaken continuously, for 12-24 hours.
6. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
7. Add 10 ml 0.075 M KCl to tube, stir lightly, and place in incubator at 37 °C for 60-120 minutes.
8. Slowly add 5 drops fresh, ice-cold 3:1 methanol-acetic acid fixative to tube, stir gently.
9. After 5 minutes, add additional 5 ml of 3:1 methanol-acetic acid fixative, stir gently.
10. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
11. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently
12. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
13. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently
14. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
15. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently
16. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
17. Add ~ 1-2 ml of 3:1 methanol-acetic acid fixative and mix till homogeneous.

Dropping Slides

18. Turn water bath to 60 °C. Place a tray in water bath with a damp paper towel on bottom.
19. Clean cooled slides (4 °C) and label.
20. Gently re-suspend cells with sterile pipette.
21. Over the water bath, release 3 drops of cell suspension over slide, making sure to drop over different areas of slide.
22. Gently tilt slides to spread suspension over entire slide.
23. Let slides sit in closed water bath till dry.

24. Once dry, remove from water bath and place in a clean fume hood at room temperature for at least 1 day.

Chromosome Visualization

25. View slides under light microscope for presence of cells and metaphase spreads.

APPENDIX 4.2 CHROMOSOME PREPARATION REAGENTS AND DETAILED PROTOCOL #2

Reagents

- Modified DMEM Culture Medium (DMEM media, 4.5 g/L NaCl, 20 g/L Urea, 30 g/L Hepes, 49.3 ml/L 7.5% sodium bicarbonate)
- 0.075 M KCl
- 0.05% colchicine
- Methanol-Acetic Acid (3:1, prepared fresh)
- 5% Giemsa
- 0.2 N HCl
- Ba(OH)₂
- 2X SSC

Sample Collection

1. Remove small pieces of tissue from animal with sterile scalpel and forceps.
2. Place tissue in sterile tube with modified culture media to a volume of 10 ml.
3. Place tube on ice or maintain at room temperature until back in the lab. Sample should be prepared as soon as possible.
4. Mince and grind tissue with culture media, removing any remaining large tissue fragments, and return to tube.

Chromosome Preparation

5. Add ~ 2 µg/ml colchicine to tissue and culture medium (40 µl of 0.05% colchicine in 10 ml volume), gently shaken continuously, for 12-24 hours.
6. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
7. Add 10 ml of warmed (37 °C) 0.075 M KCl to tube, stir lightly, and place in incubator at 37 °C for 60-120 minutes.
8. Slowly add 5 drops fresh, ice-cold 3:1 methanol-acetic acid fixative to tube, stir gently.
9. After 5 minutes, add additional 7 ml of 3:1 methanol-acetic acid fixative, stir gently.
10. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
11. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently
12. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
13. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently
14. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
15. Add ~ 1-2 ml of 3:1 methanol-acetic acid fixative and mix till homogeneous, want a slightly cloudy cell suspension.

Dropping Slides

16. Turn water bath to 60 °C. Place a tray in water bath with a damp paper towel on bottom.
17. Clean cooled slides (4 °C) and label.
18. Gently re-suspend cells with sterile pipette.
19. Over the water bath, release 3 drops of cells-fixative over slide, making sure to drop over different areas of slide.

20. Gently tilt slides to spread suspension over entire slide.
21. Let slides sit in closed water bath for a short period of time.
22. Once dry, remove from water bath and place in a clean fume hood at room temperature for at least 1 day.

Staining – Standard Giemsa

23. Rinse slides carefully with distilled water.
24. Place slides in 5% Giemsa for 10 minutes.
25. Rinse slides 2 times with distilled water and gently blot dry.

Staining – C-banding

26. Place slides in 0.2 N HCl for 30 minutes.
27. Immerse slides in a saturated solution of Ba(OH)₂ for 5 minutes at 40 °C.
28. Place slides in 2X SSC for 10 minutes at 60 °C.
29. Rinse slides with distilled water.
30. Place slides in 5% Giemsa for 10 minutes.
31. Rinse slides twice with distilled water and gently blot dry.

Chromosome Visualization

32. View slides under light microscope for presence of cells and metaphase spreads.

APPENDIX 4.3 CELL CULTURE AND CHROMOSOME PREPARATION PROTOCOL #3

Cell Culture Reagents

- Elasmobranch-modified phosphate buffered saline (phosphate buffered saline, 2.63 g NaCl, 0.12 g NaH₂PO₄)
- Elasmobranch-modified RPMI 1640 (RMPI Medium 1640 (1X) + L-glutamine, 21.62 g Urea, 11 g NaCl, 2 ml Penicillin (5000 U/ml) & Streptomycin (5000 µg/ml), 50 ml fetal bovine serum, ~ 20 g 7.5% sodium bicarbonate)

Sample Collection

1. Collect whole blood via caudal venipuncture on anesthetized animal using a sterile vacutainer tube containing lithium heparin.
2. Gently mix by inverting the tube several times.
3. Place tubes on ice till back in the lab.

Peripheral Blood Leukocyte Culture & Harvest

4. Centrifuge whole blood vacutainer tube at 50 x g for 15 minutes.
5. Sterilize hood and all equipment to be used for cell culture.
6. Remove the top PBL cell layer, leaving red blood cells behind, and place in sterile 15 ml tube.
7. Add 10 ml of elasmobranch-modified RPMI.
8. Transfer to T25 flask.
9. Check under light microscope for presence of cells.
10. Place into incubator at ~ 25 °C and 5% CO₂ for 3 days.
11. Remove ~ 5 ml of tissue culture for chromosome preparation.

Chromosome Prep Reagents

- 0.075 M KCl
- 0.05% colchicine
- Methanol-Acetic Acid (3:1, prepared fresh)

Chromosome Preparation

12. Add ~ 2 µg/ml colchicine to tissue and culture medium (20 µl of 0.05% colchicine in 5 ml volume), gently shaken continuously, for 2 hours.
13. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
14. Add 10 ml of warmed (37 °C) 0.075 M KCl to tube, stir lightly, and place in incubator at 37 °C for 60-120 minutes.
15. Slowly add 5 drops fresh, ice-cold 3:1 methanol-acetic acid fixative to tube, stir gently.
16. After 5 minutes, add additional 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
17. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
18. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
19. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
20. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
21. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.

22. Add ~ 1 ml of 3:1 methanol-acetic acid fixative and mix till homogeneous, want a slightly cloudy cell suspension.

Dropping Slides

23. Turn water bath to 60 °C. Place a tray in water bath with a damp paper towel on bottom.
24. Clean cooled slides (4 °C) and label.
25. Gently re-suspend cells with sterile pipette.
26. Over the water bath, release 3 drops of cells-fixative over slide, making sure to drop over different areas of slide.
27. Gently tilt slides to spread suspension over entire slide.
28. Let slides sit in closed water bath for a short period of time.
29. Once dry, remove from water bath and place in a clean fume hood at room temperature for at least 1 day.

Chromosome Visualization

30. View slides under light microscope for presence of cells and metaphase spreads.

APPENDIX 4.4 CELL CULTURE AND CHROMOSOME PREPARATION PROTOCOL #4

Cell Culture Reagents

- Elasmobranch-modified phosphate buffered saline (phosphate buffered saline, 2.63 g NaCl, 0.12 g NaH₂PO₄)
- Elasmobranch-modified RPMI 1640 (RMPI Medium 1640 (1X) + L-glutamine, 21.62 g Urea, 11 g NaCl, 2 ml Penicillin (5000 U/ml) & Streptomycin (5000 µg/ml), 50 ml fetal bovine serum, ~ 20 g 7.5% sodium bicarbonate)

Sample Collection

1. Collect whole blood via caudal venipuncture on anesthetized animal using a sterile vacutainer tube containing either lithium or sodium heparin.
2. Gently mix by inverting the tube several times.
3. Place tubes on ice till back in the lab.

Whole Blood Culture & Harvest (Samples 20 & 21)

4. Transfer ~ 15 ml of whole blood to 15 ml tube.
5. Add 10 ml of elasmobranch-modified RPMI to whole blood, gently mix.
6. Transfer to T25 flask, place in incubator at ~ 25 °C and 5% CO₂ for 3-4 days.
7. Gently mix cell culture and remove ~ 5 ml of culture, put in 15 ml tube for chromosome prep.

Peripheral Blood Leukocyte Culture & Harvest (Samples 15-19, 22-24)

8. Transfer blood from vacutainer to 15 ml tube.
9. Add ~ 5 ml of elasmobranch-modified PBS.
10. Centrifuge tube at 50 x g for 20 minutes, check for separation, centrifuge additional 10 minutes as needed.
11. Sterilize hood and all equipment to be used for cell culture.
12. Remove the top PBL cell layer, leaving red blood cells behind, and place in sterile 15 ml tube.
13. Add 10 ml of elasmobranch-modified RPMI.
14. Transfer to T25 flask.
15. Place into incubator at approximately 25 °C and 5% CO₂ for 3-4 days.
16. Gently mix cell culture and remove ~ 5 ml of culture, put in 15 ml tube for chromosome prep.

Chromosome Prep Reagents

- 0.075 M KCl
- 0.05% colchicine
- Methanol-Acetic Acid (3:1, prepared fresh)
- 5% Giemsa

Chromosome Preparation

17. Add ~ 2 µg/ml colchicine to cell culture (20 µl of 0.05% colchicine in 5 ml volume), gently shaken continuously, for 2 hours.
18. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
19. Add 10 ml of warmed (37 °C) 0.075 M KCl to tube, stir lightly, and place in incubator at 37 °C for 60-120 minutes.
20. Slowly add 5 drops fresh, ice-cold 3:1 methanol-acetic acid fixative to tube, stir gently.

21. After 5 minutes, add additional 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
22. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
23. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
24. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
25. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
26. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
27. Add ~ 1 ml of 3:1 methanol-acetic acid fixative and mix till homogeneous, want a slightly cloudy cell suspension.

Dropping Slides

28. Turn water bath to 60 °C.
29. Clean cooled slides (4 °C) and label.
30. Gently re-suspend cells with sterile pipette.
31. Over the water bath, release 3 drops of cells-fixative over slide, making sure to drop over different areas of slide.
32. Gently tilt slides to spread suspension over entire slide.
33. Dry slides on hot plate for at least 30 minutes.
34. Once dry, remove from water bath and place in a clean fume hood at room temperature for at least 1 day.

Chromosome Visualization

35. View slides under light microscope for presence of cells and metaphase spreads.

Staining – Standard Giemsa

36. Rinse slides carefully with distilled water.
37. Place slides in 5% Giemsa for 10 minutes.
38. Rinse slides 2 times with distilled water and gently blot dry.

Cell Viability of Cell Cultures Using Hemocytometer

39. Take 50 µl mixed cell suspension from cell culture and place in 1.5 ml tube.
40. Add 50 µl 0.4% trypan blue and mix gently.
41. Using a hemocytometer, add ~ 10 µl cell suspension-trypan blue mixture to each side.
42. Assess cell viability.

APPENDIX 4.5 CELL CULTURE WITH CONCAVALIN A ADDITION AND CHROMOSOME PREPARATION PROTOCOL #5

Cell Culture Reagents

- Elasmobranch-modified RPMI 1640 (RPMI Medium 1640 (1X) + L-glutamine, 21.62 g Urea, 11 g NaCl, 2 ml Penicillin (5000 U/ml) & Streptomycin (5000 µg/ml), 50 ml fetal bovine serum, ~ 20 g 7.5% sodium bicarbonate)
- Concanavalin A

Sample Collection

1. Collect whole blood via caudal venipuncture on unanesthetized animal using a sterile vacutainer tube containing sodium heparin.
2. Gently mix by inverting the tube several times.
3. Place tubes on ice till back in the lab.

Whole Blood Culture & Harvest

4. Take 1 ml of whole blood and mix with 9 ml elasmobranch-modified RPMI, gently mix, place in T25 flask.
5. Set up 3 flasks: add the mitogen concanavalin-A to each flask - 0 µg/ml concanavalin-A, 75 µg/ml concanavalin-A, 250 µg/ml concanavalin-A
6. Place in incubator at ~ 25 °C and 5% CO₂ for 4 days.
7. Check cell cultures each day for presence of live cells by taking an aliquot and checking for viability as outlined below.

Chromosome Prep Reagents

- 0.075 M KCl
- 0.05% colchicine
- Methanol-Acetic Acid (3:1, prepared fresh)
- 5% Giemsa

Chromosome Preparation

8. Add ~ 2 µg/ml colchicine to cell culture flask (40 µl of 0.05% colchicine in 10 ml volume), gently mix, put back in incubator for 2 hours.
9. Transfer cell culture suspension to 15 ml tube.
10. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
11. Slowly add 7 ml of warmed (30 °C) 0.075 M KCl to tube, gently stir, and place in incubator at 30 °C for 60-120 minutes.
12. Slowly add 5 drops fresh, ice-cold 3:1 methanol-acetic acid fixative to tube, stir gently.
13. After 5 minutes, slowly add additional 5 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
14. Centrifuge at 1000 rpm for 10 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
15. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.
16. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
17. Add 7 ml of 3:1 methanol-acetic acid fixative, stir gently, let sit for 30 minutes.

18. Centrifuge at 1000 rpm for 7 minutes, remove and discard supernatant, re-suspend pellet by tapping of tube.
19. Add ~ 1 ml of 3:1 methanol-acetic acid fixative and mix till homogeneous, want a slightly cloudy cell suspension.

Dropping Slides

20. Turn water bath to 60 °C.
21. Clean cooled slides (4 °C) and label.
22. Gently re-suspend cells with sterile pipette.
23. Over the water bath, release 3 drops of cells-fixative over slide, making sure to drop over different areas of slide.
24. Gently tilt slides to spread suspension over entire slide.
25. Dry slides on hot plate for at least 30 minutes.
26. Once dry, remove from water bath and place in a clean fume hood at room temperature for at least 1 day.

Chromosome Visualization

27. View slides under light microscope for presence of cells and metaphase spreads.

Staining – Standard Giemsa

28. Rinse slides carefully with distilled water.
29. Place slides in 5% Giemsa for 10 minutes.
30. Rinse slides 2 times with distilled water and gently blot dry.

Cell Viability of Cell Cultures Using Hemocytometer

31. Take 50 µl mixed cell suspension from cell culture and place in 1.5 ml tube.
32. Add 50 µl 0.4% trypan blue and mix gently.
33. Using a hemocytometer, add ~ 10 µl cell suspension-trypan blue mixture to each side.
34. Assess cell viability.

APPENDIX 4.6 CELL CULTURE AND CHROMOSOME PREPARATION PROTOCOL AT MEDICAL UNIVERSITY OF SOUTH CAROLINA, DEPARTMENT OF CYTOGENETICS #6

Cell Culture Reagents

- RPMI 1640 Media (RPMI 1640, 20 ml fetal calf serum, 1 ml L-glutamine, 1 ml Antibiotic-Antimycotic)
- Phytohaemagglutinin (PHA)
Pokeweed
- IL-2

Sample Collection

1. Collect whole blood via caudal venipuncture on unanesthetized animal using a sterile vacutainer tube containing sodium heparin.
2. Gently mix by inverting the tube several times.
3. Place tubes on ice till back in the lab.

