ASSESSMENT OF GENOMIC VARIATION USING ddRAD SEQUENCING DATA AND ITS IMPLICATIONS TOWARDS PHYLOGENOMIC RELATIONSHIPS AND RATES OF MOLECULAR EVOLUTION OF TUNAS (*Thunnus Spp*.)

A Thesis

by

GIOVANNI MADRIGAL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jaime R. Alvarado-Bremer
Committee Members,	Jessica Labonté
	Marianna Mateos
Intercollegiate Faculty Chair,	Anja Schulze

December 2020

Major Subject: Marine Biology

ABSTRACT

Tunas are an ecologically and economically important group of fishes, and have received attention by numerous researchers for decades. Despite decades of research, their phylogenomic relationships and evolutionary history remains ambiguous and controversial. Increasing fishing pressures and climate change have warranted a better understanding of these fishes, as understanding their past is imperative towards understanding their future. This study focused on: 1) providing a bioinformatic framework for dealing with missing sites using double digest restriction-site associated sequencing (ddRAD-Seq) data in phylogenomics using tunas as an example, 2) the construction of a phylogeny using diagnostically useful single nucleotide polymorphisms (SNPs) for *Thunnus spp*., 3) the revaluation of Bigeye Tuna's (*T. obesus*) phylogenomic position in relation to temperate and tropical tunas, 4) offering an average rate of molecular evolution for members of *Thunnus*, and 5) the estimation of a timeline of speciation events within *Thunnus*. Applying both concatenation and species tree methods using untrimmed and trimmed data provided the support towards concatenation and the inclusion of missing sites when using ddRAD-Seq to construct a phylogenetic tree. Despite the potential pitfalls with ddRAD-Seq, this study was able to capture informative phylogenetic markers that successfully separate individual species, placing Bigeye Tuna within the tropical Yellowfin Group. Constructing a timetree for *Thunnus* resulted in congruent results, with the timetree indicating the lineages for the temperate and tropical tunas are relatively old (> 50 million years). A relatively high rate of molecular evolution is associated with the Bigeye Tuna clade when compared to the other tunas in this study, which are estimated to have a relatively slow rate of molecular evolution when compared to several other taxonomic groups (e.g., mammals). Historic ocean cooling may

have driven Bigeye Tuna to diverge from the Yellowfin Tuna in search of resources in cooler and deeper waters, with annotated sequences in line with morphological differences between temperate and tropical tunas. Overall, this study provides evidence towards Bigeye Tuna's placement in the *Neothunnus*, and suggests this species may have undergone a relatively rapid speciation event leading to convergent evolution in *Thunnus*.

DEDICATION

To my mother, who has always been there for me from the very start. You left us all too soon, and although you are no longer here physically, I know you will always be with me spiritually. Thank you for all the love you have provided me and for ensuring that I have a bright future ahead of me. Your actions have always spoken louder than your words, and you showed me what it truly means to work towards a better future. I am sure that nothing I have stated here is new to you, so all I have left to do is to promise you that I will continue on this path you have set me on and do whatever it takes to make you proud. Love, your little boy.

ACKNOWLEDGEMENTS

I cannot underscore the amount of gratitude I have for my committee chair, Dr. Jaime R. Alvarado Bremer. Thank you for taking a chance on me and for the journey I could not have ever imagined. I still remember the day I graduated from undergrad, where you declared to my father that you will raise me up to be a fine researcher. Thank you for keeping your word.

Thank you to Dr. Jessica Labonté for showing me how vast the field of bioinformatics is and sparking a passion in me that I never knew I had. Thank you for also taking me on as an undergraduate researcher and providing me the confidence I needed to pursue this field. I'd also like to thank Dr. Mateos for the rapid support she has given me, and the quick, but much needed lessons in phylogenomics.

I'd like to give a big thank you to Janelle and Joel Espinoza for accepting me for who I really am and making sure that I live a good life. You guys are like a second family to me and I cannot stress that enough. To my lab mates Roselyn Aguila and Erica Atkins for their patience and support during my time in the lab. It was the tough times that made our friendships stronger and I learned a lot from you guys. Thank you Janelle, Roselyn, and Erica for making Dr. Alvarado's lab the best lab possible.

Thank you Yui Matsumoto for being a great friend during this project and for all the knowledge you have provided me. This project would not have been completed without your help. Further thanks need to be given to my friends Jerrimesia Hamilton and Ashley Whitehead for allowing me to sprinkle some of my passion onto your lives. Lastly, thank you to my loving parents and brothers. Your support cannot be described in words and I truly am fortunate to be given a family like you all.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee consisting of Dr. Jaime R. Alvarado Bremer as advisor and chair, and Dr. Jessica Labonté of the Marine Biology Department, and Dr. Mariana Mateos of the Wildlife and Fisheries Sciences Department.

Tuna samples used in this investigated were provided by Dr. Bruce Collette from the Smithsonian National Museum of Natural History, Dr. John Mark Dean from the University of South Carolina, and Dr. William B. Driggers from the University of New England. DNA extractions were performed by Dr. Janelle Espinoza from Victoria College, TX, Dr. Jaime R. Alvarado and Roselyn Aguila from the Marine Biology department at Texas A&M University at Galveston. Advice for Gene Ontology analyses was provided by Yui Matsumoto in the Miglietta Lab at Texas A&M University at Galveston.

All other work conducted for the thesis was completed by the student, under the advisement of Dr. Jaime R. Alvarado-Bremer.

Funding Sources

Graduate study was funded in part by the Office of Graduate Studies and Research at Texas A&M University at Galveston, and this research would not have been possible without the support from NOAA SK Grant NA15NMF4270345 awarded to my advisor Dr. Jaime R. Alvarado-Bremer.

NOMENCLATURE

4d	Fourfold Degenerative Sites
BLS	Body Length per Second
COI	Cytochrome Oxidase I
CR	Control Region
cty-b	Cytochrome b
ddRAD	Double Digest Restriction-site Associated DNA
DNA	Deoxyribonucleic Acid
EtBr	Ethidium Bromide
ESS	Effective Sample Size
GO	Gene Ontology
HKY	Hasegawa-Kishono-Yano
HPD	Highest Posterior Density
INDEL	Insertion-Deletion
ILS	Incomplete Lineage Sorting
Ma	Mega-annum
MCMC	Markov Chain Monte Carlo
Mrca	Most Recent Common Ancestor
Mt	Mitochondrial
Mya	Million years ago
N_e	Effective Population Size
NGS	Next-Generation Sequencing

NJ	Neighbor-Joining
RAD	Restriction-site Associated DNA
RefSeq	Reference Sequence
RM	Red Muscle
SAM	Sequence Alignment Map
Seq	Sequencing
SNP	Single Nucleotide Polymorphism
SST	Sea Surface Temperature
TA	Tris-Acetate

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

CONTRIBUTORS AND FUNDING SOURCES vi

NOMENCLATURE vii

TABLE OF CONTENTS ix

LIST OF FIGURES xi

LIST OF TABLES xiii

CHAPTER I INTRODUCTION 1

- Speciation in Marine Environments 1
- Literature Review of Tunas 2
- Tuna Phylogenies in the Genomics Era 11
- Objectives 11

CHAPTER II THE EFFECTS OF MISSING DATA TREATMENT IN A REFERENCED-BASED PHYLOGENOMIC ANALYSIS: RAD-TAGS SPECIES TREE OR CONCATENATION 13

- Introduction 13
- Materials and Methods 18
- Results 22
- Discussion 29

CHAPTER III ESTIMATING THE MOLECULAR EVOLUTIONARY RATE WITHIN TUNAS 32

- Introduction 32
- Materials and Methods 35
- Results 42

Discussion 55

CHAPTER IV CONCLUSIONS 63

Phylogeny Construction of *Thunnus* Using ddRAD-Seq 64
Phylogenetic Placement of Bigeye Tuna 65
Estimating the Average Molecular Evolutionary Rate in *Thunnus* 67
Treatment of Missing Data Using ddRAD-Seq in a Phylogenomics Framework 68

REFERENCES 70

APPENDIX A 107

APPENDIX B 108

LIST OF FIGURES

Page

Figure 1. The number of reads across all sequenced samples before and after assessing the quality of the data by implementing process_radtags.pl 23

Figure 2. Unrooted Neighbor-Joining (NJ) trees for all 208 loci retained after filtration of low quality and unshared loci. 25

Figure 3. Trees A-D represent species trees generated through concatenation (A & B) or partitioning (C & D) of either the untrimmed (A & C) or trimmed (B & D) dataset containing genera *Thunnus*, *Euthynnus*, and *Katsuwonus* 27

Figure 4. Trees produced across four different treatments using only the *Thunnus* samples... 28

Figure 5. Bar plot of both the number of starting raw reads for all 10 samples, along with the number of reads retained per sample after quality checking using the process_radtags perl script. 43

Figure 6. The distribution of the amount of missing data within the 531 loci across all 10 samples. 44

Figure 7. The distribution of the amount of missing data within the remaining 424 loci after removing relatively poorly characterized loci across tuna samples 44

Figure 8. Unrooted phylogenetic trees of *Thunnus* samples using the loci assembled using the ipyrad pipeline for the unfiltered (531 loci) and the filtered assembly (424 loci) 46

Figure 9. Maximum likelihood tree of *Thunnus spp*. using 424 orthologs 47

Figure 10. Summarized Bayesian tree of *Thunnus spp*. using 424 orthologs 47

Figure 11. Timetree representing the *Thunnus* samples in this study 49

Figure 12. Timetree with blue bars representing the 95% HPD intervals for the age of each node for the genus *Thunnus* 49

Figure 13. Tunas are estimated to have the lowest nuclear to mitochondrial mutation rate per Ma ratio when compared to estimates used in Allio et al. (2017) 51

Figure 14. Annotations from the biological process ontology 53

Figure 15. Annotations associated with the molecular function ontology..... 54

Figure 16. Bar graph of the annotations found for the cellular component ontology..... 55

LIST OF TABLES

Page

Table 1 Samples used for analyses, species of the sample, regions from which samples were collected, year of collection, and year sequenced. 19

Table 2 *Thunnus* samples used for analyses, species of the sample, regions from which samples were collected, year of collection, and year sequenced. 36

Table 3 Molecular evolutionary rates across several studies and taxonomic groups 50

CHAPTER I

INTRODUCTION

Speciation in Marine Environments

On a global scale, marine ecosystems are heterogeneous and dynamic, generating one of the planet's greatest source of biodiversity (Heywood & Watson 1995). Mora et al. (2011) estimated approximately 2.2 million species inhabiting marine ecosystems; however, the authors acknowledge that their estimates are conservative and that more discoveries are to be made (Crist et al. 2009). Many of these discoveries have been supported by molecular data (i.e., DNA-based), although there is no clear definition on what distinguishes a species (Saikia et al. 2008, Shapiro et al. 2016). Most surveys have relied on sequences of mitochondrial (mt) DNA genes (Hillis et al. 1996), with certain segments displaying rates of evolution sufficient to distinguish among closely related species, such as cytochrome oxidase I (COI) or cytochrome b (cyt-b), that mutate faster than other regions of the mtDNA genome (e.g., 12S rRNA), but slower than other segments (e.g., control region)(reviewed in Patwardhan et al. 2014). These relative rates, however, cannot be generalized to all species. For example, the cyt-b gene, which has been shown to have a greater number of substitutions than the control region (CR) between closely related species (Tang et al. 2006) and is often used in the field of phylogenetics, has also shown greater variability between conspecifics than heterospecifics (Chow & Kishino 1995, Ciccarese et al. 2019). This incongruity may be explained be introgression and/or incomplete lineage sorting (ILS) during rapid speciation events (Pagès et al. 2013, Meyer et al. 2016, Alexander et al. 2017). Additionally, genes involved in species divergence may also contribute to intraspecific local adaption (Wachowiak et al. 2018), inferring that not all loci diverge equally between species. Stabilizing selection may mitigate

divergence in loci that contribute to a favorable phenotypic trait shared between species (Hebert et al. 2004, Smith et al. 2011), while directional selection on other loci further speciation (Palumbi 2009, Pauers & Mckinnon 2012).

The investigation of marine organisms has provided numerous benefits to society in the form of medicine (Lloret 2010), maintenance of sustainable fisheries (Dann et al. 2013), restoration projects (van Oppen et al. 2015), and forensics (Saiki et al. 1988). Particularly, the understanding of molecular evolution has become increasingly important (Nesse 2008, Lea et al. 2017), especially in fields such as cancer (Tollis et al. 2017) and viral diseases (Wang et al. 2020a). Thus, an increased understanding of molecular evolution in the marine environment would provide novel information towards areas such as conservation efforts (Grant & Bowen 1998), medical practices (Taylor 2008), and resource sustainability (Heino & Godø 2002). The use of bioinformatic tools can aid in our ability to understand and manage marine resources, particularly marine fish of ecological and economic importance. This study seeks to add to the growing body of knowledge in the molecular evolution of marine life by illuminating ambiguity in the evolutionary relationships and history of the ecologically and economically important group of fishes commonly known as tunas while providing a bioinformatic framework on data optimization.

Literature Review of Tunas

One taxonomical group that are remarkably similar are the members of *Thunnus* (Richards & Dove 1971, Collette 1978, Chow & Kishino 1995, Ward 1995, Graham & Dickson 2000, Graham & Dickson 2004). This genus is composed of eight species and are generally known as true tunas (Collette et al. 2001). The genus is nested within the family Scombridae, residing within Tribe Thunnini (Collette 1978, Collette et al. 2001). Considered a distinctively unique group

within the Scombrids, the tribe is composed of five genera: *Allothunnus*, *Auxis*, *Euythnnus*, *Katsuwonus*, and *Thunnus* (Collette 1978, Graham & Dickson 2004). Distinguishable from most other teleost, members of the tribe maintain high metabolic and digestive rates that supports their specialized locomotion (Magnusson et al. 1978, Korsmeyer & Dewar 2001, Westneat & Wainwright 2001). With the exception of Slender Tuna (*Allothunnus fallai*), the position of their red myotomal muscle and their lateral heat-exchanging retia are characteristics that make them distinguishable from the other Scombrids (Collette 1978, Block 1991, Collette et al. 2001, Graham & Dickson 2004).

The Slender Tuna is considered the most basal member of the tribe, being established as a monotypic genus (Collette 1978). It was incorporated into the tribe due to the similarity between its central circulation system, size of haemal arches, and placement of red muscle with those of other tunas (Graham & Dickson 2000). The remaining four genera in Thunnini possess a countercurrent heat exchanging system of retia mirabilia that serves to keep metabolic heat warmer than their surrounding environment (Collette 1978). Excluding *Allothunnus* and the genus *Thunnus*, members of the tribe possess central and lateral heat exchangers (Carey et al. 1971, Graham 1973, Graham 1975, Collette 1978). Contrarily, the genus *Thunnus* have developed a more advanced lateral heat exchanger with a well-developed visceral and cranial retia mirabilia (Collette 1978, Collette et al. 2001).

Following *Allothunnus*, *Auxis* is considered the second most primitive genus within Thunnini comprising of four species, which are the smallest species in the tribe (Collette & Aadland 1996). Aside from its size, several morphological features distinguish *Auxis spp.* from the other three more advanced genera in the tribe. These characteristics include the absence of a prominent frontoparietal foramina, the structure of its ventral branch, and origin of the dorsal and

ventral branches of their cutaneous artery (Collette 1978, Collette et al. 2001). Similarly to *Euthynnus*, the liver's right lobe extends throughout the body cavity (Collette 1978). Three species of tuna currently make up *Euthynnus*: Kawakawa (*Euthynnus affinis*), Little Tunny (*E. alletteratus*), and Black Skipjack (*E. lineatus*)(Collette 1978). Although this genus is more advanced in terms of morphological features than both *Allothunnus* and *Auxis*, its cutaneous artery's ventral branch is less developed compared to more advanced tunas (Godsil 1954). Despite the fact that the three species can be distinguished morphologically, several of the distinguishable features (e.g., number of gill rakers, presence or absence of vomerine teeth) overlap between the species, making them remarkably similar (Godsil 1954, Collette 1978).

Katsuwonus mirrors *Allothunnus* in being monotypic (Kishinouye 1915, Matsumoto et al. 1984). The Skipjack Tuna (*Katsuwonus pelamis*) was separated from *Euthynnus* based on internal differences between the genera, which includes the morphology of the liver and cutaneous artery, along with the number of vertebrae (Kishinouye 1915, Matsumoto et al. 1984). Similarly to *Auxis* and *Euthynnus*, *Katsuwonus* lacks a swim bladder, possesses a straight intestine, and has no scales posterior to the corselet (Collette 1978). Despite its liver distinguishing it from the smaller tunas, the liver resembles the livers found in more advanced, tropical tunas (Matsumoto et al. 1984). Considered the second most advanced genus in Thunnini, Skipjack Tuna typically occupy tropical and subtropical waters, but can also be found in temperate environments (Collette & Nauen 1983, Arai et al. 2005).

The most advanced genus within the tribe is *Thunnus* (Collette 1978, Graham & Dickson 2004). Earlier work from Collette (1978) proposed *Thunnus* be split into two subgenera: the temperate tunas (*Thunnus*) and the tropical tunas (*Neothunnus*). Members of the temperate, Bluefin Group consist of the Atlantic Bluefin Tuna (*Thunnus thynnus*), Southern Bluefin Tuna (*T.*

maccoyii), Pacific Bluefin Tuna (*Thunnus orientalis*), and Albacore Tuna (*Thunnus alalunga*), while the tropical, Yellowfin group is composed of Yellowfin Tuna (*T. albacares*), Longtail Tuna (*T. tonggol),* and Blackfin Tuna (*T. atlanticus*). Intermediate between the two groups, resides Bigeye Tuna (*T. obesus*), which shares a number of characters with the Yellowfin group (Gibbs & Collette 1967). However, although the different species of tunas occupy different environments, fossil evidence suggests that the common ancestor of tunas inhabited tropical waters 55-65 Ma during the Paleocene Period (Graham & Dickson 2000, Dickson & Graham 2004, Monsch 2006). Accordingly, adaptations towards temperate waters are considered derived, and the Bigeye Tuna shares with the temperate Bluefin Group several morphological adaptations, including the location of striations on the liver, liver lobe sizes, and the loss of a central heat exchanger (Gibbs & Collette 1967), which justify its placement among members of that group (Collette et al. 2001). However, molecular data suggests a closer affinity of Bigeye Tuna to the tropical Yellowfin group (Alvarado Bremer et al. 1997, Díaz-Arce et al. 2016).

Members of Thunnini share a number of hydrodynamic adaptations, such as their fusiform body shape, that allows these fishes the ability to cruise at relatively high speeds over long distances when compared to other teleosts (Magnusson et al. 1978, Altringham & Shadwick 2001). Streamlined bodies, a high-aspect-ratio lunate tail and fins, posterior finlets, and small caudal peduncle enables these genera to achieve a specialized swimming style known as thunniform (Magnusson et al. 1978, Altringham & Shadwick 2001). Their body thickness to length ratio has minimized drag, with some genera (e.g., *Thunnus*, *Katsuwonus*, and *Euthynnus*) believed to have reached a ratio where little to no improvement in drag reduction can be achieved (Hertel 1966, Altringham & Shadwick 2001). In terms of speed, species in *Thunnus* have high absolute speeds (e.g., $15ms^{-1}$); however, relative speeds in BLs^{-1} (body length per second) are slower in *Thunnus*

(~1 BLs^{-1}) when compared to the smaller species in the tribe (~2 BLs^{-1}) (Magnusson et al. 1978, Wardle et al. 1989, Altringham & Shadwick 2001). While a fusiform body with high aspect ratio fins may be shared primitive characteristics (symplesiomorphies), other characteristics may be the result of convergent evolution (homoplasies) which have evolved multiple times among scombroid fishes (Finnerty & Block 1995). Homoplasies complicate the reconstruction of phylogenies and may result in the establishment of incorrect relationships (Sanderson & Hufford 1996). Molecular data can help resolve instances where morphological innovations have evolved independently several times.

Tunas constantly cruise through the water column while maintaining minimal lateral movement and primarily using their caudal fin to generate thrust (Lighthill 1970, Altringham & Shadwick 2001). Adaptions such as larger surface areas in their gills compared to other scombrids of similar size, increased water contact via swimming speed and greater ventilation volumes than other teleosts help supply tunas an adequate amount of oxygen to meet their metabolic demands (Brill & Bushnell 2001, Graham & Dickson 2004). The efficient uptake of oxygen in tunas is also coupled with relatively large hearts for high cardiac outputs (Farrell et al. 1992, Brill & Bushnell 2001), and blood cells with a high oxygen affinity to maximize oxygen delivery (Bushnell & Brill 1992, Lowe et al. 2000). Although tunas efficiently uptake oxygen, tolerance towards hypoxic environments significantly varies between species (Bushnell & Brill 1992, Brill & Bushnell 2001, Brill et al. 2005).

Tuna larvae are generally morphologically similar (Matsumoto et al. 1972, Leis et al. 1991) and broadly require similar environmental conditions to grow with some exceptions for certain species (Llopiz et al. 2010, Reglero et al. 2014, Cornic et al. 2018). All tuna species are oviparous, broadcast spawners, and require sea surface temperatures (SST) of at least 24°C to provide their

larvae a suitable habitat for survival and growth (Schaefer 2001). Despite the geographic and or temporal limit on oceanic areas where reproduction is suitable, different life history strategies have evolved throughout the tribe. For example, some populations of Yellowfin and Bigeye Tuna have been known to spawn throughout the year in certain areas (Kikawa 1966, Suzuki 1994), while other species, such as those in *Auxis* and *Euthynnus*, reproduce seasonally (Schaefer & Marr 1948). Other species of tunas (e.g., *T. thynnus*) migrate during certain times of the year to spawn in more favorable environments than their typical feeding grounds (Rooker et al. 2007).

Similar to other teleosts, tuna larvae experience high mortality rates (Strasburg 1959, Allman & Grimes 1998, Lehodey 2008, Satoh et al. 2008), but tunas have evolved several adaptions to increase larvae survival. These adaptions include the ability for larvae to elevate muscle temperature (Dickson 1994), reduce starvation via cannibalism (Reglero et al. 2011), diel movements (Tim et al. 1990), selection of relatively high energy prey items (Purcell et al. 2005), the development of an advanced visual system (Margulies 1997), and the adoption of behavioral traits associated with food-limited environments (Llopiz et al. 2010). A notable feature of tuna larvae, is the partitioning of the environment when different species live in sympatry (Boehlert & Mundy 1994, Llopiz et al. 2010, Koched et al. 2013). This splitting of resources is also mirrored in adult populations (Bernal et al. 2017).