Cell Counts on Hemocytometer

4. Take 10 μ l whole blood and mix with 190 μ l 2% acetic acid diluent.
5. Place aliquot on hemocytometer and count number of cells
6. If cell number is too high, dilute blood, and recount cell number.
7. Continue to dilute blood until average cell count per square is < 160 .

Whole Blood Culture

8. Set up 6 cultures per sample, 3 with whole blood, and 3 with diluted whole blood that has a cell count near the normal human blood cell count.
9. For whole blood cultures A-C, add 0.5 ml whole blood to 9.5 ml RPMI media.
10. For diluted blood cultures D-F, add amount of diluted blood and RPMI media as stated in table to a volume of 10 ml.
11. Add 200 μ l of mitogen, either PHA (Cultures A-D), pokeweed (Culture E), or IL-2 (Culture F).
12. Incubate culture A at 37 °C and 5% CO₂ for 3 days
13. Incubate cultures B-F at room temperature on counter, ~ 24 °C, with a tight lid, for 3 days.

Chromosome Prep Reagents

- Ethidium bromide
- Colcemid
- 0.075 M KCl
- Methanol-Acetic Acid (3:1, prepared fresh and chilled before use)

Chromosome Preparation

14. Add 100 μ l ethidium bromide to flask, place back on counter or in incubator for 65 minutes.
15. Add 100 μ l colcemid to flask, gently mix, transfer to 15 ml tube, place back on counter or in incubator for 20 minutes
16. Centrifuge at 1300 rpm for 8 minutes, remove supernatant, mix pellet well
17. Add 10 ml 0.075 M KCl (at temperature of cells), invert a few times to mix, leave on counter or place in incubator for 17 minutes.
18. Add 1 ml 3:1 methanol-acetic acid fixative, invert a few times to mix, let sit for 1 minute at room temperature.

19. Centrifuge at 1300 rpm for 8 minutes, remove supernatant, mix pellet very well.
20. Add 10 ml room temperature 3:1 methanol-acetic acid fixative.
21. Centrifuge at 1300 rpm for 8 minutes, remove supernatant, mix pellet very well.
22. Add 7 ml room temperature 3:1 methanol-acetic acid fixative.
23. Centrifuge at 1300 rpm for 8 minutes, remove supernatant, mix pellet very well.
24. Add 5 ml room temperature 3:1 methanol-acetic acid fixative.
25. Place in freezer for at least 1 hour before dropping slides.

Dropping Slides

26. Remove tubes from freezer, centrifuge at 1300 rpm for 8 minutes.
27. In thermotron (47% humidity, 80 °F)
 - a. Clean slide front and back with lens cleaner
 - b. Remove most of fixative from tube, do not touch pellet
 - c. Mix pellet well by tapping.
 - d. Slowly add room temperature 3:1 methanol-acetic acid fixative till cell suspension is almost clear.
 - e. Add 3-4 drops to slide, moving around to spread
 - f. Dab off excess liquid, let dry.

Chromosome Visualization

28. View slide under light microscope, for metaphase spreads.

APPENDIX 4.7 CONCANAVALIN A EXPERIMENT

Cell Culture Reagents

- Elasmobranch-modified RPMI 1640 (RPMI Medium 1640 (1X) + L-glutamine, 21.62 g Urea, 11 g NaCl, 2 ml Penicillin (5000 U/ml) & Streptomycin (5000 µg/ml), 50 ml fetal bovine serum, ~ 20 g 7.5% sodium bicarbonate)
- Concanavalin A
- Phosphate buffered saline (PBS)

Sample Collection

1. Collect whole blood via caudal venipuncture on unanesthetized animal using a sterile vacutainer tube containing sodium heparin.
2. Gently mix by inverting the tube several times.
3. Place tubes on ice till back in the lab.

Whole Blood Cell Counts on Hemocytometer

4. Take 10 µl whole blood and mix with 190 µl 2% acetic acid diluent.
5. Place aliquot on hemocytometer and count number of cells
6. If cell number is too high, dilute blood, and recount cell number.
7. Continue to dilute blood until average cell count per square is < 160.

Seeding 96-well Plate Whole Blood

8. Use blood concentration cell density determined above to calculate volume of blood to add to elasmobranch-modified RPMI for seeding at ~ 6250 cells per well.
9. Add 100 µl of seeding whole blood dilution to wells receiving whole blood – see tables below.

Peripheral Blood Leukocyte Preparation

10. Centrifuge remaining whole blood at 50 x g for 15 minutes.
11. Aspirate plasma, buffy coat layer, and top of red blood cell layer.
12. Centrifuge plasma, buffy coat layer, and red blood cell layer again at 50 x g for 15 minutes.
13. Aspirate plasma and buffy coat layer, leaving red blood cells behind.

Peripheral Blood Leukocyte Cell Counts on Hemocytometer

14. Take 10 µl PBL cell suspension and mix with 190 µl 2% acetic acid diluent.
15. Place aliquot on hemocytometer and count number of cells
16. If cell number is too high, dilute blood, and recount cell number.
17. Continue to dilute blood until average cell count per square is < 160.

Seeding 96-well Plate Peripheral Blood Leukocytes

18. Use PBL cell concentration cell density determined above to calculate volume of blood to add to elasmobranch-modified RPMI for seeding at ~ 6250 cells per well.
19. Add 100 µl of seeding PBL dilution to wells receiving PBL cells – see tables below.

Concanavalin A Concentrations

20. Make dilutions of concanavalin-A with PBS: 2000 µg/ml, 1250 µg/ml, 1000 µg/ml, 750 µg/ml, 500 µg/ml, 250 µg/ml
21. Add 20 µl of appropriate concanavalin A concentration to wells – see tables below.

Blanks & Controls

22. Blank: + media, + BrdU, - cells, - concanavalin A
23. Controls
 - a. + Whole blood, - concanavalin A, - BrdU
 - b. + PBL cells, - concanavalin A, - BrdU
 - c. + Media, - cells, + concanavalin A (at 6 different concentrations), + BrdU

Cell Culture

24. Incubate plate at ~ 25 °C, 5% CO₂

BrdU Cell Proliferation Assay

25. BrdU Incorporation: Add 10X BrdU solution to desired wells to a final concentration of 1X. Incubate plate at ~ 25 °C.
26. BrdU Detection & Measurement
 - a. Follow BrdU Kit Protocol (BioVision, catalog # K306-200)
 - b. Measure absorbance at 450 nm on a BioTek Synergy HT.

APPENDIX 4.8 CONCANAVALIN A EXPERIMENTAL DESIGN. *CARCHARHINUS ISODON* MALE, WHOLE BLOOD, SODIUM HEPARIN TUBE.

	1	2	3	4	5	6	7	8	9	10
A	Media+ConA 2000	Media+ConA 2000	Media+ConA 2000	Whole+ConA 2000	Whole+ConA 2000	Whole+ConA 2000	PBL+ConA 2000	PBL+ConA 2000	PBL+ConA 2000	
B		Media+ConA 1250	Media+ConA 1250	Media+ConA 1250	Whole+ConA 1250	Whole+ConA 1250	Whole+ConA 1250	PBL+ConA 1250	PBL+ConA 1250	PBL+ConA 1250
C	PBL+ConA 1000	PBL+ConA 1000	PBL+ConA 1000	Media+ConA 1000	Media+ConA 1000	Media+ConA 1000	Whole+ConA 1000	Whole+ConA 1000	Whole+ConA 1000	
D		PBL+ConA 750	PBL+ConA 750	PBL+ConA 750	Media+ConA 750	Media+ConA 750	Media+ConA 750	Whole+ConA 750	Whole+ConA 750	Whole+ConA 750
E	Whole+ConA 500	Whole+ConA 500	Whole+ConA 500	PBL+ConA 500	PBL+ConA 500	PBL+ConA 500	Media+ConA 500	Media+ConA 500	Media+ConA 500	
F		Whole+ConA 250	Whole+ConA 250	Whole+ConA 250	PBL+ConA 250	PBL+ConA 250	PBL+ConA 250	Media+ConA 250	Media+ConA 250	Media+ConA 250
G	Whole, No ConA,+BrdU	Whole, No ConA,+BrdU	Whole, No ConA,+BrdU	PBL, No ConA,+BrdU	PBL, No ConA,+BrdU	PBL, No ConA,+BrdU	Media Only	Media Only	Media Only	
H	Whole Blood Only	Whole Blood Only	Whole Blood Only					PBL Only	PBL Only	PBL Only

Media Only, No Cells, + Concanavalin A Concentration, +BrdU

Media Only, No Cells, No Concanavalin A, +BrdU

Blood Cells Only with media, No Concanavalin A, No BrdU

Blood Cells with media, Concanavalin A Concentration, +BrdU

Whole = whole blood cell culture

PBL = peripheral blood leukocytes culture

APPENDIX 4.9 CONCANAVALIN A EXPERIMENTAL DESIGN. RHIZOPRIONODON TERRAENOVAE MALE, WHOLE BLOOD, SODIUM HEPARIN TUBE.

	1	2	3	4	5	6	7	8	9
A	Media+ConA 2000	Media+ConA 2000	Media+ConA 2000	Whole+ConA 2000	Whole+ConA 2000	Whole+ConA 2000	PBL+ConA 2000	PBL+ConA 2000	PBL+ConA 2000
B	Whole+ConA 1250	Whole+ConA 1250	Whole+ConA 1250	Media+ConA 1250	Media+ConA 1250	Media+ConA 1250	PBL+ConA 1250	PBL+ConA 1250	PBL+ConA 1250
C	Whole+ConA 1000	Whole+ConA 1000	Whole+ConA 1000	PBL+ConA 1000	PBL+ConA 1000	PBL+ConA 1000	Media+ConA 1000	Media+ConA 1000	Media+ConA 1000
D	Media+ConA 750	Media+ConA 750	Media+ConA 750	Whole+ConA 750	Whole+ConA 750	Whole+ConA 750	PBL+ConA 750	PBL+ConA 750	PBL+ConA 750
E	Whole+ConA 500	Whole+ConA 500	Whole+ConA 500	Media+ConA 500	Media+ConA 500	Media+ConA 500	PBL+ConA 500	PBL+ConA 500	PBL+ConA 500
F	Whole+ConA 250	Whole+ConA 250	Whole+ConA 250	PBL+ConA 250	PBL+ConA 250	PBL+ConA 250	Media+ConA 250	Media+ConA 250	Media+ConA 250
G	Media Only	Media Only	Media Only	Whole, No ConA, +BrdU	Whole, No ConA, +BrdU	Whole, No ConA, +BrdU	PBL, No ConA, +BrdU	PBL, No ConA, +BrdU	PBL, No ConA, +BrdU
H	Whole Blood Only	Whole Blood Only	Whole Blood Only	PBL Only	PBL Only	PBL Only			

Media Only, No Cells, + Concanavalin A Concentration, +BrdU

Media Only, No Cells, No Concanavalin A, +BrdU

Blood Cells Only with media, No Concanavalin A, No BrdU

Blood Cells with media, Concanavalin A Concentration, +BrdU

Whole = whole blood cell culture

PBL = peripheral blood leukocytes culture

APPENDIX 4.10 PROTOCOL FOR SINGLE DIGEST RAD-SEQ LIBRARY PREPARATION

Library preparation protocol follows Etter et al. (2011) and modified by Gamble et al. (2015).

Reagents

- EZNA[®] Tissue DNA Kit (Omega Bio-Tek)
- RNase A: 10 mg/ml (Qiagen)
- High-quality genomic DNA: 25 ng/ μ l
- Qubit dsDNA High Sensitivity Assay Kit
- SbfI-HF Restriction enzyme (NEB)
- NEB Buffer 2
- rATP: 100 mM
- P1 Adapter: 100 nM stocks in 1x annealing buffer (AB). Single-stranded oligonucleotides were diluted to 100 μ M in a 10 μ l volume. Combine complementary adapter oligos: 2 μ l of top oligo and 2 μ l of bottom oligos combined in 16 μ l of 1x annealing buffer (10x: 500 mM NaCl, 100 mM Tris-Cl, pH 7.5-8.0). Use a boil and gradual cool program in a PCR machine to hybridize to anneal adapters together. Dilute to 100 nM concentrations in 1x annealing buffer: 1 μ l of adapter mix to 99 μ l of 1x annealing buffer. See Gamble et al. (2015) for adapter sequences.
- T4 DNA Ligase (NEB): 2,000,000 U/ml
- Covaris M220 microtubes
- QIAquick Gel Extraction Kit
- NEBNext Ultra End Repair/dA-tailing Module (NEB)
- MinElute Purification Kit
- MagNA Beads (5% Sera-Mag Magnetic Speed-beads, 18% PEG8000, 2.5 M NaCl, 10 nm Tris-Cl, 1 mM EDTA)
- MagNA Buffer (18% PEG8000, 2.5 M NaCl, 10 mM Tris-Cl, 1 mM EDTA)
- P2 Adapter: 10 μ M stock in 1 x AB prepared as P1 adapter described above. See Gamble et al. (2015) for adapter sequences.
- Kapa HiFi HotStart Ready Mix PCR
- RAD amplification primer mix: 10 μ M. Prepare 100 μ M stocks for each oligonucleotide in 1x EB. Mix together at 10 μ M.

Equipment

- Covaris M220 focused-ultrasonicator (Covaris, Inc., Woburn, Massachusetts, USA)
- Qubit[®] 2.0 Fluorometer (Life Technologies, Grand Is, New York, USA)
- Eppendorf Mastercycler[®] pro S (Hauppauge, New York, USA)

Procedure

1. Male and Female Library Preparation

I. DNA Extraction, RNase A Treatment

Clean, intact, high-quality DNA is ideal for success of protocol. Lower quality DNA can be used, but starting amount will likely need to be increased.

1. Use EZNA[®] Tissue DNA Kit to extract genomic DNA following manufacturer's instructions. Be sure to treat samples with RNase A following manufacturer's instructions during extraction to remove residual RNA.
2. Quantify DNA concentration using Qubit[®] 2.0 Fluorometer and dsDNA HS Assay Kit. Optimal concentration is 25 ng/μl or greater.
3. Run DNA on gel to check DNA integrity.

II. Restriction Endonuclease Digestion

Choice of restriction enzyme is based on several parameters such as the desired frequency of RAD sites throughout the genome, GC content, the depth of coverage necessary, and size of the genome.

1. Digest 1 μg of genomic DNA for each sample with SbfI-HF in a 50 μl reaction volume, following manufacturer's instructions. Add enzyme last, flick tubes to mix, and spin down.

<i>Reagent</i>	<i>Volume per sample</i>
10x NEB CutSmart Buffer	5 μl
Genomic DNA	1 μg
SbfI-HF	1 μl
Nuclease-free water	Up to 50 μl

2. Incubate at 37°C for one hour in a thermocycler.
3. Heat inactivate restriction enzyme at 80°C for 20 minutes, followed by 23°C for 10 min (to allow reaction to cool to ambient temperature) in thermocycler.