It is hypothesized that the diversification and radiation of tunas was facilitated by the mass extinction event during the Cretaceous-Paleogene (K-T; Carrol 1988, Monsch 2000). This event drastically reduced competitors and predators, subsequently increasing the abundance of resources available to the tunas (Dickson & Graham 2004, Friedman 2009, 2010, Near et al. 2013). At the end of the Cretaceous, changing ocean temperatures aided in the diversification of teleosts in general (Graham & Dickson 2004, Near et al. 2013). In addition to decreased competition, the

alteration of their environment through ocean cooling, gyre and thermohaline modification, and the development of new upwelling zones, may have allowed the ancestral tunas to evolve adaptations that contributed to their successful distribution on a global scale (Dickson & Graham 2004, Graham & Dickson 2004, Kumar & Kocour 2015). These hypotheses have resulted in the interpretation of *Thunnus* evolution, and thus phylogenies, in light of the niche expansion hypothesis, with tropical species considered primitive, and adaptations towards temperate environments, derived (Finnerty & Block 1995).

The fossil record of tunas extends back to the Paleocene Period (55-65 Ma), suggesting that the common ancestor to all tunas lived in tropical waters as these fossils date back prior to the cooling of ocean temperatures during this epoch (Graham & Dickson 2000, Dickson & Graham 2004, Monsch 2006). Despite some pronounced differences in size, and other morphological and physiological characteristics (Altringham & Shadwick 2001, Graham & Dickson 2001, Westneat & Wainwright 2001), most tunas occupy tropical and temperate waters as far as 45°N and S (Boyce 2004, Kumar & Kocour 2015). The ecological roles of tunas vary based on their life history (Dickson & Graham 2004, Graham & Dickson 2004, Ménard et al. 2006, Pethybridge et al. 2018). Smaller species, such as Little Tunny (*Euthynnus alletteratus*) or Blackfin Tuna, may occupy different trophic chains during their life cycle (Moore 2014) despite being considered a prey item to larger marine predators (Adams & Kerstetter 2014). Several species of tuna are regarded as apex-predators, while simultaneously being identified as prey items to sharks and killer whales (Guinet et al. 2007, Bonfil et al. 2008). Regardless of whether a tuna is categorized as an apex-predator, demonstrating top-down control (Varela et al. 2011), or supporting a higher trophic level (Hunsicker et al. 2012), tunas are critical in maintaining current ecosystem dynamics (Hinke et al. 2004).

Tunas are thought to have evolved from a common ancestor; however, each species of tuna displays a unique set of physiological and behavioral adaptations that allows it to expand its niche horizontally and vertically from the tropical environment, to colder environments (Bernal et al. 2017). Investigations on the distribution of marine fauna have shown that thermoclines do have direct and indirect effects on a species' range (Inagake & Hirano 1983, Houghton 1988, Roden & Raine 1994, Spear et al. 2001, Cartamil et al. 2011, Atienza et al. 2016). Species in *Auxis* and *Euthynnus* consistently remain above the thermocline, while some more advanced tunas are able to go beyond the thermocline for either short or extended periods (Block et al. 1997, Kitagawa et al. 2007, Schaefer & Fuller 2010, Bernal et al. 2017). This characteristic has allowed some species of tunas to travel to different regions of the ocean (vertically and horizontally) in search of resources (Kitagawa et al. 2004, Schaefer et al. 2009, Aranda et al. 2013).

Although the depth at which a thermocline is present is not temporally or spatially constant, the habitats beyond these boundaries often pose several challenges to the organisms living within them. In marine environments, thermoclines have been documented at depths as shallow as 20 meters (Leifer & Judd 2002), and as deep as 1,000 meters (Brink 1989). The difference between surface and bottom water temperatures separated by a thermocline can be greater than 20^{o}C (Conkright et al. 1998), making ventures or habitation beyond the thermocline demanding for organisms without specialized adaptations (Carey et al. 1971, Pavlov et al. 2000, Werner & Buchholz 2013). Of these adaptations, thermoregulation is one that is been reported in a variety of taxa, including teleosts (Carey & Teal 1969, Block 1991, Holland & Sibert 1994), elasmobranchs (Goldman 1997, Goldman et al. 2004), mammals (Laidre et al. 2003, Baylis et al. 2008), and birds (Ropert-Coudert et al. 2006, van Eeden et al. 2016).

All tuna species have the ability to generate heat by continuously swimming and maintaining metabolic heat within their red muscle (RM) (Graham & Dickson 2001). The relative amount and position of RM varies between tuna species, with the smaller species having proportionally more RM than larger species (Graham et al. 1983, Graham & Dickson 2001, 2004). The speed of contraction for RM in tunas is temperature dependent (Dickson 1996), which promotes propulsion through the water as their body remains stiff and movement is primarily generated through the tail (Altringham & Shadwick 2001, Graham & Dickson 2004).

These characteristics give tunas a high metabolic rate when compared to other teleost (Korsmeyer & Dewar 2001), likely contributing to the radiation of tunas to different niches in search of resources (Ciezarek et al. 2019).

As depth increases, the amount of visible light decreases exponentially (Lythgoe 1988). This has led several marine fauna to evolve traits such as light organ systems (McFall-Ngai & Montgomery 1990), proportionally large mouths (McLellan 1977), specialized olfaction structures (Jumper Jr & Baird 1991), and comparably large eyes (Nilsson et al. 2012) in order to inhabit deep waters. One notable species within Thunnini is Bigeye Tuna, which makes frequent dives below the thermocline (Schaefer & Fuller 2010). As their name implies, these tunas possess relatively large eyes that are predicted to aid in the location of prey items at depths with low visibility (Somiya et al. 2000). Given this characteristic, Bigeye Tuna behaves as a temperate species, yet morphologically, as mentioned above, it shares several characteristics with the tropical tunas, making Bigeye Tuna an intermediate between the two groups (Gibbs & Collette 1967, Le Gall et al. 1976, Collette et al. 2001).

Tuna Phylogenies in the Genomics Era

With tuna populations being increasingly affected by both overfishing and climate change, it becomes relevant to characterize the genome of these fishes to be able to formulate predictions about population status, geographic distribution, and the potential for their ecological roles to be altered. The development of next-generation sequencing (NGS) has facilitated a rapid growth of genomic and genetic research (Schuster 2007), with a number of implications in fisheries (Wenne et al. 2007, Narum et al. 2013). To date, studies have applied genomic approaches with a focus on population structure in tuna (for a review, see Kumar & Kocour 2015), with fewer investigations directed towards identifying molecular differences between species. Although several approaches have been made to understand the evolutionary history of tunas, their phylogeny remains up for debate (Alvarado Bremer et al. 1997, Santini et al. 2013, Díaz-Arce et al. 2016, Ciezarek et al. 2019).

Objectives

This study aims to use double digest Restriction-site Associated DNA Sequencing (ddRAD-Seq) data to detect genetic variants (alleles) between several members within the Thunnini tribe in order to identify SNPs that can be used to better understand the evolutionary relationships between tunas and provide an example of the ramifications of various bioinformatic pipelines in a phylogenomic framework.

Objective 1: Identify genome-wide loci useful to identify Thunnus spp. and reconstruct a phylogeny based on genomic data using single nucleotide polymorphisms (SNPs).

Hypothesis: Tunas have accumulated a suite of SNPs through random mutation and natural selection after diverging from a common ancestor, that has led to the formation of each species. These SNPs can be detected and used to delimit *Thunnus* species.

Objective 2: Determine the controversial phylogenetic position of Bigeye Tuna in relationship to the temperate (Bluefin Group) and the tropical (Yellowfin Group) tunas more closely related to the temperate (Bluefin Group) or the tropical (Yellowfin Group) tunas.

Hypothesis: Recent genomic studies suggest the Bigeye Tuna reside within the Yellowfin Group and a further insight into this specific topic can validate its phylogenetic position within *Thunnus*.

Objective 3: Estimate the average rate of molecular evolution within *Thunnus*.

Hypothesis: Tunas are morphologically and genetically similar and will display a relatively slow rate of molecular evolution when compared to other taxonomic groups.

Objective 4: Offer a timeline of speciation within *Thunnus* based on a calibrated molecular clock using a combination of fossil records and genomic data.

Hypothesis: Speciation events between the temperate and tropical tunas will coincide with historic changes to ocean temperatures and circulation.

CHAPTER II

THE EFFECTS OF MISSING DATA TREATMENT IN A REFERENCED-BASED PHYLOGENOMIC ANALYSIS: RAD-TAGS SPECIES TREE OR CONCATENATION

Introduction

The availability of Next Generation Sequencing (NGS) has provided researchers the ability to develop and test omics-related hypotheses at an unprecedent rate (Zhang et al. 2011, Koboldt et al. 2013). As the price of sequencing per sample has decreased, the number of library preparation methods have increased, with library designs focused on whole genome sequencing or other methods geared towards specific regions within a genome (reviewed in Jiang et al. 2016). Although a hypothesis may present clear goals, the appropriate sequencing approach may still remain ambiguous (Scheben et al. 2017). When investigating non-model species, methods that require no prior knowledge of the organism's genome prove to be a useful tool towards obtaining meaningful results (Burr et al. 1983, Andrews et al. 2016). Genotyping by sequencing (GBS) approaches, include a variety of methods that provide an overall assessment of the genomic variation of an organism. Most genomic variation is attributable to single nucleotide polymorphisms (SNPs), which therefore, offer the highest resolution for tracking disease genes and population history (Altshuler et al. 2000). Restriction-Site Associated DNA Sequencing (RAD-Seq) is one of various methods available of reduced representation sequencing that has the advantage over shearing methods of size-selection during library preparation of increasing the potential to identify orthologs among samples by utilizing restriction endonucleases to generate uniformly-sized small fragments required by many sequencing platforms (e.g., Illumina), while at

the same time keeping assays affordable and without prior genomic resources for focal samples (Baird et al. 2008).

Since its introduction, different variants of RAD-Seq have been introduced, each possessing its own particular set of advantages and disadvantages (Andrews et al. 2014, Andrews et al. 2016). Double digest Restriction-site Associated DNA Sequencing (ddRAD-Seq) is one of these variants which involves a set of two restriction enzymes, a rare (e.g., six-base cutter) and a common (four-base cutter), which when used in combination with unique adaptors to cut and capture genomic regions flanking the restriction sites, allow the characterization of hundreds of thousands of orthologous reads across a large number of individuals (Peterson et al. 2012). Initially designed for intraspecific characterization of genomic variation, its utility in the field of phylogenomics has been demonstrated (Leache et al. 2015, de Oca et al. 2017, Esquerré et al. 2019). Using RAD-Seq datasets has allowed the detection of introgression in freshwater fish within the genus *Rhodeus* (Takahashi et al. 2020) and between different genera of sea turtles (Arantes et al. 2020), as well as to delimit species of corals (Arrigoni et al. 2020) and newts (Rancilhac et al. 2019), and to test phylogenetic hypotheses (de Oca et al. 2017, Near et al. 2018).

Despite the numerous benefits RAD-Seq provides researchers, each variant of RAD-Seq is prone to different biases and issues (Puritz et al. 2014), particularly the dispersion of missing data across samples (Arnold et al. 2013). Previous phylogenetic work has suggested that researchers include missing sites within their datasets (Eaton et al. 2017), as the addition of more data can provide enough phylogenetic signal to distinguish the focal taxa (Tripp et al. 2017); however, this assumption has been questioned on whether the high node support is caused by true phylogenetic signal or an artifact of missing data (Dell'Ampio et al. 2014). Possible sources of missing data may be linked to true variations in the restriction sites of samples, preventing the

capture of orthologous sites; howbeit, technical issues such as library preparation and inadequate sequencing have been anticipated to have a larger effect on allelic dropout (Rivera-Colón et al. 2020). Regardless of these factors during sequencing, the quality of the tissue has a significant role on whether the sample's DNA is successfully extracted for downstream analysis and application (Cho et al. 2017). This bestows a challenge to researchers investigating rare or endangered species (Kohn & Wayne 1997), especially with organisms with few to no genomic resources (Ekblom & Galindo 2011), and of specimens captured in the field at distant locations. While many species of commercial fishes are freshly available through local fisheries, when dealing with highly migratory fishes high quality tissue sample availability is often constrained by the large distances and long periods (i.e., weeks to months) between date of capture and landing at port. While many vessels freshly freeze the samples, the temperatures are often not sufficiently low to guarantee the integrity of high molecular weight DNA. Further, even when observers are present to sample tunas on board of vessels, it is not uncommon for fish captured in longlines to have spent considerable 'soaking' times already dead, often in warm waters, before the fishes are landed, resulting in degraded DNA samples even if they are immediately preserved immediately in liquid nitrogen.

NGS has provided labs thousands of nucleotide sequences (Slatko et al. 2018), which has driven the need for best practices on how to handle the battery of sequencing information (Van der Auwera et al. 2013, Brownstein et al. 2014, Olson et al. 2015). With respect to RAD-Seq, studies show that including more loci is only marginally beneficial when trying to differentiate populations once a sufficient amount of polymorphic sites across samples have been identified (Wang et al. 2013, Díaz-Arce & Rodríguez-Ezpeleta 2019). Ideally, smaller datasets should contain enough information to represent an unbiased assessment of the genetic variation contained throughout the genome of the entire population while alleviating computational burden (Jeong et al. 2017). In

contrast to the Sanger sequencing era, the large amount of data makes manual inspection of the sequence reads impractical, compelling researchers to trust that any systematic errors are masked by the data's biological signal (Lemmon & Lemmon 2013). Methods such as the shortening of sequence reads through trimming has indicated an improvement in the quality of the sequences in both *de novo* and reference based methods (Chen et al. 2014), making this method favorable towards decreasing systematic errors.

When genomic resources are not available for the specific species being investigated, many researchers have been able to utilize genomic resources from related taxa (Clucas et al. 2016, Ng et al. 2017, Wang et al. 2019, Ito et al. 2020). Despite potential issues of utilizing a reference genome that may be too divergent from the target individuals (Günther & Nettelblad 2019, Bohling 2020), references as divergent as confamilials produced reliable estimates of nucleotide diversity and heterozygosity measurements (Galla et al. 2019). Aligning to a reference genome does not insure the absence of missing data, and best practices, such as the trimming low quality bases, are still recommended when implementing these types of approaches (Torkamaneh et al. 2016). However, the use of a reference helps minimize the presence of extraneous sequences resulting from contamination, and provides a certain level of assurance about the inclusion of orthologous loci in the analyses. Another issue to consider when dealing with multi-locus data in phylogenomics is whether to treat each locus separately (partitioning), or whether all the loci should be treated as a single locus (concatenation) when assessing phylogenies (McVay & Carstens 2013).

Concatenation methods strive towards maximizing phylogenetic signal while neglecting to take into account independent genealogies (Gadagkar et al. 2005), while species tree schemes take into account separate gene histories (i.e., gene trees), which may not reflect the true species history

(Degnan & Rosenberg 2009). Both approaches remain vulnerable to biological phenomenon such as incomplete lineage sorting (ILS; Warnow 2015), introgression (McVay et al. 2017), and recombination (Maddison 1997, Thiergart et al. 2014). Although studies are limited, both methods have produced similar results empirically (Townsend et al. 2011, Thompson et al. 2014) and through simulations (Tonini et al. 2015). Empirical evidence has also demonstrated that these methods can produce opposing topologies using the same dataset (Xi et al. 2014). While both methodologies have their advantages and disadvantages (McVay & Carstens 2013), both strategies are still vulnerable to biases caused by missing data (Hosner et al. 2016), and the effects of these needs to be investigated.

Using several members of Thunnini, this study seeks to demonstrate the utility of the reduction of missing data generated using a confamilial reference and manual curation of RAD-Seq data, to identify phylogenetically informative sites for a commercially and ecologically important taxonomic group (Macfadyen & Defaux 2016, ISSF 2018, Tidd et al. 2018). Since tunas have been documented to have high levels of genetic similarity within but not across genera (Chow & Kishino 1995), they serve as an excellent example to test for the effects of missing data on tree topologies. Although previous work has been conducted towards testing concatenation versus species trees schemes in regards to missing data (Hosner et al. 2016), this study seeks to minimize missing data using two approaches: by requiring locus presence across samples and by minimizing missing sites within shared loci. As technologies such as ddRAD-Seq improve in the amount of sequences generated, manual inspection of putative loci is becoming less routine (see Lemmon & Lemmon 2013). The need for manual inspection of NGS data has been recognized (Cantalupo & Pipas 2019), further warranting the need of proper guidelines towards data processing and validity. Given that RAD tags are of relatively small size when compared to entire genes, it is hypothesized

that concatenation methods would produce phylogenetic trees with greater branch support than species tree methods due to the overall phylogenetic signal, with even greater support when missing data is minimized.

Materials and Methods

Sample Collection and DNA Isolation

In this study, seven species of tuna differing in migratory behavior, trophic ecology, generation rate, longevity, and metabolic rate were chosen for ddRAD-Sequencing (Young & Davis 1990, Block et al. 1993, Korsmeyer & Dewar 2001, Schaefer 2001, Graham & Dickson 2004, Schaefer et al. 2009). The individuals chosen for sequencing consist of the following species: Atlantic Bluefin Tuna (*Thunnus thynnus*), Bigeye Tuna (*T. obesus*), Yellowfin Tuna (*T. albacares*), Blackfin Tuna (*T. atlanticus*), Albacore Tuna (*T. alalunga*), Skipjack Tuna (*Katsuwonus pelamis*), and Little Tunny (*Euthynnus alletteratus*). All the species were represented by a single individual with the exception of Bigeye Tuna, Yellowfin Tuna, and Little Tunny. Here, four Bigeye Tuna samples, three Yellowfin Tuna samples, and two Little Tunny samples were processed for DNA extraction and isolation, resulting in a total of 13 samples.

Prior to the start of the investigation, several DNA isolation methods were implemented for each sample in order to produce the highest quality sequences, where the isolation with the highest yield was chosen for sequencing. DNA isolation methods consisted of phenol-chloroform extraction (Sambrook et al. 1989), Qiagen Gentra Puregene Kit (Qiagen, Inc. Valencia, CA), and Zymo Quick-DNA Universal Kit (Zymo Research, Irvine, California, USA) protocols. Following, 5 μL of the DNA isolates were subject to gel electrophoresis. Ethidium bromide was used to stain the 1% tris-acetate (TA) agarose gel to assess DNA quality and quantity under a UV

transilluminator. A Qubit 2.0 fluorometer was used to quantify the amount of DNA per isolation following the manufacturer's recommendations (Life Technologies, Grand Island, NY, USA) implementing the Qubit dsDNA HS Assay kit (Invitrogen, MA), whereas a NanoDrop® 2000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA) was used to evaluate the purity of the isolations. The DNA extracts were then sent to the Texas A&M AgriLife Genomics and Bioinformatics facility (Texas A&M, College Station, TX) for additional quality and quantity evaluation using a Fragment Analyzer (Advanced Analytics), library preparation, and paired-end ddRAD-Sequencing using the *Msp*I and *Pst*I restriction enzymes for digestion (Appendix A). *Thunnus* samples were sequenced in 2015 on an Illumina HiSeq 2500, while the other two genera were sequenced in 2018 on the Illumina NovaSeq platform. Table 1 presents a synopsis of the sample's information.

Table 1: Samples used for analyses, species of the sample, regions from which samples were collected, year of collection, and year sequenced.

Sample	Species	Region of Collection	Year of Collection	Year Sequenced
Tobe0003	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0007	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0008	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0012	*Thunnus obesus*	North Atlantic	1993	2015
Talb0566	*Thunnus albacares*	Gulf of Mexico	2003	2015
Talb0569	*Thunnus albacares*	Gulf of Mexico	2003	2015
Talb0577	*Thunnus albacares*	Gulf of Mexico	2003	2015
Tatl0028	*Thunnus atlanticus*	North Atlantic	1998	2015
Tala0277	*Thunnus alalunga*	Atlantic Ocean	1994	2015
Tthy906	*Thunnus thynnus*	Ionian Sea	1999	2015
Kpel0001	*Katsuwonus pelamis*	North Pacific	1998	2018
Ealt0001	*Euthynnnus alletteratus*	North Atlantic	1998	2018
Ealt0002	*Euthynnnus alletteratus*	North Atlantic	1998	2018

Read Assessment and Clustering

The raw reads were demultiplexed, implementing the bcl2fastq Conversion Software (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-

software.html), subjected to cutadapt 1.8 (Martin 2011) for barcode removal, and assessed for quality using FastQC (Andrews 2010) by the sequencing facility prior to the release of the sequence reads to the investigators. Once the sequences were obtained, a sliding window was used to remove low quality sequences when a window comprising of 15% of the sequence contained an average lower than a 10 phred score, using the process_radtags.pl script from the STACKS v2.3 suite (Catchen et al. 2013, Rochette et al. 2019). Paired-end data generated in 2015 had forward and reverse reads split into separate files depending on the barcode used for the loci, resulting in two files per read direction. Files from the same sample and direction were compiled together in order to standardize the bioinformatic workflow.

In order to identify phylogenetically informative reads across the taxa, the sequences were aligned to the *T. orientalis* reference genome (Accession GCA_009176245.1; Suda et al. 2019) using ipyrad v.0.9.33 in "pairddrad" mode (Eaton & Overcast 2020). To ensure an adequate amount of representation for each species, every sample with the exception either of the *E. alletteratus* samples was required to contain the locus in order for the locus to be incorporated into the final assembly. The ipyrad software merges paired-end reads using VSEARCH (Rognes et al. 2016) prior to mapping reads against a reference genome, in which it implements bwa v0.7.017 (Li & Durbin 2010) and bedtools v2.29.2 (Quinlan & Hall 2010) to account for genomic features such as indels. MUSCLE v3.8.1551 (Edgar 2004) was then employed to align clustered sequences for the final assembly.

Loci Selection

To further minimize missing data in the assembly and increase the number of tuna species in the assembly, orthologs from the Pacific Bluefin reference genome were incorporated into the

assembly and loci that either were not shared between all 13 sequenced samples or contained at least 60% missing sites for seven or more samples were removed. Separate loci may possess different amounts of phylogenetic information and experience contrasting genealogies (Degnan & Rosenberg 2009, Yoder et al. 2013). Furthermore, orthologs may not necessarily evolve equivalently after divergence (Heger & Ponting 2007). As this study's purpose is not to resolve the phylogenetic relationships between scombrid members, but aims to demonstrate the implications of certain treatments on the sequence data, only a subset of informative loci were selected.