III. P1 Adapter Ligation

Ligate barcoded, restriction site overhang-specific P1-adapters onto complementary compatible ends on digested DNA.

1. To each digest (50 μl) add following reagents for a 60 μl total volume. Be sure to add adapters to reaction before ligase. Add ligase last. Flick to mix and spin down tube. Make a mix of buffer, water, and rATP. Add adapter and ligase separately.

<i>Reagent</i>	<i>Volume (μL) per sample</i>
NEB Buffer 2 (10X)	1
Nuclease-free water	5.4
Barcoded P1 adapter (100 nM)	2.5
rATP (100 mM)	0.6
T4 DNA ligase (2,000,000 U/ml)	0.5

2. Incubate reaction at room temperature (23°C) overnight in a thermocycler.
3. Heat inactivate T4 DNA ligase for 10 minutes at 65°C, followed by 23°C for 10 min in thermocycler.

IV. Sample Multiplexing and DNA Shearing

Shear DNA to an average size of 500 bp.

1. Quantify DNA concentration of each sample using the Qubit.
2. Combine barcoded samples at equimolar ratios, making a male library and a female library. If needed, add nuclease-free water for a total volume of 130 μ l, for 1-2 μ g total DNA.
3. Shear multiplexed sample to an average size of 500 bp using a Covaris M220 focused-ultrasonicator.
4. Run 3 μ l of sheared DNA on a 1% agarose gel to check size of DNA fragments.
5. Clean up sheared DNA using MinElute PCR Purification Kit following manufacturer's instructions, except elute DNA in two rounds of 15 μ l elution buffer.

V. Size Selection/Agarose Gel Extraction

This step removes free unligated or concatemerized P1 adapters and restricts the size range of tags to those that can be sequenced efficiently on Illumina sequencer. Select for fragments >200 bp

1. Run the entire sheared library sample (30 μ l) in 6x Loading Dye (10 μ l; total volume 40 μ l) on a 1.25% agarose, 0.5x TBE gel, 0.5x gel red (75 ml gel), for 45 min at 100 V, next to 5 μ l 100 bp DNA Ladder (with 10 μ l loading dye). Run ladder in lane between the samples until the 300 and 500 bp ladder bands are sufficiently resolved from 200 to 600 bp bands.
2. Use a fresh razor blade, cut a slice of the gel spanning 250-700 bp.
3. Extract DNA using the QIAquick Gel Extraction kit following manufacturer's instructions, except melt agarose gel slices in the supplied buffer at room temperature with agitation until dissolved.
4. Elute in 30 μ l EB and combine columns for each library.

VI. End Repair and 3' -dA Overhang Addition

Convert 5' and 3' overhangs, created by shearing, into phosphorylated blunt ends using T4 DNA polymerase and T4 polynucleotide kinase. Need to add an -A base to the 3' ends of DNA fragments. Prepares DNA for ligation to the P2 adapter, which possesses a single T base overhang at the 3' end of its bottom strand.

1. Using the NEB Ultra End Repair/dA-tailing module, mix the following reagents with DNA library for a 65 μ l volume reaction. Mix by pipetting, followed by a quick spin.

<i>Reagent</i>	<i>Volume (μL) per sample</i>
End Prep Enzyme Mix	3
End Repair Reaction Buffer	6.5
DNA library (5ng-1ug)	55.5

2. Incubate in thermocycler for 30 minutes at 20°C, then 30 minutes at 65°C, (hold at 4°C). Heated lid on.

- Purify with a MinElute PCR Purification Kit. Elute in 45 μ l of EB into a microcentrifuge tube containing 5 μ l 10x NEB Buffer 2.

VII. P2 Adapter Ligation

Ligate the P2 adapter, a Y adapter with divergent ends that contains a 3' dT overhang, onto the ends of the DNA fragments with 3' dA overhangs to create RAD seq library template ready for amplification. P2 adapters also have unique barcodes for male and female libraries.

- To each sample from previous step (50 μ l), add:

<i>Reagent</i>	<i>Volume (μL) per sample</i>
Barcoded P2 adapter (10 μ M)	1
rATP (100 mM)	0.5
T4 DNA ligase (2,000,000 U/ml)	0.5

- Incubate the reaction at room temperature (23°C) overnight in a thermocycler.
- Purify with a MinElute PCR Purification Kit. Elute in 50 μ l of EB.
- Quantify each library using the Qubit.

VIII. RAD Tag Amplification

High-fidelity PCR amplification is used to enrich for RAD tags that contain both adaptors.

- Perform a test amplification to determine library quality. Want library to be between 5-10 ng for a 25 μ l reaction. In a PCR tube, combine the following reagents into a 25 μ l reaction:

<i>Reagent</i>	<i>Volume (μL) per sample</i>
Nuclease-free water	9.5
KAPA HiFi HotStart Ready Mix	12.5
RAD amplification P1 primer (10 μ M)	1
RAD amplification P2 primer (10 μ M)	1
RAD library template	1

- Incubate in thermocycler with the following program: 45 sec 98°C, 18x (15 sec 98°C, 30 sec 65°C, 45 sec 72°C), 5 min 72°C, 10 min 4°C.
- Run 5 μ l PCR product in 6x loading dye out on a 1% agarose, 1X TBE gel, 0.5X GelRed dye, next to 1 μ l RAD library template and 2 μ l 100 bp DNA ladder.
- If amplified product is at least twice as bright as the template, perform a larger volume amplification (100 μ l) but with fewer cycles (12-14). If amplification looks poor, use more library template in a second test PCR reaction.
- In a PCR tube, combine the following reagents into a 100 μ l reaction. May need to alter RAD library template volume.

<i>Reagent</i>	<i>Volume (μL) per sample</i>
Nuclease-free water	38
KAPA HiFi HotStart Ready Mix	50
RAD amplification P1 primer (10 μ M)	4
RAD amplification P2 primer (10 μ M)	4
RAD library template	4

6. Incubate in thermocycler with the following program: 45 sec 98°C, 12-14x (15 sec 98°C, 45 sec 65°C, 45 sec 72°C), 5 min 72°C, 10 min 4°C.
7. Purify the PCR product using the MinElute PCR Purification Kit. Elute DNA in 20 μ l of EB.
8. Use MagNa beads to clean up PCR reaction.
 - a. Use a 0.75 beads to DNA volume ratio. For 100 μ l, add 75 μ l beads/buffer, mix, let sit for 10 minutes.
 - b. Place tube on magnetic plate, let sit for 5 minutes.
 - c. Remove supernatant.
 - d. Add 186 μ l freshly prepared 70% ethanol to beads, let sit 1 minute, remove supernatant.
 - e. Repeat for 2 ethanol washes.
 - f. Remove residual ethanol, let air dry for 5 minutes.
 - g. Elute DNA in 20 μ l of water, mix, let sit for a few minutes.
 - h. Place tube on magnetic plate, transfer supernatant to a new tube.
9. Purification/Agarose gel extraction to remove artefactual bands.
 - a. Load entire sample (20 μ l) in 6x Loading dye (10 μ l) on a 1.25% agarose, 0.5x TBE gel, with 0.5x GelRed, and run for 45 min at 100 V, next to 5 μ l of 100 bp DNA Ladder in 10 μ l loading dye.
 - b. Use a fresh razor to cut a slice of the gel spanning ~300-600 bp.
 - c. Extract DNA using QIAquick gel extraction kit following manufacturer's instructions, but melt agarose gel at room temperature
 - d. Elute in 30 μ l of EB. (If more than one column is used per sample, elute in 30 μ l of EB and combine.
10. If needed, use MagNA beads to concentrate library.
 - a. Use a 0.75 beads to DNA volume ratio.
 - b. Follow same protocol as above. Elute DNA in 20 μ l of water, transfer to new tube.
11. Quantify DNA using the Qubit. Concentrations should range from 1 to 20 ng/ μ l.

LIST OF REFERENCES

- Abozaid, H., Wessels, S., & Hörstgen-Schwark, G. (2011). Effect of rearing temperatures during embryonic development on the phenotypic sex in zebrafish (*Danio rerio*). *Sexual Development*, 5(5), 259-265.
- Aichino, D. R., Pastori, M. C., Roncati, H. A., Ledesma, M. A., Swarca, A. C., & Fenocchio, A. S. (2013). Characterization and description of a multiple sex chromosome system in *Potamotrygon motoro* (Chondrichthyes, Myliobatiformes) from the Parana River, Argentina. *Genetics and Molecular Research*, 12(3), 2368-2375.
- Alföldi, J., & Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Research*, 23(7), 1063-1068.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389-3402.
- Alves, A. L., Oliveira, C., Nirchio, M., Granado, Á., & Foresti, F. (2006). Karyotypic relationships among the tribes of Hypostominae (Siluriformes: Loricariidae) with description of XO sex chromosome system in a Neotropical fish species. *Genetica*, 128(1), 1-9.
- Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., . . . Postlethwait, J. H. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science*, 282(5394), 1711-1714.
- Amores, A., Suzuki, T., Yan, Y. L., Pomeroy, J., Singer, A., Amemiya, C., & Postlethwait, J. H. (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Research*, 14(1), 1-10.
- Anderson, J. L., Rodriguez Mari, A., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., & Postlethwait, J. H. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One*, 7(7), e40701.
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81-92.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-m., Dehal, P., . . . Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297(5585), 1301-1310.

- Arlt, D., Bensch, S., Hansson, B., Hasselquist, D., & Westerdahl, H. (2004). Observation of a ZZW female in a natural population: implications for avian sex determination. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(Suppl 4), S249.
- Arnason, U., Gullburg, A., & Janke, A. (2001). Molecular phylogenetics of gnathostomous (jawed) fishes: old bones, new cartilage. *Zoologica Scripta*, 30(4), 249-255.
- Aronesty, E. (2011). Command-line tools for processing biological sequencing data. ea-utils <http://code.google.com/p/ea-utils/>: Expression Analysis, Durham, NC.
- Asahida, T., & Ida, H. (1989). Karyological notes on four sharks in the order Carcharhiniformes. *Japanese Journal of Ichthyology*, 36(2), 275-280.
- Asahida, T., & Ida, H. (1990). Karyotypes of Two Rays, *Torpedo tokionis* and *Dasyatis matsubarae*, and Their Systematic Relationships. *Japanese Journal of Ichthyology*, 37(1), 71-75.
- Asahida, T., & Ida, H. (1995). Karyotype and cellular DNA content of a guitarfish, *Rhinobatos schlegelii*. *Kromosomo*, 79, 2725-2730.
- Asahida, T., Ida, H., & Inoue, S. (1987). Karyotypes of three rays in the order Myliobatiformes. *Japanese Journal of Ichthyology*, 33(4), 426-430.
- Asahida, T., Ida, H., & Inoue, T. (1988). Karyotypes and cellular DNA contents of two sharks in the family Scyliorhinidae. *Japanese Journal of Ichthyology*, 35(2), 215-219.
- Asahida, T., Ida, H., Terashima, H., & Chang, H.-Y. (1993). The karyotype and cellular DNA content of a ray, *Mobula japonica*. *Japanese Journal of Ichthyology*, 40(3), 317-322.
- Avise, J. C. (2004). *Molecular markers, natural history, and evolution. 2nd Edition*. Sunderland, MA: Sinauer Associates.
- Avise, J. C., & Mank, J. E. (2009). Evolutionary perspectives on hermaphroditism in fishes. *Sexual Development*, 3(2-3), 152-163.
- Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., . . . Vamosi, J. C. (2014). Sex determination: why so many ways of doing it? *PLoS Biology*, 12(7), e1001899.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., . . . Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3(10), e3376.
- Ballard, J. W., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4), 729-744.

- Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., . . . Philippe, H. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences*, 99(3), 1414-1419.
- Baroiller, J. F., Clota, F., & Geraz, E. (1995). Temperature sex determination in two tilapia species, *Oreochromis niloticus* and the red tilapia (Red Florida strain): effect of high or low temperatures. In F. W. Goetz & P. Thomas (Eds.), *Proceedings of the 5th International Symposium on the Reproductive Physiology of Fish*. Austin, Texas: The University of Texas, Austin.
- Baroiller, J. F., & D'Cotta, H. (2001). Environment and sex determination in farmed fish. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 130(4), 399-409.
- Baroiller, J. F., D'Cotta, H., & Saillant, E. (2009a). Environmental effects on fish sex determination and differentiation. *Sexual Development*, 3(2-3), 118-135.
- Baroiller, J. F., D'Cotta, H., Bezault, E., Wessels, S., & Hoerstgen-Schwark, G. (2009b). Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 153(1), 30-38.
- Baroiller, J. F., & Guiguen, Y. (2001). Endocrine and environmental aspects of sex differentiation in gonochoristic fish. In G. Scherer & M. Schmid (Eds.), *Genes and Mechanisms in Vertebrate Sex Determination* (pp. 177-201): Birkhauser Verlag Basel, Switzerland.
- Baroiller, J. F., Guiguen, Y., & Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *Cellular and Molecular Life Sciences*, 55(7), 910.
- Baroiller, J. F., Nakayama, I., Foresti, F., & Chourrout, D. (1996). Sex determination studies in two species of teleost fish, *Oreochromis niloticus* and *Leporinus elongatus*. *Zoological Studies*, 35(4), 279-285.
- Barón, B. S., Bückle, R. F., & Espina, S. (2002). Environmental factors and sexual differentiation in *Poecilia sphenops* Valenciennes (Pisces: Poeciliidae). *Aquaculture Research*, 33(8), 615-619.
- Baxter, S. W., Davey, J. W., Johnston, J. S., Shelton, A. M., Heckel, D. G., Jiggins, C. D., & Blaxter, M. L. (2011). Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One*, 6(4), e19315.
- Benton, M. J., & Ayala, F. J. (2003). Dating the tree of life. *Science*, 300(5626), 1698-1700.
- Benton, M. J., Donohue, P. C. J., & Asher, R. J. (2009). Calibrating and constraining molecular clocks. In S. B. Hedges & S. Kumar (Eds.), *The Timetree of Life* (pp. 35-86): Oxford University Press.