The Neighbor-Joining (NJ) method rapidly constructs phylogenies while allowing rate variation between branches (Saitou & Nei 1987), which may be practical when analyzing relatively young lineages (Masta 2000, Mihaescu et al. 2009). NJ trees for each locus were assembled using the R packages ape (Paradis et al. 2004, Paradis & Schliep 2019) and Phangorn (Schliep 2011, Schliep et al. 2016) in R (Team 2013). Ten loci that exhibited differentiation between species were selected and imported into MEGA X (Kumar et al. 2018) for manual trimming. Positions where seven or more samples contained an "N" were removed, with the exception of regions filled with Ns linking forward and reverse reads, in which three Ns were left as a remnant of these features.

Phylogeny Construction

Loci within both untrimmed and trimmed datasets were concatenated and subjected to a jModelTest (Posada 2008) in the R environment using the software Phangorn. Model selection was based on a shared top model between the two dataset when models are sorted primarily by the corrected Akaike Information Criterion (AICc; Akaike 1973, Sugiura 1978, Hurvich & Chih-Ling

1989) and secondarily by the Bayesian Information Criterion (BIC; Schwarz 1978) in increasing order for a fair comparison. To evaluate the efficacy of the trimmed loci compared to the untrimmed loci, phylogenies were inferred using IQTree2 (Minh et al. 2020) using a HKY (Hasegawa-Kishono-Yano) model (Hasegawa et al. 1985) with the parameters employing a gamma distribution (G) and the consideration of invariant sites (I) was used to create two phylogenetic trees following either a concatenation and species tree method for comparison. A series of 10,000 ultrafast bootstraps (Hoang et al. 2018) were implemented to assess branch support for each tree.

Previous work has shown that members of *Thunnus* are exceptionally similar morphologically (Gibbs & Collette 1967, Collette 1979) and genetically (Chow & Kishino 1995). The design of the NJ method is intended to produce unrooted phylogenies (Saitou & Nei 1987), which may not be appropriate when including highly divergent taxa (Zhang et al. 2014). Given this information, tree construction of untrimmed vs trimmed, and concatenation vs partitioning schemes were repeated using only *Thunnus* samples. Nucleotides were re-ran through the jModelTest to determine the whether the model selected when including all 14 samples was considered a top model for the *Thunnus* dataset following the same criteria in model ranking to increase comparability between results.

Results

Read Assessment and Clustering

The number of sequences generated per sample varied remarkably, with one sample only containing approximately 3.6% of the same amount of reads as the sample with the greatest amount of sequences retrieved (Figure 1). Altogether, 46,112,168 sequence reads were available for

analysis. An average of 909 reads were removed per sample following QC using process_radtags.pl, resulting in 46,100,348 radtags remaining (Figure 1). After alignment and clustering using the ipyrad pipeline, the data assembly resulted in 92,575 nucleotide sites representing 241 individual loci.

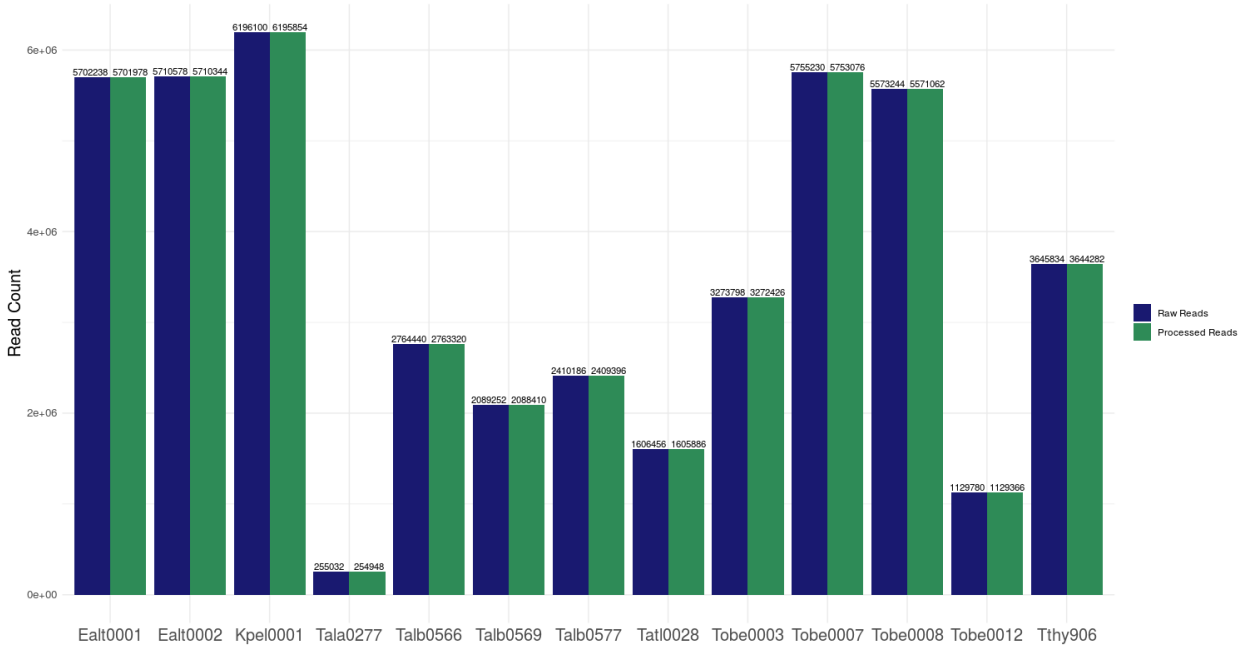

Figure 1: The number of reads across all sequenced samples before and after assessing the quality of the data by implementing process_radtags.pl. Bars are labeled at the bottom with the sample name, and at the top with the number of reads at both stages of the pipeline.

Loci Selection

Examination of the ipyrad assembly revealed 14 loci where the 13 sequenced individuals did not all contain information. Additionally, 20 loci (one of which is not genotyped for all 13 sequenced samples) were found to be of low quality (missing sites composing equal or greater than 60% of the locus across seven or more samples). Once these loci were removed from the dataset, 208 loci were retained, representing 79,978 nucleotide sites across all 14 samples. The majority of

the remaining loci did not show strong phylogenetic differentiation between species (**Figure 2**). NJ trees for loci numbers 10, 13, 38, 60, 89, 114, 170, 188, 195, and 237 were inferred to contain informative phylogenetic signal towards species delimitation within tunas. These ten loci amounted to 3,336 total sites, with all samples containing at least two-thirds of its bases genotyped. Post trimming resulted in a reduction to 2,407 sites, with the greatest percentage of missing data found to be roughly 4.61% for one sample.

Figure 2: Unrooted Neighbor-Joining (NJ) trees for all 208 loci retained after filtration of low quality and unshared loci. Ten loci (trees colored in dark red) demonstrate structuring across species and were selected for further analysis.

Phylogeny Construction

An HKY+G+I was found in the top five models when using the untrimmed and trimmed loci for the 14 samples. When excluding the *Euthynnus* and *Katsuwonus* genera, this model remained a shared top (within top 10) model between both *Thunnus* datasets. In regards to the untrimmed and trimmed datasets for all 14 samples, the topologies produced were all similar with the exception for the tree produced through gene partitioning using the trimmed assembly (**Figure 3**), which shows paraphyly between the tropical and temperate tunas (see Chapter 2). There are slight improvements in branch support on a few nodes between trees shown in **Figure 3A** and **Figure 3B** using a concatenation approach, with more branches seeing a decrease in branch support post trimming. When comparing the trees generated using the untrimmed data between the concatenation and partitioning method, nodes tend to encompass higher branch support for a similar topology following a partitioning scheme as opposed to when the loci are concatenated.

The omission of the Little Tunny and Skipjack Tuna samples resulted in comparable tree topologies for the *Thunnus* dataset regardless of the approach used (**Figure 4**). Branch support and lengths were also found to be relatively equivalent across trees (**Figure 4**). The longest branch for all four trees is representative of the Pacific Bluefin Tuna, followed by the branch leading to the Bluefin Tuna group (Pacific + Atlantic; **Figure 4**). The scale bars for trees in **Figure 4** also show a six-fold reduction in distance measured once the *Euthynnus* and *Katsuwonus* individuals were removed.

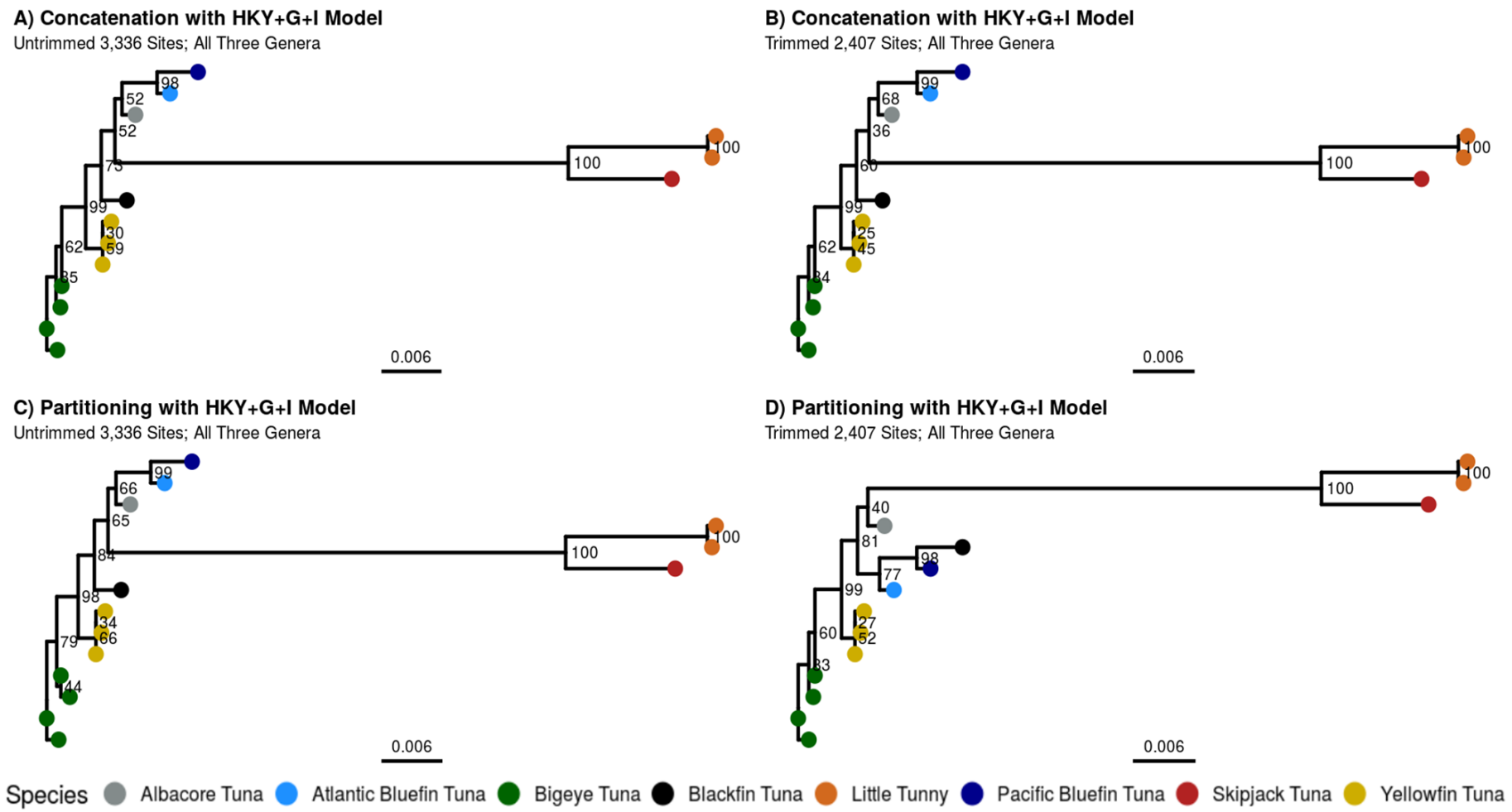

Figure 3: Trees A-B represent species trees generated through concatenation (A & B) or partitioning (C & D) of either the untrimmed (A & C) or trimmed (B & D) dataset containing genera *Thunnus*, *Euthynnus*, and *Katsuwonus*. The tree depicted in D is the most distinct of the four trees, where the Blackfin Tuna sample is now more closely related to the Pacific Bluefin Tuna. Branch support is not high (< 95) for most branches for all four trees, but topologies are generally consistent across treatments.

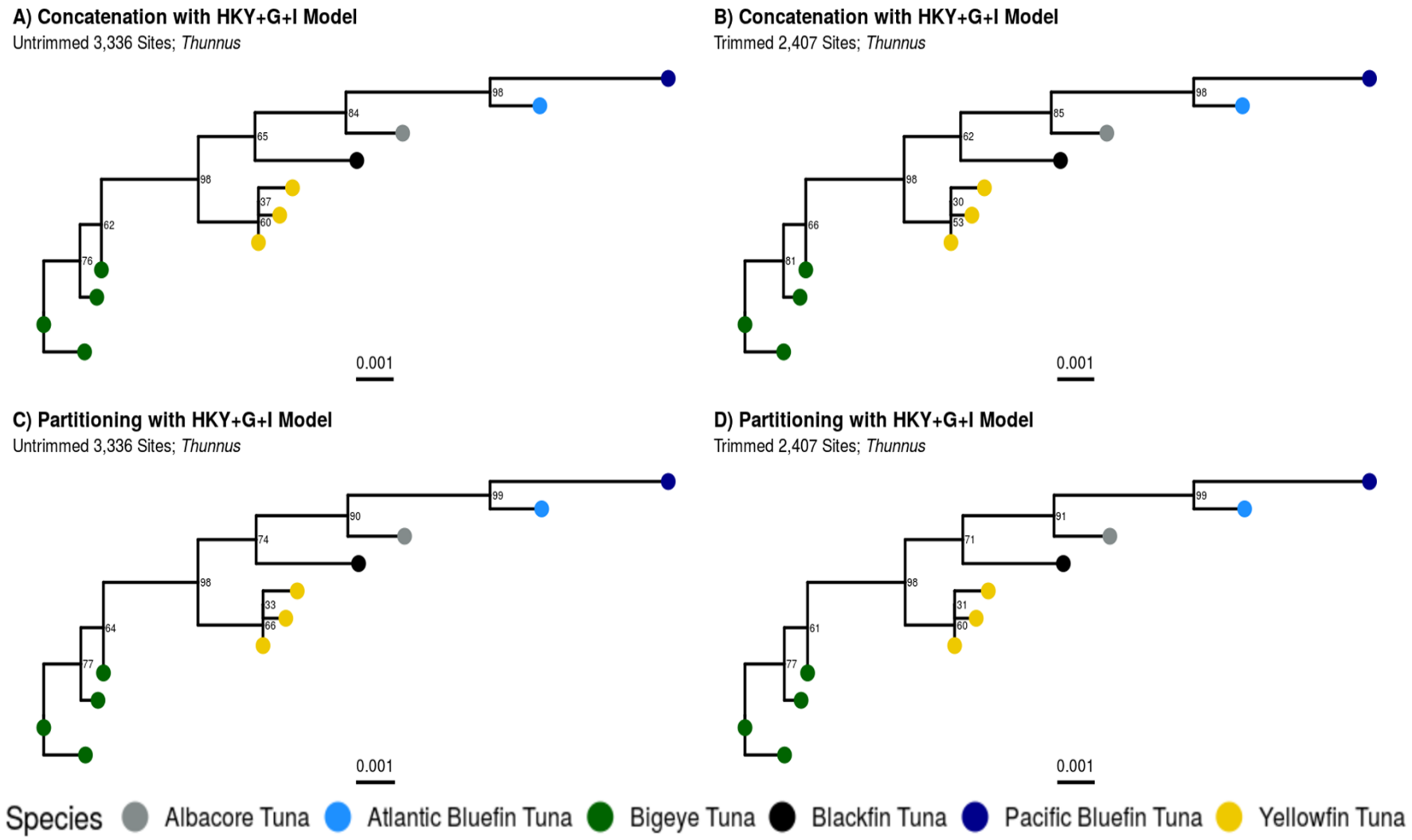

Figure 4: Trees produced across four different treatments using only the *Thunnus* samples. Trees A and C were generated following a concatenation methodology, while trees B and D followed a partitioning scheme. Despite the disparities in the number of sites representative per sample, the HKY+G+I model was able to recreate equivalent trees for all four analyzes.

Discussion

The introduction of NGS technology has greatly benefited the field of biology (Zhang et al. 2011, Koboldt et al. 2013); however, the plethora of information has presented researchers a number of challenges (Editorial 2009, Ruffalo et al. 2011, Muir et al. 2016). Experiments often begin with large datasets that are filtered down to reduce noise, preserving enough data representing the biological signal pertaining to the experiment's objectives (Edgar & Flyvbjerg 2015, Laehnemann et al. 2016). Although common quality checking and data merging methods exist, there is no consensus on a standard pipeline (Davis-Turak et al. 2017). Here, we demonstrate that different methodologies can result in similar results when the phylogenetic signal is strong.

Samples sequenced on the same machines often result in disproportional amounts of information, which can be attributed to a number of factors (Hicks et al. 2018, Rivera-Colón et al. 2020). Despite the large disparity in sequences retrieved between samples, this study was able to capture phylogenetically informative markers across relatively recent (Santini et al. 2013) and divergent taxa (see Graham & Dickson 2004). The phylogeny of tunas remains subject to debate; however, the majority of *Thunnus spp.* have been categorized into the Bluefin or Yellowfin Group (Sharp & Dizon 1972, Collette et al. 2001, Díaz-Arce et al. 2016, Ciezarek et al. 2019). Using this topology as a proxy, we find the reduction in the dataset is able to differentiate these different taxonomical groups (**Figures 3 and 4**) despite some of these organisms being reported as remarkably genetically similar (Sharp & Dizon 1972, Chow & Kishino 1995, Pujolar et al. 2003).

Similar to what other studies have concluded (Lemmon & Lemmon 2013, Davis-Turak et al. 2017), this investigation illustrates that there is no single pipeline that can be implemented towards NGS projects. Sequencing errors and biases are known contributors towards data heterogeneity (Johnson & Slatkin 2008, Taub et al. 2010), making biological heterogeneity

difficult to precisely characterize. For simplicity, this analysis selected a subset of reads (140 and 110 for the full and *Thunnus* datasets respectively) that would minimize missing data and maximize phylogenetic signal in order to better qualitatively assess the implementations of two commonly used approaches in phylogenomics: concatenation and species tree/gene trees methods. While manual selection of a particular set of loci satisfying a priori is not encouraged for empirical studies (Peterman et al. 2016, Baxter 2020), especially when results are not confidently known beforehand, selection of such data allows for the examination of phylogenetic signal through different modelling schemes.

With the exception of one tree (**Figure 3D**), the topologies produced, regardless of treatment method following the HKY+G+I model, were similar (**Figures 3 and 4**). In spite of **Figure 3C** containing more missing data than **Figure 3D**, their analogous topologies render support towards the inclusion of missing data during phylogeny reconstruction (Eaton et al. 2017, Hodel et al. 2017). **Figure 3D** displayed the most dissimilar topology and was constructed with 10 loci ranging in lengths of from 127 to 255 basepairs (bp). Genes have been documented up to lengths of greater than 2 million bp (Piovesan et al. 2019), suggesting that when considered separately (i.e. through partitioning), short loci may not provide enough phylogenetic signal to overcome systematic errors in the assembly of the data or differentiate taxonomic groups (Xi et al. 2015).

Heterogeneity in the amount of information per sample post sequencing is a known issue (Hicks et al. 2018, Cerca et al. 2020), with rare, informative alleles likely not captured in relatively under sequenced samples (Marroni et al. 2011). This possibility is even greater when applying reduce representation methods (DaCosta & Sorenson 2014). Using ddRAD-Sequencing, this study was able to identify relevant phylogenetic markers across a number of ecologically and

economically species (Hinke et al. 2004, Fromentin & Powers 2005, Pillai 2009, Macfadyen & Defaux 2016), while illustrating the utility of a unbalanced dataset across several methodologies. Although manual inspection of millions of sequence reads is impractical, researchers should evaluate and test their data when practical to ensure results are not subjected to systematic errors, producing false signals of differentiation across samples.

CHAPTER III

ESTIMATING THE MOLECULAR EVOLUTIONARY RATE WITHIN TUNAS

Introduction

The ability to detect molecular variants in the form of single nucleotide polymorphisms (SNPs), insertions-deletions (INDELS), and inversions as aided researchers in understanding the evolutionary history for numerous organisms (see Delsuc et al. 2005). Restriction-site Associated DNA Sequencing (RAD-Seq; Baird et al. 2008) has emerged as the alternative of choice for population and systematic studies of natural populations due to the ample unbiased coverage of the genome that improve the chances of characterizing ortholog sequences among samples of individuals of the same species, or among samples of different species. RAD-Seq has the advantage of being affordable, allowing the researcher to sequence large sample sizes at relatively low cost (Baird et al. 2008). Restriction endonucleases are used to generate a large number of fragments uniform in size (< 500 basepairs) that are amenable to current sequencing platforms (e.g., Illumina). Adaptors containing barcodes can be attached to DNA strands that allow sample identification, permitting multiple samples to be sequenced at once. Applying this approach allows the capture of orthologous loci between samples, providing investigators pertinent data for molecular marker detection and identification without the need of a reference genome.

Since its introduction, RAD-Seq has served as a basis for the development of other alternative reduced representation sequencing methods. Newer technologies include double digest RAD (ddRAD; Peterson et al. 2012), ezRAD (Toonen et al. 2013), 2b-RAD (Wang et al. 2012), and 3RAD (Glenn et al. 2017). These protocols vary in sample and library preparation, giving researchers options ranging from the number of restriction enzymes used, to the target volume of

sequence data estimated to be generated after sequencing. For example, ddRAD uses two restriction enzymes to cut the genome and avoids random shearing, improving some of the short comings of the original RAD-Seq protocol. Although each method encompasses its own advantages and disadvantages, the detection of molecular variance between focal groups generally remains the objective of their design (Andrews et al. 2016).

RAD-Seq is often employed in the investigation of natural populations (Davey & Blaxter 2010), with the capability of revealing population structure (Blanco-Bercial & Bucklin 2016, Vendrami et al. 2017), detecting gene flow (Dierickx et al. 2015), and identifying putative loci under selection (Benestan et al. 2016). Outside of population genomics, RAD-Seq has also been applied in the field of phylogenomics, resulting in the detection of introgression (Eaton & Ree 2013, Hou et al. 2015), species delimitation (Razkin et al. 2016), and to resolve phylogenetic relationships (Darwell et al. 2016, Near et al. 2018). Phylogenomic analysis using RAD-Seq technologies has been applied on a variety of taxonomic groups, including divergences as old as 68 Ma (Herrera et al. 2014), or as recent as 15,000 years (Wagner et al. 2013). While missing data is an issue when applying RAD-Seq approaches to survey phylogenies (Rubin et al. 2012), a number of studies have been able to resolve phylogenetic relationships with a relatively high percentage of missing data (Huang & Knowles 2016, Eaton et al. 2017, Hodel et al. 2017, Tripp et al. 2017), and produce results that have similar resolving power as other sequencing approaches (Leache et al. 2015, Manthey et al. 2016).