- Bergey, C. M., Pozzi, L., Disotell, T. R., & Burrell, A. S. (2013). A New Method for Genome-wide Marker Development and Genotyping Holds Great Promise for Molecular Primatology. *International Journal of Primatology*, 34(2), 303-314.
- Bergthorsson, U., Adams, K. L., Thomason, B., & Palmer, J. D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, 424(6945), 197-201.
- Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., & Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. *Nature*, 348(6300), 448-450.
- Betancur, R. R., Li, C., Munroe, T. A., Ballesteros, J. A., & Orti, G. (2013). Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology*, 62(5), 763-785.
- Bigelow, H. B., & Schroeder, W. C. (1950). New and little known cartilaginous fishes from the Atlantic. *Bulletin of the Museum of Comparative Zoology at Harvard College*, 103(7), 385-409.
- Bigelow, H. B., & Schroeder, W. C. (1953). Sharks, sawfishes, guitarfishes, skates and rays. Chimaeroids. In J. Tee-Van, C. M. Breder, S. F. Hildebrand, A. E. Parr, & W. C. Schroeder (Eds.), *Fishes of the Western North Atlantic. Part 2*. (pp. 1-514). Yale University, New Haven: Sears Foundation for Marine Research.
- Blanquart, S., & Lartillot, N. (2008). A site- and time-heterogeneous model of amino acid replacement. *Molecular Biology and Evolution*, 25(5), 842-858.
- Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4, e1660.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., . . . Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 10(4), e1003537.
- Bradley, K. M., Breyer, J. P., Melville, D. B., Broman, K. W., Knapik, E. W., & Smith, J. R. (2011). An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3 (Bethesda)*, 1(1), 3-9.
- Brandley, M. C., Wang, Y., Guo, X., de Oca, A. N., Feria-Ortiz, M., Hikida, T., & Ota, H. (2011). Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of Plestiodon (Eumeces) lizards. *Systematic Biology*, 60(1), 3-15.
- Bromham, L. D., & Hendy, M. D. (2000). Can fast early rates reconcile molecular dates with the Cambrian explosion? *Proceedings of the Royal Society B: Biological Sciences*, 267(1447), 1041-1047.

- Bromham, L., & Penny, D. (2003). The modern molecular clock. *Nature Reviews Genetics*, 4(3), 216-224.
- Brykov, V. A. (2015). Mechanisms of sex determination in fish: evolutionary and practical aspects. *Russian Journal of Marine Biology*, 40(6), 407-417.
- Bull, J. J. (1980). Sex determination in reptiles. *The Quarterly Review of Biology*, 55(1), 3-21.
- Bull, J. J. (1983). *Evolution of sex determining mechanisms*. Menlo Park, California: The Benjamin/Cummings Publishing Company, Inc.
- Bull, J. J. (1985). Sex determining mechanisms: An evolutionary perspective. *Experientia*, 41(10), 1285-1296.
- Burtis, K. C., & Baker, B. S. (1989). Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, 56(6), 997-1010.
- Cao, Y., Janke, A., Waddell, P. J., Westerman, M., Takenaka, O., Murata, S., . . . Hasegawa, M. (1998). Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *Journal of Molecular Evolution*, 47(3), 307-322.
- Cappetta, H. (2012). Handbook of Paleoichthyology, Vol. 3E: Chondrichthyes · Mesozoic and Cenozoic Elasmobranchii: Teeth. *Verlag Dr. Friedrich Pfeil*, 1-512.
- Carmichael, S. N., Bekaert, M., Taggart, J. B., Christie, H. R., Bassett, D. I., Bron, J. E., . . . Sturm, A. (2013). Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. *PLoS One*, 8(10), e77832.
- Carranza, S., Arnold, E. N., Mateo, J. A., & Geniez, P. (2002). Relationships and evolution of the North African geckos, *Geckonia* and *Tarentola* (Reptilia: Gekkonidae), based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, 23(2), 244-256.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes|Genomes|Genetics*, 1(3), 171-182.
- Chang, C. Y., & Witschi, E. (1956). Genic control and hormonal reversal of sex differentiation in *Xenopus*. *Proceedings of the Society for Experimental Biology and Medicine*, 93(1), 140-144.
- Chapman, D. D., Firchau, B., & Shivji, M. S. (2008). Parthenogenesis in a large-bodied requiem shark, the blacktip *Carcharhinus limbatus*. *Journal of Fish Biology*, 73(6), 1473-1477.

- Chapman, D. D., Shivji, M. S., Louis, E., Sommer, J., Fletcher, H., & Prodohl, P. A. (2007). Virgin birth in a hammerhead shark. *Biology Letters*, 3(4), 425-427.
- Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, 6(2), 149-162.
- Charlesworth, D., & Mank, J. E. (2010). The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. *Genetics*, 186(1), 9-31.
- Chen, J., Fu, Y., Xiang, D., Zhao, G., Long, H., Liu, J., & Yu, Q. (2008a). XX/XY heteromorphic sex chromosome systems in two bullhead catfish species, *Liobagrus marginatus* and *L. styani* (Amblycipitidae, Siluriformes). *Cytogenetic and Genome Research*, 122(2), 169-174.
- Chen, M., Zou, M., Yang, L., & He, S. (2012). Basal jawed vertebrate phylogenomics using transcriptomic data from Solexa sequencing. *PLoS One*, 7(4), e36256.
- Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., . . . Wang, J. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nature Genetics*, 46(3), 253-260.
- Chen, S.-L., Deng, S.-P., Ma, H.-Y., Tian, Y.-S., Xu, J.-Y., Yang, J.-F., . . . Zhai, J.-m. (2008b). Molecular marker-assisted sex control in half-smooth tongue sole (*Cynoglossus semilaevis*). *Aquaculture*, 283(1-4), 7-12.
- Chen, S.-L., Li, J., Deng, S.-P., Tian, Y.-S., Wang, Q.-Y., Zhuang, Z.-M., . . . Xu, J.-Y. (2007). Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). *Marine Biotechnology*, 9(2), 273-280.
- Chen, S. L., Tian, Y. S., Yang, J. F., Shao, C. W., Ji, X. S., Zhai, J. M., . . . Wang, N. (2009). Artificial gynogenesis and sex determination in half-smooth tongue sole (*Cynoglossus semilaevis*). *Marine Biotechnology*, 11(2), 243-251.
- Chen, T. R., & Reisman, H. M. (1970). A comparative chromosome study of the North American species of sticklebacks (Teleostei: Gasterosteidae). *Cytogenetics*, 9(5), 321-332.
- Chiu, C. H., Amemiya, C., Dewar, K., Kim, C. B., Ruddle, F. H., & Wagner, G. P. (2002). Molecular evolution of the HoxA cluster in the three major gnathostome lineages. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 99(8), 5492-5497.
- Chiu, C. H., Dewar, K., Wagner, G. P., Takahashi, K., Ruddle, F., Ledje, C., . . . Amemiya, C. T. (2004). Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Research*, 14(1), 11-17.
- Cioffi, M. B., Liehr, T., Trifonov, V., Molina, W. F., & Bertollo, L. A. (2013). Independent sex chromosome evolution in lower vertebrates: a molecular cytogenetic overview in the Erythrinidae fish family. *Cytogenetic and Genome Research*, 141(2-3), 186-194.

- Claeson, K. M., Underwood, C. J., & Ward, D. J. (2013). †*Tingitanius tenuimandibulus*, a new platyrhinid batoid from the Turonian (Cretaceous) of Morocco and the cretaceous radiation of the Platyrhinidae. *Journal of Vertebrate Paleontology*, 33(5), 1019-1036.
- Cnaani, A., Lee, B. Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., . . . Kocher, T. D. (2008). Genetics of sex determination in tilapia species. *Sexual Development*, 2(43-54).
- Coates, M. I., Gess, R. W., Finarelli, J. A., Criswell, K. E., & Tietjen, K. (2017). A symmoriiform chondrichthyan braincase and the origin of chimaeroid fishes. *Nature*, 541(7636), 208-211.
- Coates, M. I., & Sequeira, S. E. K. (2001). A new stethacanthid chondrichthyan from the lower Carboniferous of Bearsden, Scotland. *Journal of Vertebrate Paleontology*, 21(3), 438-459.
- Cole, N. J., & Currie, P. D. (2007). Insights from sharks: evolutionary and developmental models of fin development. *Developmental Dynamics*, 236(9), 2421-2431.
- Combosch, D. J., & Vollmer, S. V. (2015). Trans-Pacific RAD-seq population genomics confirms introgressive hybridization in Eastern Pacific Pocillopora corals. *Molecular Phylogenetics and Evolution*, 88, 154-162.
- Compagno, L. J. V., Stehmann, M., & Ebert, D. A. (1990). *Rhinochimaera africana*, a new longnose chimaera from southern Africa, with comments on the systematics and distribution of the genus *Rhinochimaera* Garman, 1901 (Chondrichthyes, Chimaeriformes, Rhinochimaeridae). *South African Journal of Marine Science*, 9, 201-222.
- Conant, G. C., & Lewis, P. O. (2001). Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. *Molecular Biology and Evolution*, 18(6), 1024-1033.
- Conroy, C. J., & van Tuinen, M. (2003). Extracting time from phylogenies: positive interplay between fossil and genetic data. *Journal of Mammalogy*, 84(2), 444-455.
- Crews, D. (1996). Temperature-Dependent Sex Determination: The Interplay of Steroid Hormones and Temperature. *Zoological Science*, 13(1), 1-13.
- Cruz, V. P. d., Shimabukuro-Dias, C. K., Oliveira, C., & Foresti, F. (2011). Karyotype description and evidence of multiple sex chromosome system X1X1X2X2/X1X2Y in *Potamotrygon* aff. *motoro* and *P. falkneri* (Chondrichthyes: Potamotrygonidae) in the upper Paraná River basin, Brazil. *Neotropical Ichthyology*, 9, 201-208.
- Cummings, M. P., Otto, S. P., & Wakeley, J. (1995). Sampling properties of DNA sequence data in phylogenetic analysis. *Molecular Biology and Evolution*, 12(5), 814-822.

- Cunningham, C. W. (1997). Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology*, 46(3), 464-478.
- Cuny, G., Rieppel, O., & Sander, P. M. (2001). The shark fauna from the Middle Triassic (Anisian) of North-Western Nevada. *Zoological Journal of the Linnean Society*, 133(3), 285-301.
- Cutting, A., Chue, J., & Smith, C. A. (2013). Just how conserved is vertebrate sex determination? *Developmental Dynamics*, 242(4), 380-387.
- Daga, R. R., Thode, G., & Amores, A. (1996). Chromosome complement, C-banding, Ag-NOR and replication banding in the zebrafish *Danio rerio*. *Chromosome Research*, 4(1), 29-32.
- Davey, A. J. H., & Jellyman, D. J. (2005). Sex determination in freshwater eels and management options for manipulation of sex. *Reviews in Fish Biology and Fisheries*, 15(1-2), 37-52.
- Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*, 22(11), 3151-3164.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499-510.
- Davidson, W. S., Huang, T. K., Fujiki, K., von Schalburg, K. R., & Koop, B. F. (2009). The sex determining loci and sex chromosomes in the family salmonidae. *Sexual Development*, 3(2-3), 78-87.
- de Beer, G. R., & Moy-Thomas, J. A. (1935). V I.—On the skull of Holocephali. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 224(514), 287.
- de La Cruz-Agüero, J., García-Rodríguez, F. J., Cota-Gómez, V. M., Melo-Barrera, F. N., & González-Armas, R. (2012). Morphometric and molecular data on two mitochondrial genes of a newly discovered chimaeran fish (*Hydrolagus melanophasma*, Chondrichthyes). *Ocean Science Journal*, 47(2), 147-153.
- de Souza Valentim, F. C., Porto, J. I., Bertollo, L. A., Gross, M. C., & Feldberg, E. (2013). XX/XO, a rare sex chromosome system in *Potamotrygon* freshwater stingray from the Amazon Basin, Brazil. *Genetica*, 141(7-9), 381-387.
- de Valentim, F. C., Falcao, J., Porto, J. I., & Feldberg, E. (2006). Chromosomes of three freshwater stingrays (Rajiformes Potamotrygonidae) from the Rio Negro basin, Amazon, Brazil. *Genetica*, 128(1-3), 33-39.
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361-375.

- Delvin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*, 208, 191-364.
- Demming, D. C. (2004). Prevalence of TSD in Crocodylians. In N. Valenzuela & V. A. Lance (Eds.), *Temperature-Dependent Sex Determination in Vertebrates* (pp. 33-41). Washington, D.C.: Smithsonian Institution Press.
- Desprez, D., & Mélard, C. (1998). Effect of ambient water temperature on sex determinism in the blue tilapia *Oreochromis aureus*. *Aquaculture*, 162(1-2), 79-84.
- Diaz, N., & Piferrer, F. (2015). Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC Genomics*, 16, 679.
- Didier, D. A. (1995). Phylogenetic systematics of extant chimaeroid fishes (Holocephali, Chimaeroidei). *American Museum Novitates*, 3119, 1-86.
- Didier, D. A., Kemper, J. M., & Ebert, D. A. (2012). Phylogeny, Biology, and Classification of Extant Holocephalans. In J. M. Carrier, J.; Heithaus, M. (Ed.), *Biology of sharks and their relatives, edition 2* (pp. 97-124). Boca Raton, FL: CRC Press.
- Didier, D. A., Last, P. R., & White, W. T. (2008). Three new species of the genus *Chimaera* Linnaeus (Chimaeriformes: Chimaeridae) from Australia. *CSIRO Marine and Atmospheric Research Paper*, 22, 327-340.
- Donahue, W. H. (1974). A karyotypic study of three species of Rajiformes (Chondrichthyes, Pisces). *Canadian Journal of Genetics and Cytology*, 16(1), 203-211.
- Dornburg, A., Beaulieu, J. M., Oliver, J. C., & Near, T. J. (2011). Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Systematic Biology*, 60(4), 519-527.
- Dornburg, A., Brandley, M. C., McGowen, M. R., & Near, T. J. (2012). Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution*, 29(2), 721-736.
- Dornburg, A., Townsend, J. P., Friedman, M., & Near, T. J. (2014). Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol Biol*, 14(1), 169.
- Douady, C. J., Dosay, M., Shivji, M. S., & Stanhope, M. J. (2003). Molecular phylogenetic evidence refuting the hypothesis of Batoidea (rays and skates) as derived sharks. *Molecular Phylogenetics and Evolution*, 26, 215-221.
- Driskell, A. C., Ane, C., Burleigh, J. G., McMahon, M. M., O'Meara B, C., & Sanderson, M. J. (2004). Prospects for building the tree of life from large sequence databases. *Science*, 306(5699), 1172-1174.

- Drummond, A. J., Ho, S. Y., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology*, *4*(5), e88.
- Dudgeon, C. L., Coulton, L., Bone, R., Ovenden, J. R., & Thomas, S. (2017). Switch from sexual to parthenogenetic reproduction in a zebra shark. *Scientific Reports*, *7*, 40537.
- Duffin, C. J., & Ward, D. J. (1983). Teeth of a new neoselachian shark from the British lower Jurassic. *Palaeontology*, *26*(4), 839-844.
- Ebersberger, I., Strauss, S., & von Haeseler, A. (2009). HaMStR: profile hidden markov model based search for orthologs in ESTs. *BMC Evol Biol*, *9*, 157.
- Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the Academy of Natural Sciences of Philadelphia*, *107*(37), 16196-16200.
- Eshel, O., Shirak, A., Dor, L., Band, M., Zak, T., Markovich-Gordon, M., . . . Ron, M. (2014). Identification of male-specific amh duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics*, *15*, 774.
- Eshel, O., Shirak, A., Weller, J. I., Slossman, T., Hulata, G., Cnaani, A., & Ron, M. (2011). Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Animal Genetics*, *42*(2), 222-224.
- Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology*, *772*, 157-178.
- Ezaz, T., Sarre, S. D., O'Meally, D., Graves, J. A., & Georges, A. (2009). Sex chromosome evolution in lizards: independent origins and rapid transitions. *Cytogenetic and Genome Research*, *127*(2-4), 249-260.
- Feldheim, K. A., Chapman, D. D., Sweet, D., Fitzpatrick, S., Prodohl, P. A., Shivji, M. S., & Snowden, B. (2010). Shark virgin birth produces multiple, viable offspring. *Journal of Heredity*, *101*(3), 374-377.
- Felip, A., Young, W. P., Wheeler, P. A., & Thorgaard, G. H. (2005). An AFLP-based approach for the identification of sex-linked markers in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, *247*(1-4), 35-43.
- Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics*, *22*, 521-565.
- Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.

- Ferguson-Smith, M. (2007). The evolution of sex chromosomes and sex determination in vertebrates and the key role of DMRT1. *Sexual Development*, 1(1), 2-11.
- Fields, A. T., Feldheim, K. A., Poulakis, G. R., & Chapman, D. D. (2015). Facultative parthenogenesis in a critically endangered wild vertebrate. *Current Biology*, 25(11), R446-447.
- Finnilä, S., Lehtonen, M. S., & Majamaa, K. (2001). Phylogenetic network for European mtDNA. *American Journal of Human Genetics*, 68(6), 1475-1484.
- Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., . . . Nusbaum, C. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *genome Biology*, 12(1), R1.
- Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic zoology*, 19(2), 99-113.
- Foresti, F., Oliveira, C., & Foresti de Almeida-Toledo, L. (1993). A method for chromosome preparations from large fish specimens using in vitro short-term treatment with colchicine. *Experientia*, 49(9), 810-813.
- Foster, J. W., Brennan, F. E., Hampikian, G. K., Goodfellow, P. N., Sinclair, A. H., Lovell-Badge, R., . . . Marshall Graves, J. A. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. *Nature*, 359(6395), 531-533.
- Foster, P. G. (2004). Modeling compositional heterogeneity. *Systematic Biology*, 53(3), 485-495.
- Foster, P. G., & Hickey, D. A. (1999). Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, 48(3), 284-290.
- Fowler, B. L., & Buonaccorsi, V. P. (2016). Genomic characterization of sex-identification markers in *Sebastes carnatus* and *Sebastes chrysomelas* rockfishes. *Molecular Ecology*, 25(10), 2165-2175.
- Francis, R. C. (1984). The effects of bidirectional selection for social dominance on agonistic behavior and sex ratios in the paradise fish (*Macropodus opercularis*). *Behaviour*, 90(1), 25-44.
- Francis, R. C. (1990). Temperament in a fish: a longitudinal study of the development of individual differences in aggression and social rank in the midas cichlid. *Ethology*, 86(4), 311-325.
- Francis, R. C., & Barlow, G. W. (1993). Social control of primary sex differentiation in the Midas cichlid. *Proceedings of the National Academy of Sciences of the United States of America*, 90, 10673-10675.

- Frandsen, P. B., Calcott, B., Mayer, C., & Lanfear, R. (2015). Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. *BMC Evolutionary Biology*, *15*, 13.
- Furlan, E., Stoklosa, J., Griffiths, J., Gust, N., Ellis, R., Huggins, R. M., & Weeks, A. R. (2012). Small population size and extremely low levels of genetic diversity in island populations of the platypus, *Ornithorhynchus anatinus*. *Ecology and Evolution*, *2*(4), 844-857.
- Galtier, N., & Gouy, M. (1995). Inferring phylogenies from DNA sequences of unequal base compositions. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 11317-11321.
- Galtier, N., & Gouy, M. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Molecular Biology and Evolution*, *15*(7), 871-879.
- Galtier, N., Nabholz, B. N., Glemin, S., & Hurst, G. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, *18*(22), 4541-4550.
- Gamble, T. (2016). Using RAD-seq to recognize sex-specific markers and sex chromosome systems. *Molecular Ecology*, *25*(10), 2114-2116.
- Gamble, T., Castoe, T. A., Nielsen, S. V., Banks, J. L., Card, D. C., Schield, D. R., . . . Booth, W. (2017). The discovery of XY sex chromosomes in a boa and python. *Current Biology*, *27*(14), 2148-2153.
- Gamble, T., Coryell, J., Ezaz, T., Lynch, J., Scantlebury, D. P., & Zarkower, D. (2015). Restriction site-associated DNA sequencing (RAD-seq) reveals an E]xtraordinary number of transitions among gecko sex-determining systems. *Molecular Biology and Evolution*, *32*(5), 1296-1309.
- Gamble, T., & Zarkower, D. (2014). Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Molecular Ecology Resources*, *14*(5), 902-913.
- Garamszegi, L. Z., & Gonzalez-Voyer, A. (2014). Working with the Tree of Life in Comparative Studies: How to Build and Tailor Phylogenies to Interspecific Datasets. In L. Z. Garamszegi (Ed.), *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice* (pp. 19-48). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Garman, S. (1901). Genera and families of the chimaeroids. *Proceedings of the New England Zoological Club*, *2*, 75-77.
- Ghigliotti, L., Cheng, C.-H. C., & Pisano, E. (2014). Sex determination in Antarctic notothenioid fish: chromosomal clues and evolutionary hypotheses. *Polar Biology*, *39*(1), 11-22.

- Gill, T. (1862). Note on some genera of fishes of western North America. *Proceedings of the Academy of Natural Sciences of Philadelphia*, 14, 329-332.
- Godwin, J. (2009). Social determination of sex in reef fishes. *Seminars in Cell & Developmental Biology*, 20(3), 264-270.
- Goode, G. B., & Bean, T. H. (1895). On Harriotta, a new type of chimaeroid fish from the deeper waters of the northwestern Atlantic. *Proceedings of the United States National Museum*, 17(1014), 471-473.
- Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., . . . Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, 15(6), 1304-1315.
- Griffiths, R., & Orr, K. (1999). The use of amplified fragment length polymorphism (AFLP) in the isolation of sex-specific markers. *Molecular Ecology*, 8(4), 671-674.
- Grogan, E. D., Lund, R., & Greenfest-Allen, E. (2012). The Origin and Relationships of Early Chondrichthyans. In J. Carrier, J. Musick, & M. Heithaus (Eds.), *Biology of Sharks and their Relatives, Edition 2* (pp. 3-30). Boca Raton, FL: CRC Press.
- Grutzner, F., Rens, W., Tsend-Ayush, E., El-Mogharbel, N., O'Brien, P. C., Jones, R. C., . . . Marshall Graves, J. A. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature*, 432(7019), 913-917.
- Guan, G., Kobayashi, T., & Nagahama, Y. (2000). Sexually dimorphic expression of two types of DM (Doublesex/Mab-3)-domain genes in a teleost fish, the tilapia (*Oreochromis niloticus*). *Biochemical and Biophysical Research Communications*, 272(3), 662-666.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., . . . Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346(6281), 245-250.
- Haag, E. S., & Doty, A. V. (2005). Sex determination across evolution: connecting the dots. *PLoS Biology*, 3(1), e21.
- Hamaguchi, S., Toyazaki, Y., Shinomiya, A., & Sakaizumi, M. (2004). The XX-XY sex-determination system in *Oryzias luzonensis* and *O. mekongensis* revealed by the sex ratio of the progeny of sex-reversed fish. *Zoological Science*, 21(10), 1015-1018.
- Harlow, P. S. (2004). Temperature-dependent sex determination in lizards. In N. Valenzuela & V. A. Lance (Eds.), *Temperature-dependent sex determination in vertebrates* (pp. 42-52). Washington, D.C.: Smithsonian Books.
- Harmon, T. S., Kamerman, T. Y., Corwin, A. L., & Sellas, A. B. (2016). Consecutive parthenogenetic births in a spotted eagle ray *Aetobatus narinari*. *Journal of Fish Biology*, 88(2), 741-745.

- Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., . . . Strussmann, C. A. (2012). A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), 2955-2959.
- Hattori, R. S., Oura, M., Sakamoto, T., Yokota, M., Watanabe, S., & Strüssmann, C. A. (2010). Establishment of a strain inheriting a sex-linked SNP marker in Patagonian pejerrey (*Odontesthes hatcheri*), a species with both genotypic and temperature-dependent sex determination. *Animal Genetics*, 41(1), 81-84.
- Hayes, T. B. (1998). Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. *Journal of Experimental Zoology*, 281(5), 373-399.
- Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3), 239-257.
- Heinicke, M. P., Naylor, G. J. P., & Hedges, S. B. (2009). Cartilaginous fishes (Chondrichthyes). In S. B. Hedges & S. Kumar (Eds.), *The Timetree of Life*.
- Heiskanen, M., Peltonen, L., & Palotif, A. (1996). Visual mapping by high resolution FISH. *Trends in Genetics*, 12(10), 379-382.
- Heist, E. J. (2012). Genetics of Sharks, Skates, and Rays. In J. C. Carrier, J. A. Musick, & M. R. Heithaus (Eds.), *Biology of Sharks and their Relatives, Edition 2* (pp. 487-504). Boca Raton, Florida: CRC Press.
- Helfman, G., Collette, B. B., Facey, D. E., & Bowen, B. W. (2009). *The Diversity of Fishes: Biology, Evolution, and Ecology, 2nd Edition*: Wiley-Blackwell.
- Hess, J. E., Campbell, N. R., Close, D. A., Docker, M. F., & Narum, S. R. (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, 22(11), 2898-2916.
- Heule, C., Salzburger, W., & Bohne, A. (2014). Genetics of sexual development: an evolutionary playground for fish. *Genetics*, 196(3), 579-591. doi:10.1534/genetics.114.161158
- Hillis, D. M., & Green, D. M. (1990). Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. *Journal of Evolutionary Biology*, 3(1-2), 49-64.
- Hillis, D. M., & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology*, 42(2), 182-192.
- Ho, S. Y., & Jermini, L. (2004). Tracing the decay of the historical signal in biological sequence data. *Systematic Biology*, 53(4), 623-637.
- Ho, S. Y., & Larson, G. (2006). Molecular clocks: when times are a-changin'. *Trends in Genetics*, 22(2), 79-83.

- Ho, S. Y. W., & Phillips, M. J. (2009). Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, 58(3), 367-380.
- Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, 16(5), 1264-1278.
- Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, 11, 117-122.
- Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLOS Genetics*, 6(2), e1000862.
- Hohenlohe, P. A., Catchen, J., & Cresko, W. A. (2012). Population Genomic Analysis of Model and Nonmodel Organisms Using Sequenced RAD Tags. In F. Pompanon & A. Bonin (Eds.), *Data Production and Analysis in Population Genomics: Methods and Protocols* (pp. 235-260). Totowa, NJ: Humana Press.
- Holland, B. S., & Hadfield, M. G. (2004). Origin and diversification of the endemic Hawaiian tree snails (Achatinellidae: Achatinellinae) based on molecular evidence. *Molecular Phylogenetics and Evolution*, 32(2), 588-600.
- Hug, L. A., & Roger, A. J. (2007). The impact of fossils and taxon sampling on ancient molecular dating analyses. *Molecular Biology and Evolution*, 24(8), 1889-1897.
- Hunter, G. A., Donaldson, E. M., Goetz, F. W., & Edgell, P. R. (1982). Production of all-female and sterile coho salmon, and experimental evidence for male heterogamety. *Transactions of the American Fisheries Society*, 111(3), 367-372.
- Hunter, G. A., Donaldson, E. M., Stoss, J., & Baker, I. (1983). Genetics in aquaculture production of monosex female groups of chinook salmon (*Oncorhynchus tshawytscha*) by the fertilization of normal ova with sperm from sex-reversed females. *Aquaculture*, 33(1), 355-364.
- Ida, H., Murofushi, M., Fujiwara, S.-i., & Fujino, K. (1978). Preparation of fish chromosomes by *in vitro* colchicine treatment. *Japanese Journal of Ichthyology*, 24(4), 281-284.
- Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., . . . Roest Crolius, H. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431(7011), 946-957.
- Janvier, P. (1996). *Early Vertebrates*. Oxford: Oxford University Press.

- Janzen, F. J., & Krenz, J. G. (2004). Phylogenetics: Which was first, TSD or GSD? In N. Valenzuela & V. A. Lance (Eds.), *Temperature-Dependent Sex Determination in Vertebrates* (pp. 121-130). Washington, DC: Smithsonian Books.
- Janzen, F. J., & Paukstis, G. L. (1991). Environmental sex determination in reptiles: ecology, evolution, and experimental design. *The Quarterly Review of Biology*, *66*(2), 149-179.
- Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, *22*(4), 225-231.
- Johnson, P. A., Kent, T. R., Urick, M. E., & Giles, J. R. (2008). Expression and regulation of anti-mullerian hormone in an oviparous species, the hen. *Biology of Reproduction*, *78*(1), 13-19.
- Johnstone, R., Simpson, T. H., Youngson, A. F., & Whitehead, C. (1979). Sex reversal in salmonid culture. *Aquaculture*, *18*(1), 13-19.
- Johnstone, R., & Youngson, A. F. (1984). The progeny of sex-inverted female Atlantic salmon (*Salmo salar* L.). *Aquaculture*, *37*(2), 179-182.
- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275-282.
- Josso, N., di Clemente, N., & Gouedard, L. (2001). Anti-Mullerian hormone and its receptors. *Molecular and Cellular Endocrinology*, *179*, 25-32.
- Kallman, K. D. (1973). The Sex-Determining Mechanism of the Platyfish, *Xiphophorus maculatus*. In J. H. Schröder (Ed.), *Genetics and Mutagenesis of Fish* (pp. 19-28). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., . . . Kikuchi, K. (2012). A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLOS Genetics*, *8*(7), e1002798.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772-780.
- Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., . . . Seehausen, O. (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, *22*(11), 2848-2863.
- Kemper, J. M., Ebert, D. A., Naylor, G. J. P., & Didier, D. A. (2014). *Chimaera carophila* (Chondrichthyes: Chimaeriformes: Chimaeridae), a new species of chimaera from New Zealand. *Bulletin of Marine Science*, *91*(1), 63-81.