Prior to the development of NGS and newer technologies (e.g., 4th generation sequencing), researchers have investigated the intraspecific relationships of tunas using a suite of molecular markers including restriction fragment polymorphisms (RFLPs; Chow & Inoue 1993), allozymes (Sharp & Pirages 1978, Elliott & Ward 1995, Ward et al. 1995), mitochondrial genes (Block et al.

1993, Block & Finnerty 1994, Chow & Kishino 1995, Alvarado Bremer et al. 1997), and microsatellites (Takagi et al. 1999). In spite of decades of research, the phylogenetic relationships of tunas remains controversial, and several competing phylogenies have been proposed (Chow & Kishino 1995, Tseng et al. 2012, Santini et al. 2013, Díaz-Arce et al. 2016, Bayona-Vásquez et al. 2017, Ciezarek et al. 2019). Recently Díaz-Arce et al. (2016) produced a highly supported tree using RAD-Seq; however, even though *de novo* assemblies have produced similar results to referenced based methods (Fitz-Gibbon et al. 2017), the exclusion of unidentified paralogs in their analysis cannot be ruled out.

In addition to Tribe Thunnini sharing many morphological characteristics (Sharp & Dizon 1972), they are more similar genetically than other examples of congeneric teleosts (Chow & Kishino 1995, Elliott & Ward 1995). Speciation can occur at different time scales, with anagenesis requiring a couple thousands of years (Momigliano et al. 2017), or millions of years (Lalu et al. 2010). These speciation events are expected to be coupled with changes to the organisms' genomes; however, phenotypic divergence is not always coupled with molecular divergence (Sturmbauer & Meyer 1992, Meyer 1993), just as phenotypic convergence is not necessarily mirrored by molecular convergence (Corbett-Detig et al. 2020), and not all genes are equally involved in the process (Coyne 1992). Given that tunas are morphologically similar (Sharp & Dizon 1972, Collette et al. 2001, Graham & Dickson 2004) and considered a relatively recent lineage (Santini et al. 2013, Ciezarek et al. 2019), members of Tribe Thunnini serve as an excellent model towards better understanding the micro and macroevolutionary mechanisms operating in marine teleosts. Given that adaptability is presumed to aid a species or population evade extinction (Hoffmann & Sgrò 2011), estimating the rate members of *Thunnus* evolve is a crucial component when developing fisheries models. The purpose of this study is three-fold. First, to reconstruct the

Thunnus phylogeny to test whether the origin of temperate (Bluefin Group) and tropical tunas (Yellowfin Group) are reciprocally monophyletic. Second, to determine whether Bigeye Tuna is more closely related to the Bluefin Group, as per morphological interpretations, or to the Yellowfin Group as suggested by mtDNA gene trees. The third aim of this study is to utilize SNPs contained in loci obtained using ddRAD-Sequencing across several *Thunnus* species to provide an estimate of their molecular evolutionary rate. Altogether, we speculate that Bigeye Tuna will be nested within the Yellowfin Group, which together with the temperate tunas, will demonstrate a relatively slow rate of molecular evolution.

Materials and Methods

Sample Collection and DNA Isolation

A total of 10 samples, representing five species of tuna, were selected for DNA isolation and sequencing (**Table 2**). Although the sample sizes are small, sample sizes as low as two individuals have been demonstrated to produce reliable estimates, such as F_{st} values, when using greater than 1,000 bi-allelic markers (Willing et al. 2012, Nazareno et al. 2017), and are able to resolve phylogenetic relationships within populations (Guo et al. 2019); accordingly, small sample sizes may be sufficient to resolve species-level relationships. Samples were collected from Bigeye Tuna, Albacore Tuna, and Blackfin Tuna from the Atlantic Ocean. Yellowfin Tuna samples were collected in the Gulf of Mexico, and the Atlantic Bluefin Tuna sample was obtained from the Ionian Sea. The reference genome for the Pacific Bluefin Tuna (*T. orientalis*) was obtained from GenBank (Accession GCA_009176245.1; Suda et al. 2019). Although some samples were collected two decades prior to sequencing, age estimation of tunas have been recorded at longer timespans (Gunn et al. 2008, Shimose et al. 2009). Along with the large effective population sizes

of tunas (Ely et al. 2005) and their predicted mutation rate (Alvarado Bremer et al. 2005), twenty years does not present a sufficient amount of time for novel adaptations to become fixed in these populations.

Table 2: *Thunnus* samples used for analyses, species of the sample, regions from which samples were collected, year of collection, and year sequenced.

Sample	Species	Region of Collection	Year of Collection	Year Sequenced
Tobe0003	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0007	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0008	*Thunnus obesus*	North Atlantic	1993	2015
Tobe0012	*Thunnus obesus*	North Atlantic	1993	2015
Talb0566	*Thunnus albacares*	Gulf of Mexico	2003	2015
Talb0569	*Thunnus albacares*	Gulf of Mexico	2003	2015
Talb0577	*Thunnus albacares*	Gulf of Mexico	2003	2015
Tatl0028	*Thunnus atlanticus*	North Atlantic	1998	2015
Tala0277	*Thunnus alalunga*	Atlantic Ocean	1994	2015
Tthy906	*Thunnus thynnus*	Ionian Sea	1999	2015

DNA was isolated using a variety of methods in order to maximize DNA quality and yield (see Appendix B). These methods include the use of a Qiagen Gentra Puregene Kit (Qiagen, Inc. Valencia, CA), a Zymo Quick-DNA Universal Kit (Zymo Research, Irvine, California, USA), and a phenol-chloroform extraction protocol (Sambrook et al. 1989). The quality and quantity of supercoiled DNA was evaluated via electrophoresing of 5 µL of DNA isolate through 1% tris-acetate (TA) agarose gels stained with ethidium bromide (EtBr) followed by digital documentation through a UV transilluminator. The purity of the DNA extracts was assessed for purity using a NanoDrop® 2000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA), and the amount of DNA was quantified using a Qubit 2.0 fluorometer following the manufacturer's recommendations (Life Technologies, Grand Island, NY, USA) with the Qubit dsDNA HS Assay kit (Invitrogen, MA).

Double-Digested Restriction-Site Associated DNA (ddRAD) Sequencing

DNA samples that met the minimum standards of concentration (> 50 ng/μL) and purity (OD_{260}/OD_{280} & $OD_{260}/OD_{230} \approx 1.8$) required for massive parallel sequencing were sent to Texas A&M AgriLife Genomics and Bioinformatics facility (Texas A&M, College Station, TX) for library preparation and paired-end ddRAD sequencing using restriction enzymes *Msp*I and *Pst*I (Appendix A). These enzymes were chosen owing to demonstrating the ability to producing a desirable amount of fragments of the correct size to generate genomic data in marine fishes (Herrera et al. 2015). Tuna samples were then sequenced on an Illumina HiSeq 2500.

ddRAD Loci Assembly

Raw reads were demultiplexed using the bcl2fastq Conversion Software (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html), had barcodes removed using cutadapt 1.8 (Martin 2011), and quality checked using FASTQC (Andrews 2010) at the AgriLife Genomics and Bioinformatics facility (Texas A&M, College Station, TX). Post demultiplexing, the data was checked for demultiplexing balance, rate of prefiltering, and the presence of the technical adapter sequence at the ends of the reads. Sequences were then made available for investigation.

Errors in bioinformatic processing are recognized for negatively affecting loci recovery in ddRAD-seq (DaCosta & Sorenson 2014). For further refinement, reads were processed using process_radtags.pl in STACKS v2.3 (Catchen et al. 2013, Rochette et al. 2019). During this process, a sliding window evaluating 15% of the read's sequence was used to drop low quality sequences containing an average phred score < 10 in a window, and once low quality reads were removed, forward and reverse reads were phased. Given the current technology at the time, the

samples sequenced resulted in an additional file per pair-end, consisting of sequences possessing the alternative barcode during sequencing. FASTQ files of forward reads (R1) were added with forward reads from the same sample containing the alternative barcode, incorporating all forward reads of one sample into one file. The same process was replicated for reverse reads (R2).

Processed sequences were then aligned to the *T. orientalis* genome (Accession GCA_009176245.1; Suda et al. 2019) using the "reference" method in ipyrad v.0.9.57 (Eaton & Overcast 2020). "pairddrad" was set as the datatype and to ensure an adequate amount of representation for each species, every sample was required to contain the locus in order for the locus to be processed in the final assembly. The remaining parameters were kept at default. Briefly, bwa v0.7.17 (Li & Durbin 2010) and bedtools v2.29.2 (Quinlan & Hall 2010) were used by ipyrad to map reads that were merged using VSEARCH (Rognes et al. 2016) onto the reference genome to generate clusters of sequences, which were aligned using MUSCLE v3.8.1551 (Edgar 2004).

After the loci were assembled, the corresponding orthologous regions from the reference *T. orientalis* genome were added into the dataset in order increase the number of taxonomic groups represented within the genus. The inclusion of missing data using RAD-Seq has been shown to improve phylogenetic signal (Huang & Knowles 2016, Eaton et al. 2017, Tripp et al. 2017); however, missing sites have been shown to bias estimations in substitution rate (Excoffier & Yang 1999). To produce a more precise estimation of substitution rates among tuna genes, loci that either were not shared between all 10 samples or contained at least 50% missing sites for six or more samples were removed.

Whether sequences should be concatenated into a supermatrix or treated separately is subject to debate (McVay & Carstens 2013), with simulation and empirical studies suggesting concatenation over gene trees in order to overcome gene tree discordance (Gatesy & Springer

2014, Guevara & Steiper 2014, Rivers et al. 2016). Although the nucleotides from an individual are derived from the same genome, individual genes are subject to different mechanisms that promote variation, such as recombination (Lanier & Knowles 2012), heterotachy (Pagel & Meade 2008), incomplete lineage sorting (ILS) and hybridization (García et al. 2017). If loci are to be treated separately, it is recommended that researchers use loci that provided enough phylogenetic signal for the study's goal; however, short fragments may not harbor as many phylogenetic informative sites as opposed to longer stretches of sequences (Rivers et al. 2016). Given this information, the retained reads were concatenated into a supermatrix for rate estimation.

Phylogeny Reconstruction

To verify the supermatrix contained enough phylogenetic information to differentiate samples by species, the concatenated sequences for both the unfiltered and filtered datasets were subject to a jModelTest (Posada 2008) using the R package Phangorn (Schliep 2011, Schliep et al. 2016) in R (Team 2013). Models were sorted by the lowest corrected Akaike Information Criterion (AICc; Akaike 1973, Sugiura 1978, Hurvich & Chih-Ling 1989) and Bayesian Information Criterion (BIC; Schwarz 1978) for model selection. Two unrooted trees were then created using IQTree2 (Minh et al. 2020), implementing a HKY (Hasegawa-Kishono-Yano) model (Hasegawa et al. 1985), with the inclusion of the invariant sites parameter. A total of 10,000 ultrafast bootstraps (Hoang et al. 2018) were used to assess branch support. The R packages ggtree (Yu et al. 2017, Yu et al. 2018) and treeio (Wang et al. 2020b) were used to visualize the generated trees along with their corresponding sequences.