- Kettlewell, J. R., Raymond, C. S., & Zarkower, D. (2000). Temperature-dependent expression of turtle *Dmrt1* prior to sexual differentiation. *genesis*, 26(3), 174-178.
- Kikuchi, K., Kai, W., Hosokawa, A., Mizuno, N., Suetake, H., Asahina, K., & Suzuki, Y. (2007). The sex-determining locus in the tiger pufferfish, *Takifugu rubripes*. *Genetics*, 175(4), 2039-2042.
- Kikugawa, K., Katoh, K., Kuraku, S., Sakurai, H., Ishida, O., Iwabe, N., & Miyata, T. (2004). Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biology*, 2(1), 3.
- Kikuno, T., & Ojima, Y. (1987). A karyotypic studies of a guitar fish, *Rhinobatos hyinnicephalus* Richardson (Pisces, Rajiformes). *La Kromosomo*, 2(47-48), 1538-1544.
- Kikuno, T., & Ojima, Y. (1987). A karyotypic studies of a guitarfish, *Rhinobatus hynnicephalus* Richardson (Pisces, Rajiformes). *La Kromosomo II*, 47-48, 1538-1544.
- Kitano, J., & Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental Biology of Fishes*, 94(3), 549-558.
- Klug, S. (2010). Monophyly, phylogeny and systematic position of the †Synchondontiformes (Chondrichthyes, Neoselachii). *Zoologica Scripta*, 39(1), 37-49.
- Kobayashi, T., & Nagahama, Y. (2009). Molecular aspects of gonadal differentiation in a teleost fish, the Nile tilapia. *Sexual Development*, 3(2-3), 108-117.
- Kolaczkowski, B., & Thornton, J. W. (2008). A mixed branch length model of heterotachy improves phylogenetic accuracy. *Molecular Biology and Evolution*, 25(6), 1054-1066.
- Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., . . . Schartl, M. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Research*, 16(7), 815-826.
- Kondo, M., Nanda, I., Hornung, U., Asahina, S., Shimizu, N., Mitani, H., . . . Schartl, M. (2003). Absence of the candidate male sex-determining gene *dmrt1b(Y)* of medaka from other fish species. *Current Biology*, 13, 416-420.
- Kondo, M., Nanda, I., Hornung, U., Schmid, M., & Schartl, M. (2004). Evolutionary origin of the medaka Y chromosome. *Current Biology*, 14(18), 1664-1669.
- Kooke, R., Johannes, F., Wardenaar, R., Becker, F., Etcheverry, M., Colot, V., . . . Keurentjes, J. J. (2015). Epigenetic basis of morphological variation and phenotypic plasticity in *Arabidopsis thaliana*. *Plant Cell*, 27(2), 337-348.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature*, 351(6322), 117-121.

- Koopman, P., Munsterberg, A., Capel, B., Vivian, N., & Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature*, *348*(6300), 450-452.
- Koshimizu, E., Strüssmann, C. A., Okamoto, N., Fukuda, H., & Sakamoto, T. (2010). Construction of a genetic map and development of DNA markers linked to the sex-determining locus in the Patagonian pejerrey (*Odontesthes hatcheri*). *Marine Biotechnology*, *12*(1), 8-13.
- Kraak, S. B. M., & De Looze, E. M. A. (1992). A new hypothesis on the evolution of sex determination in vertebrates; big females ZW, big males XY. *Netherlands Journal of Zoology*, *43*(3), 260-273.
- Krefft, G. (1990). Chimaeridae. In J. C. Quero, J. C. Jureau, C. P. Karrer, & L. Saldanha (Eds.), *Check-list of the fishes of the eastern tropical Atlantic* (pp. 111-113): Unesco.
- Kruger, F. (2012). Trim Galore! Available at: [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/].
- Kuck, P., & Longo, G. C. (2014). FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology*, *11*(1), 81.
- Lacepède, B. G. E. (1798). Histoire naturelle des poissons. *I*, 1-532.
- Lampert, K. P., & Scharl, M. (2008). The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philosophical Transactions of the Royal Society B*, *363*(1505), 2901-2909.
- Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695-1701.
- Lang, J. W., & Andrews, H. V. (1994). Temperature-dependent sex determination in crocodylians. *The Journal of Experimental Zoology*, *270*, 28-44.
- Larsson, T. A., Tay, B. H., Sundstrom, G., Fredriksson, R., Brenner, S., Larhammar, D., & Venkatesh, B. (2009). Neuropeptide Y-family peptides and receptors in the elephant shark, *Callorhynchus milii* confirm gene duplications before the gnathostome radiation. *Genomics*, *93*(3), 254-260.
- Lartillot, N., & Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, *21*(6), 1095-1109.
- Le, H. L., Lecointre, G., & Perasso, R. (1993). A 28S rRNA-based phylogeny of the gnathostomes: first steps in the analysis of conflict and congruence with morphologically based cladograms. *Molecular Phylogenetics and Evolution*, *3*(1), 31-51.
- Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, *25*(7), 1307-1320.

- Lee, B. Y., Penman, D. J., & Kocher, T. D. (2003). Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics*, *34*(5), 379-383.
- Levinton, J. S. (2008). The Cambrian explosion: how do we use the evidence. *BioScience*, *58*(9), 855-864. doi:10.1641/B580912
- Li, C., Hofreiter, M., Straube, N., Corrigan, S., & Naylor, G. J. (2013). Capturing protein-coding genes across highly divergent species. *Biotechniques*, *54*(6), 321-326.
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., . . . Wang, D. (2015). A tandem duplicate of anti-mullerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLOS Genetics*, *11*(11), e1005678.
- Liao, X., Xu, G., & Chen, S. L. (2014). Molecular method for sex identification of half-smooth tongue sole (*Cynoglossus semilaevis*) using a novel sex-linked microsatellite marker. *International Journal of Molecular Sciences*, *15*(7), 12952-12958.
- Liew, W. C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K. R., & Orban, L. (2012). Polygenic sex determination system in zebrafish. *PLoS One*, *7*(4), e34397.
- Linnaeus, C. (1758). *Systema Naturae per regna tria naturae, regnum animale, secundum classes, ordines, genera, species, cum characteribus differentiis synonymis, locis. Ed. X.* Stockholm (L. SALVIUS).
- Liu, Y. G., Bao, B. L., Liu, L. X., Wang, L., & Lin, H. (2008). Isolation and characterization of polymorphic microsatellite loci from RAPD product in half-smooth tongue sole (*Cynoglossus semilaevis*) and a test of cross-species amplification. *Molecular Ecology Resources*, *8*(1), 202-204.
- Liu, Y. G., Sun, X. Q., Gao, H., & Liu, L. X. (2007). Microsatellite markers from an expressed sequence tag library of half-smooth tongue sole (*Cynoglossus semilaevis*) and their application in other related fish species. *Molecular Ecology Notes*, *7*(6), 1242-1244.
- Lockhart, P. J., Howe, C. J., Bryant, D. A., Beanland, T. J., & Larkum, A. W. D. (1992). Substitutional bias confound inference of cyanelle origins from sequence data. *Journal of Molecular Evolution*, *34*, 153-162.
- Lockhart, P. J., Steel, M. A., Hendy, M. D., & Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution*, *11*(4), 605-612.
- Lopez, D. M., Sigel, M. M., & Lee, J. C. (1974). Phylogenetic studies on T cells: I. lymphocytes of the shark with differential response to phytohemagglutinin and concanavalin A. *Cellular Immunology*, *10*(2), 287-293.

- Lopez, P., Casane, D., & Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, 19(1), 1-7.
- Lukoschek, V., Scott Keogh, J., & Avise, J. C. (2012). Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: a comparison of three approaches. *Systematic Biology*, 61(1), 22-43.
- Lund, R., & Grogan, E. D. (1997). Relationships of the Chimaeriformes and the basal radiation of the Chondrichthyes. *Reviews in Fish Biology and Fisheries*, 7(1), 65-123.
- Luzio, A., Santos, D., Fontainhas-Fernandes, A. A., Monteiro, S. M., & Coimbra, A. M. (2016). Effects of 17alpha-ethinylestradiol at different water temperatures on zebrafish sex differentiation and gonad development. *Aquatic Toxicology*, 174, 22-35.
- Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, 46(3), 523-536.
- Maddison, W. P., & Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, 55(1), 21-30.
- Maddison, W. P., & Maddison, D. R. (2017). Mesquite: a modular system for evolutionary analysis. Version 3.2. Retrieved from <http://mesquiteproject.org>
- Maddock, M. B., & Schwartz, F. J. (1996). Elasmobranch Cytogenetics: Methods and Sex Chromosomes. *Bulletin of Marine Science*, 58(1), 147-155.
- Mair, G. C., Scott, A. G., Penman, D. J., Skibinski, D. O. F., & Beardmore, J. A. (1991). Sex determination in the genus *Oreochromis* 2. Sex reversal, hybridisation, gynogenesis and triploid in *O. aureus* Steindachner. *Theoretical and Applied Genetics*, 82, 153-160.
- Maisey, J. G. (1984). Chondrichthyan phylogeny: a look at the evidence. *Journal of Vertebrate Paleontology*, 4(3), 359-371.
- Maisey, J. G. (1986). Heads and tails: a chordate phylogeny. *Cladistics*, 2(3), 201-256.
- Mallatt, J., & Winchell, C. J. (2007). Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. *Molecular Phylogenetics and Evolution*, 43(3), 1005-1022.
- Mank, J. E., Promislow, D. E. L., & Avise, J. C. (2006). Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society*, 87, 83-93.
- Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J.-J., Bernot, A., . . . Guiguen, Y. (2000). DMRT1 expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1493(1-2), 180-187.
- Marín, I., & Baker, B. S. (1998). The Evolutionary Dynamics of Sex Determination. *Science*, 281(5385), 1990.

- Marshall, C. R. (1990). The fossil record and estimating divergence times between lineages: maximum divergence times and the importance of reliable phylogenies. *Journal of Molecular Evolution*, 30(5), 400-408.
- Marshall, C. R. (2008). A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist*, 171(6), 726-742.
- Martinez, P., Vinas, A. M., Sanchez, L., Diaz, N., Ribas, L., & Piferrer, F. (2014). Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Frontiers in Genetics*, 5, 340.
- Matsubara, K., Tarui, H., Toriba, M., Yamada, K., Nishida-Umehara, C., Agata, K., & Matsuda, Y. (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(48), 18190-18195.
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., . . . Sakaizumi, M. (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417(559-563).
- Matsuda, M., Sato, T., Toyazaki, Y., Nagahama, Y., Hamaguchi, S., & Sakaizumi, M. (2003). *Oryzias curvinotus* has DMY, a gene that is required for male development in the medaka, *O. latipes*. *Zoological Science*, 20(2), 159-161.
- Meadows, J. R. S., & Lindblad-Toh, K. (2017). Dissecting evolution and disease using comparative vertebrate genomics. *Nature Reviews Genetics*.
- Mello, B., & Schrago, C. G. (2014). Assignment of calibration information to deeper phylogenetic nodes is more effective in obtaining precise and accurate divergence time estimates. *Evolutionary Bioinformatics Online*, 10, 79-85.
- Meyer, A. (1994). Shortcomings of the cytochrome b gene as a molecular marker. *Trends in Ecology and Evolution*, 9(8), 278-280.
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 6.
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010, 14 Nov. 2010). *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*. Paper presented at the Proceedings of the Gateway Computing Environments Workshop (GCE).
- Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2), 240-248.
- Modi, W. S., & Crews, D. (2005). Sex chromosomes and sex determination in reptiles. *Current Opinion in Genetics & Development*, 15(6), 660-665.

- Moore, E. C., & Roberts, R. B. (2013). Polygenic sex determination. *Current Biology*, 23(12), R510-R512.
- Moreno, C., Lazar, J., Jacob, H. J., & Kwitek, A. E. (2008). Comparative Genomics for Detecting Human Disease Genes. *Advances in Genetics*, 60, 655-697.
- Mulley, J. F., Zhong, Y.-F., & Holland, P. W. H. (2009). Comparative genomics of chondrichthyan Hoxa clusters. *BMC Evolutionary Biology*, 9(1), 218.
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., . . . Sakaizumi, M. (2012). Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics*, 191(1), 163-170.
- Nabhan, A. R., & Sarkar, I. N. (2012). The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Briefings in Bioinformatics*, 13(1), 122-134.
- Nabholz, B., Kunstner, A., Wang, R., Jarvis, E. D., & Ellegren, H. (2011). Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Molecular Biology and Evolution*, 28(8), 2197-2210.
- Nah, G. S. S., Lim, Z. W., Tay, B.-H., Osato, M., & Venkatesh, B. (2014). Runx family genes in a cartilaginous fish, the elephant shark (*Callorhynchus milii*). *PLoS One*, 9(4), e93816.
- Nakamura, M. (2010). The mechanism of sex determination in vertebrates—are sex steroids the key-factor? *Journal of Experimental Zoology. Part A, Ecological Genetics and Physiology*, 313(7), 381-398.
- Nakamura, M., Kobayashi, T., Chang, X.-T., & Nagahama, Y. (1998). Gonadal sex differentiation in teleost fish. *Journal of Experimental Zoology*, 281(5), 362-372.
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., . . . Schartl, M. (2002). A duplicated copy of dmrt1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), 11778-11783.
- Nanda, I., Shan, Z., Schartl, M., Burt, D. W., Koehler, M., Nothwang, H.-G., . . . Schmid, M. (1999). 300 million years of conserved synteny between chicken Z and human chromosome 9. *Nature Genetics*, 21(3), 258-259.
- Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J., & Roberts, I. N. (2000). Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International Journal of Systematic and Evolutionary Microbiology*, 50(5), 1931-1942.
- Navarro-Martin, L., Vinas, J., Ribas, L., Diaz, N., Gutierrez, A., Di Croce, L., & Piferrer, F. (2011). DNA methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLOS Genetics*, 7(12), e1002447.

- Naylor, G., Caira, J., Jensen, K., Rosana, K., Straube, N., & Lakner, C. (2012a). Elasmobranch phylogeny: A mitochondrial estimate based on 595 species. In J. Carrier, J. Musick, & M. Heithaus (Eds.), *Biology of sharks and their relatives, 2 edition*. Boca Raton, FL: CRC Press.
- Naylor, G. J. P., Caira, J., Jensen, K., Rosana, K., White, W., & Last, P. (2012b). A DNA sequence-based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology. *Bulletin of the American Museum of Natural History*, 367, 1-262.
- Naylor, G. J. P., Ryburn, J. A., Fedrigo, O., & López, A. (2005). Phylogenetic relationships among the major lineages of modern Elasmobranchs. In W. Hamlett & B. Amieson (Eds.), *Reproductive Biology and Phylogeny of Chondrichthyes*: University of Queensland Press.
- Near, T. J., Bolnick, D. I., Wainwright, P. C., & Yoder, A. (2005). Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution*, 59(8), 1768-1782.
- Nelson, J. S., Grande, T. C., & Wilson, M. V. H. (2016). *Fishes of the World* (5 ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Nelson, N. J., Cree, A., Thompson, M. B., Keall, S. N., & Daucherty, C. H. (2004). Temperature-dependent sex determination in tuatara. In N. Valenzuela & V. A. Lance (Eds.), *Temperature-Dependent Sex Determination in Vertebrates* (pp. 53-58). Washington: Smithsonian Institution.
- Nesov, L., & Averianov, A. E. (1996). Early Chimaeriformes of Russia, Ukraine, Kazakhstan and Middle Asia II. Description of new taxa. *Vestnik Sankt-Peterburgskogo Universiteta, Seriya 7*, 3(21), 2-10.
- Neyt, C., Jagla, K., Thisse, C., Thisse, B., Haines, L., & Currie, P. D. (2000). Evolutionary origins of vertebrate appendicular muscle. *Nature*, 408(6808), 82-86.
- Nickle, D. C., Heath, L., Jensen, M. A., Gilbert, P. B., Mullins, J. I., & Kosakovsky Pond, S. L. (2007). HIV-specific probabilistic models of protein evolution. *PLoS One*, 2(6), e503.
- Non, A. L., Kitchen, A., & Mulligan, C. J. (2007). Identification of the most informative regions of the mitochondrial genome for phylogenetic and coalescent analyses. *Molecular Phylogenetics and Evolution*, 44(3), 1164-1171.
- Ogata, M., Hasegawa, Y., Ohtani, H., Mineyama, M., & Miura, I. (2008). The ZZ/ZW sex-determining mechanism originated twice and independently during evolution of the frog, *Rana rugosa*. *Heredity*, 100(1), 92-99.
- Ogawa, A., Murata, K., & Mizuno, S. (1998). The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proceedings of the National Academy of Sciences*, 95(8), 4415-4418.
- Ohno, S. (1974). *Animal Cytogenetics*. Berlin: Gebruder Borntraeger

- Oldfield, R. G., McCrary, J., & McKaye, K. R. (2006). Habitat use, social behavior, and female and male size distributions of juvenile Midas cichlids, *Amphilophus cf. citrinellus*, in Lake Apoyo, Nicaragua. *Caribbean Journal of Science*, 42(2), 197-207.
- Ospina-Alvarez, N., & Piferrer, F. (2008). Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS One*, 3(7), e2837.
- Ota, K., Tateno, Y., & Gojobori, T. (2003). Highly differentiated and conserved sex chromosome in fish species (*Aulopus japonicus*: Teleostei, Aulopidae). *Gene*, 317, 187-193.
- Page, R. D. M., & Holmes, E. C. (1998). *Molecular Evolution: a phylogenetic approach*. Oxford: Blackwell Science.
- Pagel, M., & Meade, A. (2004). A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Systematic Biology*, 53(4), 571-581
- Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M. E., Oral, M., Taggart, J. B., . . . Migaud, H. (2013a). Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics*, 14, 566.
- Palaiokostas, C., Bekaert, M., Khan, M. G., Taggart, J. B., Gharbi, K., McAndrew, B. J., & Penman, D. J. (2013b). Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. *PLoS One*, 8(7), e68389.
- Pamilo, P., & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, 5(5), 568-583.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2), 289-290.
- Patterson, C. (1965). The phylogeny of chimaeroids. *Philosophical Transactions of the Royal Society of London*, 249(757), 101-219.
- Patwardhan, A. R., S.; Roy, A. (2014). Molecular markers in phylogenetic studies - a review. *Phylogenetics and Evolutionary Biology*, 2(2), 1000131.
- Peaston, A. E., & Whitelaw, E. (2006). Epigenetics and phenotypic variation in mammals. *Mammalian Genome*, 17(5), 365-374.
- Peichel, C. L., Ross, J. A., Matson, C. K., Dickson, M., Grimwood, J., Schmutz, J., . . . Kingsley, D. M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology*, 14(16), 1416-1424.
- Pelegri, F., & Schulte-Merker, S. (1999). A gynogenesis-based screen for maternal-effect genes in the zebrafish, *Danio rerio*. *Methods in Cell Biology*, 60, 1-20.

- Pereira, S. L. (2000). Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology*, 23, 745-752.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, 9(3), e1000602.
- Philippe, H., Chenuil, A., & Adoutte, A. (1994). Can the Cambrian explosion be inferred through molecular phylogeny? *Development*, 1994(Supplement), 15.
- Philippe, H., Delsuc, F., Brinkmann, H., & Lartillot, N. (2005a). Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics*, 36(1), 541-562.
- Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., & Delsuc, F. (2005b). Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol*, 5, 50.
- Phillips, M. J., Delsuc, F., & Penny, D. (2004). Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution*, 21(7), 1455-1458.
- Phillips, R. B., Amores, A., Morasch, M. R., Wilson, C., & Postlethwait, J. H. (2006). Assignment of zebrafish genetic linkage groups to chromosomes. *Cytogenetic and Genome Research*, 114(2), 155-162.
- Piferrer, F., Blázquez, M., Navarro, L., & González, A. (2005). Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *General and Comparative Endocrinology*, 142(1-2), 102-110.
- Pijnacker, L. P., & Ferwerda, M. A. (1995). Zebrafish chromosome banding. *Genome*, 38(5), 1052-1055.
- Pollock, D. D., & Bruno, W. J. (2000). Assessing an unknown evolutionary process: effect of increasing site-specific knowledge through taxon addition. *Molecular Biology and Evolution*, 17(12), 1854-1858.
- Pollock, D. D., Zwickl, D. J., McGuire, J. A., & Hillis, D. M. (2002). Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology*, 51(4), 664-671.
- Portnoy, D. S., Hollenbeck, C. M., Johnston, J. S., Casman, H. M., & Gold, J. R. (2014). Parthenogenesis in a whitetip reef shark *Triaenodon obesus* involves a reduction in ploidy. *Journal of Fish Biology*, 85(2), 502-508.
- Pradel, A., Langer, M., Maisey, J. G., Geffard-Kuriyama, D., Cloetens, P., Janvier, P., & Tafforeau, P. (2009). Skull and brain of a 300-million-year-old chimaeroid fish revealed by synchrotron holotomography. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), 5224-5228.

- Qiu, S., Bergero, R., Guirao-Rico, S., Campos, J. L., Cezard, T., Gharbi, K., & Charlesworth, D. (2016). RAD mapping reveals an evolving, polymorphic and fuzzy boundary of a plant pseudoautosomal region. *Molecular Ecology*, *25*(1), 414-430.
- Quinn, A. E., Sarre, S. D., Ezaz, T., Marshall Graves, J. A., & Georges, A. (2011). Evolutionary transitions between mechanisms of sex determination in vertebrates. *Biology Letters*, *7*(3), 443-448.
- Rannala, B., & Yang, Z. (2007). Inferring speciation times under an episodic molecular clock. *Systematic Biology*, *56*(3), 453-466.
- Ravi, V., Lam, K., Tay, B.-H., Tay, A., Brenner, S., & Venkatesh, B. (2009). Elephant shark (*Callorhynchus milii*) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proceedings of the National Academy of Sciences*, *106*(38), 16327-16332.
- Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J., & Zarkower, D. (1999). Expression of *dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Developmental Biology*, *215*(2), 208-220.
- Raymond, C. S., Shamu, C. E., Shen, M. M., Seifert, K. J., Hirsch, B., Hodgkin, J., & Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature*, *391*(6668), 691-695.
- Rees, J. A. N. (2005). Neoselachian shark and ray teeth from the Valanginian, lower Cretaceous, of Wawal, central Poland. *Palaeontology*, *48*(2), 209-221.
- Rens, W., Grützner, F., O'Brien, P. C. M., Fairclough, H., Graves, J. A. M., & Ferguson-Smith, M. A. (2004). Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(46), 16257-16261.
- Renz, A. J., Meyer, A., & Kuraku, S. (2013). Revealing less derived nature of cartilaginous fish genomes with their evolutionary time scale inferred with nuclear genes. *PLoS One*, *8*(6), e66400.
- Rey, R., Lukas-Croisier, C., Lasala, C., & Bedecarrás, P. (2003). AMH/MIS: what we know already about the gene, the protein and its regulation. *Molecular and Cellular Endocrinology*, *211*(1-2), 21-31.
- Rhen, T., & Schroeder, A. (2010). Molecular mechanisms of sex determination in reptiles. *Sexual Development*, *4*(1-2), 16-28.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S. D., . . . Birol, I. (2010). De novo assembly and analysis of RNA-seq data. *Nature Methods*, *7*(11), 909-912.
- Robinson, D. M., Jones, D. T., Kishino, H., Goldman, N., & Thorne, J. L. (2003). Protein evolution with dependence among codons due to tertiary structure. *Molecular Biology and Evolution*, *20*(10), 1692-1704.

- Robinson, D. P., Baverstock, W., Al-Jaru, A., Hyland, K., & Khazanehdari, K. A. (2011). Annually recurring parthenogenesis in a zebra shark *Stegostoma fasciatum*. *Journal of Fish Biology*, 79(5), 1376-1382.
- Robinson-Rechavi, M., Boussau, B., & Laudet, V. (2004). Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. *Molecular Biology and Evolution*, 21(3), 580-586.
- Rocco, L. (2013). Sex-related genomic sequences in cartilaginous fish: an overview. *Cytogenetic and Genome Research*, 141(2-3), 169-176.
- Rocco, L., Bencivenga, S., Archimandritis, A., & Stingo, V. (2009). Molecular characterization and chromosomal localization of spermatogenesis related sequences in *Torpedo torpedo* (Chondrichthyes, Torpediniformes). *Marine Genomics*, 2(2), 99-102.
- Rodrigue, N., Philippe, H., & Lartillot, N. (2006). Assessing site-interdependent phylogenetic models of sequence evolution. *Molecular Biology and Evolution*, 23(9), 1762-1775.
- Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B. F., Philippe, H., & Anderson, F. (2007). Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic Biology*, 56(3), 389-399.
- Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, 22(5), 939-946.
- Rokas, A., & Abbot, P. (2009). Harnessing genomics for evolutionary insights. *Trends in Ecology and Evolution*, 24(4), 192-200.
- Rokas, A., & Carroll, S. B. (2005). More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution*, 22(5), 1337-1344.
- Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 425(6960), 798-804.
- Romer, U., & Beisenherz, W. (1996). Environmental determination of sex in *Apistogramma* (Cichlidae) and two other freshwater fishes (Teleostei). *Journal of Fish Biology*, 48, 714-725.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., . . . Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539-542.
- Rosenberg, M. S., & Kumar, S. (2003). Taxon sampling, bioinformatics, and phylogenomics. *Systematic Biology*, 52(1), 119-124.
- Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., & Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLOS Genetics*, 5(2), e1000391.

- Rubin, B. E. R., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. *PLoS One*, 7(4), e33394.
- Rubin, D. A. (1985). Effect of pH on sex ratio in cichlids and a poeciliid (Teleostei). *Copeia*, 1985(1), 233-235.
- Rubinoff, D., & Holland, B. S. (2005). Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, 54(6), 952-961.
- Rubinoff, D., & Sperling, F. A. H. (2004). Mitochondrial DNA sequence, morphology and ecology yield contrasting conservation implications for two threatened buckmoths (*Hemileuca*: Saturniidae). *Biological Conservation*, 118(3), 341-351.
- Saitou N. F.; Nei, M. (1986). The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence. *Journal of Molecular Evolution*, 24(1-2), 189-204.
- Saladin, B., Leslie, A. B., Wuest, R. O., Litsios, G., Conti, E., Salamin, N., & Zimmermann, N. E. (2017). Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology*, 17(1), 95.
- Sanders, K. L., & Lee, M. S. (2007). Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. *Biology Letters*, 3(3), 275-279.
- Sandra, G.-E., & Norma, M.-M. (2009). Sexual determination and differentiation in teleost fish. *Reviews in Fish Biology and Fisheries*, 20(1), 101-121.
- Sansom, I. J., Smith, M. M., & Smith, M. P. (1996). Scales of thelodont and shark-like fishes from the Ordovician of Colorado. *Nature*, 379, 628.
- Satta, Y., Klein, J., & Takahata, N. (2000). DNA archives and our nearest relative: the trichotomy problem revisited. *Molecular Phylogenetics and Evolution*, 14(2), 259-275.
- Sawatari, E., Shikina, S., Takeuchi, T., & Yoshizaki, G. (2007). A novel transforming growth factor-beta superfamily member expressed in gonadal somatic cells enhances primordial germ cell and spermatogonial proliferation in rainbow trout (*Oncorhynchus mykiss*). *Developmental Biology*, 301(1), 266-275.
- Schaeffer, B., & Williams, M. (1977). Relationships of Fossil and Living Elasmobranchs. *American Zoologist*, 17(2), 293-302.
- Schartl, M. (2004). Sex chromosome evolution in non-mammalian vertebrates. *Current Opinion in Genetics & Development*, 14(6), 634-641.
- Schmid, M., Steinlein, C., Bogart, J. P., Feichtinger, W., León, P., La Marca, E., . . . Hedges, S. B. (2010). The chromosomes of terraranan frogs. Insights into vertebrate cytogenetics. *Cytogenetic and Genome Research*, 130-131(1-8), 1-14.

- Schulte, J. A. (2013). Undersampling taxa will underestimate molecular divergence dates: An example from the South American lizard clade Liolaemini. *International Journal of Evolutionary Biology*, 2013, 1-12.
- Schultheis, C., Bohne, A., Scharl, M., Volff, J. N., & Galiana-Arnoux, D. (2009). Sex determination diversity and sex chromosome evolution in poeciliid fish. *Sexual Development*, 3(2-3), 68-77.
- Schultheis, C., Zhou, Q., Froschauer, A., Nanda, I., Selz, Y., Schmidt, C., . . . Volff, J. N. (2006). Molecular analysis of the sex-determining region of the platyfish *Xiphophorus maculatus*. *Zebrafish*, 3(3), 299-309.
- Schwartz, F. J., & Maddock, M. B. (2002). Cytogenetics of the elasmobranchs: genome evolution and phylogenetic implications. *Marine and Freshwater Research*, 53(2), 491-502.
- Schwartz, R. J., & Maddock, M. B. (1986). Comparisons of karyotypes and cellular DNA contents within and between major lines of elasmobranchs. In T. Uyeno, R. Arai, T. Taniuchi, & K. Matsuura (Eds.), *Indo Pacific Fish Biology* (pp. 148-157). Tokyo: Ichthyological Society of Japan.
- Schwartz, R. S., & Mueller, R. L. (2010). Branch length estimation and divergence dating: estimates of error in Bayesian and maximum likelihood frameworks. *BMC Evolutionary Biology*, 10, 5.
- Ser, J. R., Roberts, R. B., & Kocher, T. D. (2010). Multiple interacting loci control sex determination in lake Malawi cichlid fish. *Evolution*, 64(2), 486-501.
- Shang, E. H. H., Yu, R. M. K., & Wu, R. S. S. (2006). Hypoxia affects sex differentiation and development, leading to a male-dominated population in zebrafish (*Danio rerio*). *Environmental Science & Technology*, 40(9), 3118-3122.
- Shao, C.-W., Chen, S.-L., Scheuring, C. F., Xu, J.-Y., Sha, Z.-X., Dong, X.-L., & Zhang, H.-B. (2010). Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. *Marine Biotechnology*, 12(5), 558-568.
- Sheffield, N. C. (2013). The interaction between base compositional heterogeneity and among-site rate variation in models of molecular evolution. *ISRN Evolutionary Biology*, 2013, 8.
- Sheffield, N. C., Song, H., Cameron, S. L., & Whiting, M. F. (2009). Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Systematic Biology*, 58(4), 381-394.
- Shen, M. M., & Hodgkin, J. (1988). *mab-3*, a gene required for sex-specific yolk protein expression and a male-specific lineage in *C. elegans*. *Cell*, 54(7), 1019-1031.
- Shibata, K., Takase, M., & Nakamura, M. (2002). The *Dmrt1* expression in sex-reversed gonads of amphibians. *General and Comparative Endocrinology*, 127(3), 232-241.

- Siegfried, K. R. (2010). In search of determinants: gene expression during gonadal sex differentiation. *Journal of Fish Biology*, 76(8), 1879-1902.
- Silva, W. A., Jr., Bonatto, S. L., Holanda, A. J., Ribeiro-Dos-Santos, A. K., Paixao, B. M., Goldman, G. H., . . . Zago, M. A. (2002). Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *American Journal of Human Genetics*, 71(1), 187-192.
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research*, 19(6), 1117-1123.
- Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., . . . Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346(6281), 240-244.
- Siverson, M., & Cappetta, H. (2001). A skate in the lowermost Maastrichtian of southern Sweden. *Palaeontology*, 44(3), 431-445.
- Smit, S., Widmann, J., & Knight, R. (2007). Evolutionary rates vary among rRNA structural elements. *Nucleic Acids Res*, 35(10), 3339-3354.
- Smith, C. A. (2010). Sex determination in birds: a review. *Emu*, 110(4), 364-377.
- Smith, C. A., McClive, P. J., Western, P. S., Reed, K. J., & Sinclair, A. H. (1999). Conservation of a sex-determining gene. *Nature*, 402(601-602).
- Smith, C. A., Roeszler, K. N., Hudson, Q. J., & Sinclair, A. H. (2007). Avian sex determination: what, when and where? *Cytogenetic and Genome Research*, 117(1-4), 165-173.
- Smith, C. A., Roeszler, K. N., Ohnesorg, T., Cummins, D. M., Farlie, P. G., Doran, T. J., & Sinclair, A. H. (2009). The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature*, 461(7261), 267-271.
- Smith, C. A., & Sinclair, A. H. (2004). Sex determination: insights from the chicken. *Bioessays*, 26(2), 120-132.
- Soltis, D. E., & Soltis, P. S. (2003). The Role of Phylogenetics in Comparative. *Plant Physiology*, 132(4), 1790-1800.
- Som, A. (2015). Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics*, 16(3), 536-548.
- Song, S., Pursell, Z. F., Copeland, W. C., Longley, M. J., Kunkel, T. A., & Mathews, C. K. (2005). DNA precursor asymmetries in mammalian tissue mitochondria and possible contribution to mutagenesis through reduced replication fidelity. *Proceedings of the National Academy of Sciences of the United States of America*, 102(14), 4990-4995.

- Spencer, M., Susko, E., & Roger, A. J. (2005). Likelihood, parsimony, and heterogeneous evolution. *Molecular Biology and Evolution*, 22(5), 1161-1164.
- Springer, M. S. (1995). Molecular clocks and the incompleteness of the fossil record. *Journal of Molecular Evolution*, 41(5), 531-538.
- Stahl, B. J. (1999). Mesozoic holocephalans. In G. Arratia & H. P. Schultze (Eds.), *Mesozoic Fishes. Systematics and the Fossil Record*. (Vol. 2, pp. 9-19). Munich: Verlag Dr. Friedrich Pfeil.
- Stahl, B. J., & Chatterjee, S. (1999). A Late Cretaceous chimaerid (Chondrichthyes, Holocephali) from Seymour Island, Antarctica. *Palaeontology*, 42(6), 979-989.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Stingo, V., & Rocco, L. (2001). Selachian cytogenetics: a review. *Genetica*, 111(1), 329-347.
- Stingo, V., Rocco, L., Odierna, G., & Bellitti, M. (1995). NOR and heterochromatin analysis in two cartilaginous fishes by C-, Ag- and RE (restriction endonuclease)-banding. *Cytogenetics and Cell Genetics*, 71(3), 228-234.
- Straube, N., Lampert, K. P., Geiger, M. F., Weiss, J. D., & Kirchhauser, J. X. (2016). First record of second-generation facultative parthenogenesis in a vertebrate species, the whitespotted bambooshark *Chiloscyllium plagiosum*. *Journal of Fish Biology*, 88(2), 668-675.
- Sullivan, J., & Swofford, D. L. (2001). Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Systematic Biology*, 50(5), 723-729.
- Sullivan, J., Swofford, D. L., & Naylor, G. J. P. (1999). The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Molecular Biology and Evolution*, 16(10), 1347-1347.
- Sumner, A. T., Evans, H. J., & Buckland, R. A. (1971). New technique for distinguishing between human chromosomes. *Nature New Biology*, 232(27), 31-32.
- Swofford, D. (2002). *PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4*. Sunderland, MA: Sinauer Associates.
- Swofford, D. L., Olsen, G. J., Waddell, P. J., & Hillis, D. M. (1996). Phylogenetic Inference. In D. M. Hillis, D. Moritz, & B. K. Mable (Eds.), *Molecular Systematics* (pp. 407-514). Sunderland, MA: Sinauer Associates.
- Takehana, Y., Hamaguchi, S., & Sakaizumi, M. (2008). Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Research*, 16(5), 801-811.

- Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., & Kumar, S. (2012). Estimating divergence times in large molecular phylogenies. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19333-19338.
- Tanaka, K., Takehana, Y., Naruse, K., Hamaguchi, S., & Sakaizumi, M. (2007). Evidence for different origins of sex chromosomes in closely related *Oryzias* fishes: substitution of the master sex-determining gene. *Genetics*, *177*(4), 2075-2081.
- Tanaka, M., Munsterberg, A., Anderson, W. G., Prescott, A. R., Hazon, N., & Tickle, C. (2002). Fin development in a cartilaginous fish and the origin of vertebrate limbs. *Nature*, *416*(6880), 527-531.
- Teixeira, J., Maheswaran, S., & Donahoe, P. K. (2001). Müllerian inhibiting substance: An instructive developmental hormone with diagnostic and possible therapeutic applications. *Endocrine Reviews*, *22*(5), 657-674.
- Thorgaard, G. (1977). Heteromorphic sex chromosomes in male rainbow trout. *Science*, *196*(4292), 900-902.
- Thorgaard, G. H., Allendorf, F. W., & Knudsen, K. L. (1983). Gene-centromere mapping in rainbow trout: high interference over long map distances. *Genetics*, *103*(4), 771-783.
- Tong, S.-K., Hsu, H.-J., & Chung, B.-c. (2010). Zebrafish monosex population reveals female dominance in sex determination and earliest events of gonad differentiation. *Developmental Biology*, *344*(2), 849-856.
- Traut, W., & Winking, H. (2001). Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Research*, *9*(8), 659-672.
- Triantaphyllopoulos, K. A., Ikononopoulos, I., & Bannister, A. J. (2016). Epigenetics and inheritance of phenotype variation in livestock. *Epigenetics & Chromatin*, *9*(1), 31.
- Trukhina, A. V., Lukina, N. A., Wackerow-Kouzova, N. D., & Smirnov, A. F. (2013). The Variety of Vertebrate Mechanisms of Sex Determination. *BioMed Research International*, *2013*, 587460.
- Tuppen, H. A. L., Blakely, E. L., Turnbull, D. M., & Taylor, R. W. (2010). Mitochondrial DNA mutations and human disease. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1797*(2), 113-128.
- Uchida, D., Yamashita, M., Kitano, T., & Iguchi, T. (2004). An aromatase inhibitor or high water temperature induce oocyte apoptosis and depletion of P450 aromatase activity in the gonads of genetic female zebrafish during sex-reversal. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *137*(1), 11-20.
- Uller, T., & Helanterä, H. (2011). From the origin of sex-determining factors to the evolution of sex-determining systems. *The Quarterly Review of Biology*, *86*(3), 163-180.

- Underwood, C. J. (2002). Sharks, rays and a chimaeroid from the Kimmeridgian (Late Jurassic) of Ringstead, Southern England. *Palaeontology*, 45(2), 297-325.
- Underwood, C. J. (2006). Diversification of the Neoselachii (Chondrichthyes) during the Jurassic and Cretaceous. *Paleobiology*, 32(2), 215-235.
- Underwood, C. J., Ward, D. J., King, C., Antar, S. M., Zalmout, I. S., & Gingerich, P. D. (2011). Shark and ray faunas in the Middle and Late Eocene of the Fayum Area, Egypt. *Proceedings of the Geologists' Association*, 122(1), 47-66.
- Uwa, H., & Ojima, Y. (1981). Detailed and Banding Karyotype Analyses of the Medaka, *Oryzias latipes* in Cultured Cells. *Proceedings of the Japan Academy, Series B*, 57(2), 39-43.
- Valenzuela, N. (2004). Introduction. In N. Valenzuela & V. A. Lance (Eds.), *Temperature-Dependent Sex Determination in Vertebrates* (pp. 1-4). Washington: Smithsonian Institution.
- Van Den Bussche, R. A., Baker, R. J., Huelsenbeck, J. P., & Hillis, D. M. (1998). Base compositional bias and phylogenetic analyses: A test of the "Flying DNA" hypothesis. *Molecular Phylogenetics and Evolution*, 10(3), 408-416.
- Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., . . . Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5(3), 247-252.
- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., & Chatain, B. (2007). A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics*, 176(2), 1049-1057.
- Velez-Zuazo, X., & Agnarsson, I. (2011). Shark tales: a molecular species-level phylogeny of sharks (Selachimorpha, Chondrichthyes). *Molecular Phylogenetics and Evolution*, 58(2), 207-217.
- Venkatesh, B., Erdmann, M. V., & Brenner, S. (2001). Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20), 11382-11387.
- Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., . . . Brenner, S. (2006). Ancient noncoding elements conserved in the human genome. *Science*, 314(5807), 1892.
- Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., Lee, A. P., Johnson, J., . . . Brenner, S. (2007). Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biology*, 5(4), e101.
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., . . . Warren, W. C. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature*, 505(7482), 174-179.

- Venkatesh, B., Tay, A., Dandona, N., Patil, J. G., & Brenner, S. (2005). A compact cartilaginous fish model genome. *Current Biology*, *15*(3), R82-R83.
- Viets, B. E., Ewert, M. A., Talent, L. G., & Nelson, C. E. (1994). Sex-determining mechanisms in squamate reptiles. *Journal of Experimental Zoology*, *270*(1), 45-56.
- Volff, J.-N., & Schartl, M. (2001). Variability of genetic sex determination in poeciliid fishes. *Genetica*, *111*(1), 101-110.
- Volff, J. N., Nanda, I., Schmid, M., & Schartl, M. (2007). Governing sex determination in fish: regulatory putches and ephemeral dictators. *Sexual Development*, *1*(2), 85-99.
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., . . . Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, *22*(3), 787-798.
- Wang, J., Lee, A. P., Kodzius, R., Brenner, S., & Venkatesh, B. (2009). Large number of ultraconserved elements were already present in the jawed vertebrate ancestor. *Molecular Biology and Evolution*, *26*(3), 487-490.
- Wang, L., & Tsai, C. (2000). Effects of temperature on the deformity and sex differentiation of tilapia, *Oreochromis mossambicus*. *Journal of Experimental Zoology*, *286*(5), 534-537.
- Ward, D., & Duffin, C. J. (1989). Mesozoic chimaeroids. 1. A new chimaeroid from the early Jurassic of Gloucestershire, England. *Mesozoic Research*, *2*, 45-51.
- Ward, R. D., Holmes, B. H., White, W. T., & Last, P. R. (2008). DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research*, *59*, 57-71.
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1847-1857.
- White, W. T., Corrigan, S., Yang, L., Henderson, A. C., Bazinet, A. L., Swofford, D. L., & Naylor, G. J. P. (2017). Phylogeny of the manta and devilrays (Chondrichthyes: Mobulidae), with an updated taxonomic arrangement for the family. *Zoological Journal of the Linnean Society*, 1-26.
- Widmann, J., Harris, J. K., Lozupone, C., Wolfson, A., & Knight, R. (2010). Stable tRNA-based phylogenies using only 76 nucleotides. *RNA*, *16*(8), 1469-1477.
- Wiedmann, R. T., Smith, T. P., & Nonneman, D. J. (2008). SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC Genetics*, *9*, 81.
- Wilhelm, D., Palmer, S., & Koopman, P. (2007). Sex determination and gonadal development in mammals. *Physiological Reviews*, *87*(1), 1-28.

- Willing, E. M., Hoffmann, M., Klein, J. D., Weigel, D., & Dreyer, C. (2011). Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics*, 27(16), 2187-2193.
- Wilson, C. A., High, S. K., McCluskey, B. M., Amores, A., Yan, Y. L., Titus, T. A., . . . Postlethwait, J. H. (2014). Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. *Genetics*, 198(3), 1291-1308.
- Winchell, C. J., Martin, A. P., & Mallatt, J. (2004). Phylogeny of elasmobranchs based on LSU and SSU ribosomal RNA genes. *Molecular Phylogenetics and Evolution*, 31(1), 214-224.
- Wray, G. A., Levinton, J. S., & Shapiro, L. H. (1996). Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science*, 274(5287), 568.
- Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1), 1-7.
- Yamamoto, T. (1969). Sex differentiation. In W. S. Hoar & D. J. Randall (Eds.), *Fish Physiology* 3 (pp. 117-175). New York: Academic Press.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-314.
- Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology and Evolution*, 11(9), 367-372.
- Yang, Z., Nielsen, R., & Hasegawa, M. (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution*, 15(12), 1600-1611.
- Yang, Z., & Rannala, B. (2006). Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution*, 23(1), 212-226.
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., . . . Guiguen, Y. (2012). An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology*, 22(15), 1423-1428.
- Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., & Guiguen, Y. (2013). The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications*, 6(3), 486-496.
- Yao, H. H. C., & Capel, B. (2005). Temperature, genes, and sex: a comparative view of sex determination in *Trachemys scripta* and *Mus musculus*. *The Journal of Biochemistry*, 138(1), 5-12.

- Yi, W., Ross, J. M., & Zarkower, D. (2000). Mab-3 is a direct tra-1 target gene regulating diverse aspects of *C. elegans* male sexual development and behavior. *Development*, *127*(20), 4469.
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., . . . Ito, M. (2008). A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(7), 2469-2474.
- Yu, W. P., Rajasegaran, V., Yew, K., Loh, W. L., Tay, B. H., Amemiya, C. T., . . . Venkatesh, B. (2008). Elephant shark sequence reveals unique insights into the evolutionary history of vertebrate genes: A comparative analysis of the protocadherin cluster. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 3819-3824.
- Zhu, L., Wilken, J., Phillips, N. B., Narendra, U., Chan, G., Stratton, S. M., . . . Weiss, M. A. (2000). Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes & Development*, *14*(14), 1750-1764.
- Zhuang, Z. M., Wu, D., Zhang, S. C., Pang, Q. X., Wang, C. L., & Wan, R. J. (2006). G-banding patterns of the chromosomes of tonguefish *Cynoglossus semilaevis* Günther, 1873. *Journal of Applied Ichthyology*, *22*(5), 437-440.
- Zwickl, D. J., & Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology*, *51*(4), 588-598.