Using the smaller dataset, a maximum likelihood tree was generated using raxml-ng (Kozlov et al. 2019) as follows. A single maximum likelihood tree was initially constructed using

an HKY+I model, with branch support computed by utilizing 1,000 bootstrap replicates. Additionally, a phylogeny was built using MrBayes 3.2.7a (Ronquist et al. 2012) with the same substitution model implemented in the assembly of the maximum likelihood tree. The Markov Chain Monte Carlo (MCMC) chain was run for a total of 10,000 generations, sampling every 10. Sampled trees and branch lengths were then summarized with a 25% burnin to create a single representative tree. Both maximum likelihood and Bayesian trees were visualized in the FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) software.

Evolutionary Rate Analyses

To estimate the substitution rate for the assembled reads, the filtered loci were imported into BEAUTi2 to prepare the parameters for BEAST2 (Bouckaert et al. 2014). To allow the clock rate to vary, the "Automatic set clock rate" parameter was disabled. A HKY+I model was selected for the site model, while the clock model was set to a log normal relaxed clock (Drummond et al. 2006). For a tree prior, a Calibrated Yule Model (Heled & Drummond 2012) was chosen. Two log normal priors were used to calibrate the time tree. As previous studies predict species within *Thunnus* to share a common ancestor during the Paleocene (Graham & Dickson 2000, Graham & Dickson 2004, Monsch 2006), one prior was composed of all samples and was centered approximately 56 Mega-annum (Ma), with a 95% probability range between 54.1 and 58.5 Ma. The timing of the formation of the Isthmus of Panama is unsettled (Jaramillo et al. 2017), with estimates as recent as 2.8 Ma (O'Dea et al. 2016), but also including pre-Pliocene dates ranging ~6-10M a Ma (Bacon et al. 2015), as well as older dates (Montes et al. 2015). In the ensuing analyses an age of 3.5 Ma was chosen as a reasonable median to the approximate timeline of ~3.5-4.0 Ma for the closure of the Central American Seaway (Grant 1987). Accordingly, the second

prior was set to a median of 3.5 million years ago (mya; 95% probability range of 2.25-5.44 mya) as a proxy of the formation of the Isthmus of Panama, and consisted of the two sister Bluefin species, *T. orientalis* and *T. thynnus*. To ensure adequate tree sampling, the MCMC chain was set to 500 million replications with sampling occurring every thousand. The xml file was then analyzed using BEAST2 in the CIPRES Science Gateway (Miller et al. 2010).

The resulting log file was inputted into Tracer v1.7.1 (http://tree.bio.ed.ac.uk/software/tracer/) to evaluate the Effective Sample Size (ESS) for all resulting posterior parameters. Trees constructed by BEAST2 were then used as input to generate a maximum clade credibility tree, with 10% of the trees set as burnin using TreeAnnotator as part of the BEAST2 package. The maximum clade credibility tree was then visualized using FigTree v1.4.4. A literature review was then conducted in to compare the mean molecular evolution rate of tunas to other taxonomical groups. Studies using mitochondrial markers were excluded as this locus generally experiences higher mutation rates than nuclear chromosomes in animals (Allio et al. 2017) with the exception of one study on tunas. Additionally, data from Allio et al. (2017) and references therein were used to compare the nuclear and mitochondrial substitution rates across several classes of animals to the nuclear and mitochondrial substitution rate for tunas estimated from this study and Ely et al. (2001) respectively.

Gene Ontology

To associate the retained orthologs to known proteins, sequences used for molecular evolutionary rate estimation were subject to a BLASTX analysis (Altschul et al. 1990) using DIAMOND (Buchfink et al. 2015) against the NCBI non-redundant protein database. A maximum of 20 BLAST hits with an e-value threshold of e^{-15} with a tolerance of zero high-scoring segment

pairs (HSPs) were allowed. A custom python script was used to filter BLAST hits to a single representation for each query sequence. In short, BLAST hits without Reference Sequence (RefSeq v200) accessions were filtered out and the BLAST hits with descriptions associated with hypothetical proteins or uncharacterized proteins were removed. RefSeq accessions were chosen as they are generally considered high quality annotations as it is a highly curated and managed project, representing data for over 55,000 organisms across a wide range of taxonomic groups (O'Leary et al. 2016). Any query sequences that did not meet these criteria was returned back into the dataset, at which point the blast hit for each query with the lowest e-value was selected as the single representation for that query sequence. Accession IDs were then used to retrieve the Gene Ontology (GO) information associated with the IDs using all the available datasets in the ensemble database (Hubbard et al. 2002) via the R package biomaRt (Durinck et al. 2009). To avoid false overrepresentation of GO terms due to repetitive mapping, each RefSeq accession was searched across the resulting data frame using a custom python script ensuring that no accession retained information from multiple datasets. GO annotations were then visualized using the R package ggplot2 (Wickham 2016) for all three ontologies.

Results

Double-Digested Restriction-Site Associated DNA (ddRAD) Sequencing

A total of 28,503,252 sequences were obtained after paired-end ddRAD-Sequencing, with the number of sequences per sample ranging as low as 255,032, up to 5,755,230 (**Figure 5**). After filtering for low-quality reads, 28,492,172 sequence reads remained, with the average of 11,080 reads lost per sample (**Figure 5**). Evaluation of the Sequence Alignment Map (SAM) files generated when mapping the processed sequences to the *T. orientalis* genome revealed 100% of

the reads remaining after filtration in ipyrad's pipeline were properly paired and aligned. At the end of the pipeline, a total of 531 loci were retained, consisting of 249,856 sites harboring 5,463 SNPs.

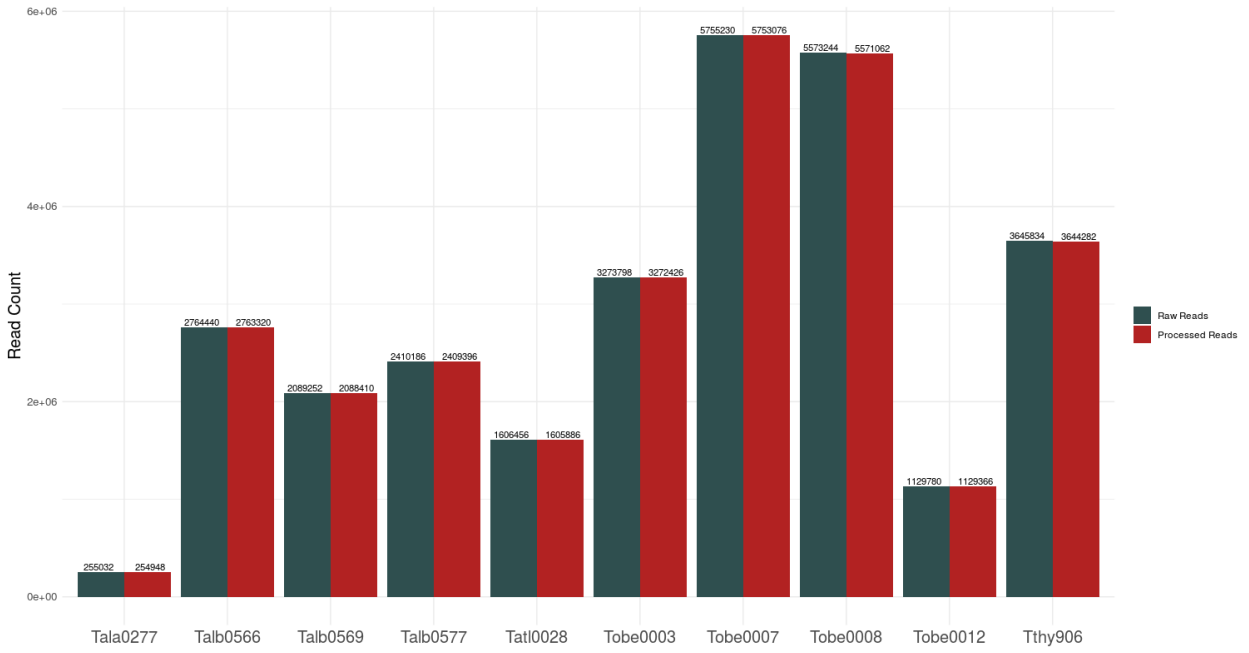

Figure 5: Bar plot of both the number of starting raw reads for all 10 samples, along with the number of reads retained per sample after quality checking using the process_radtags perl script. The values on top of the bars indicates the number of reads.

The amount of missing data within assembled loci varied across samples (**Figure 6)**. The Albacore Tuna sample remained the most under-sequenced sample, with approximately 57.48% of its sites genotyped in the alignment. The remaining samples had at least two-thirds of their nucleotides identified in the sequence matrix). A total of 107 loci were filtered out, resulting in 424 remaining loci composing of 157,171 sites, with the average amount of missing sites per locus below 25% for most samples (**Figure 7**).

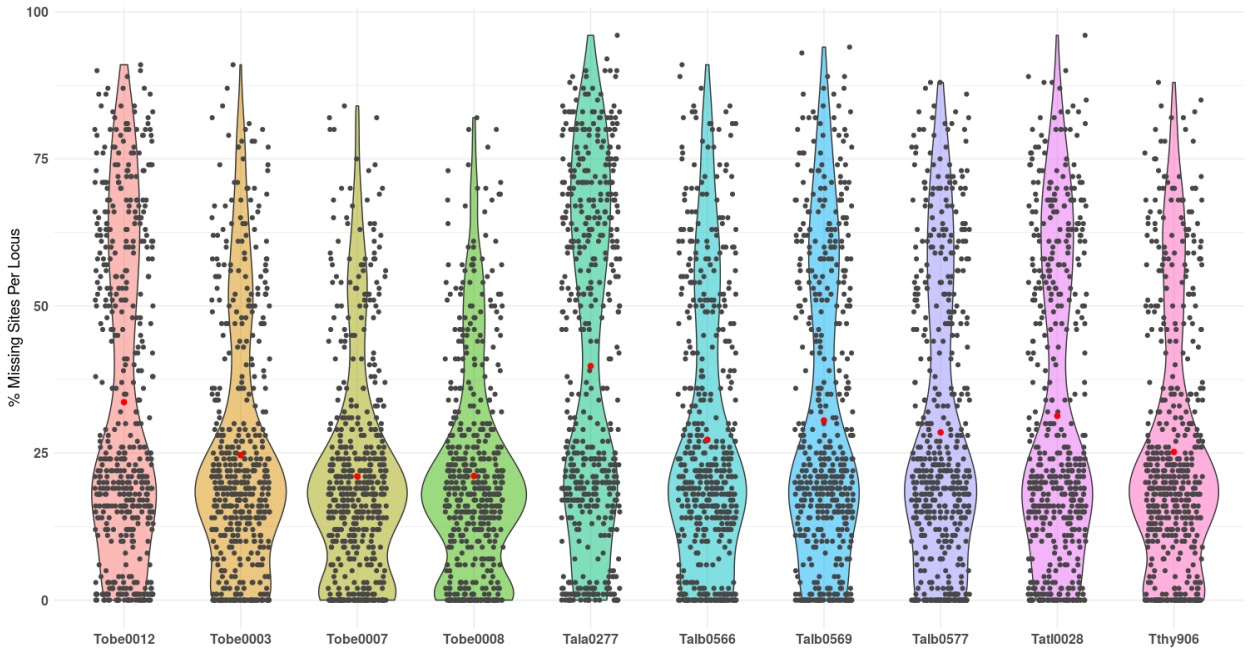

Figure 6: The distribution of the amount of missing data within the 531 loci across all 10 samples. Each point corresponds to a locus, with the red point representing the average amount of missing data for that sample.

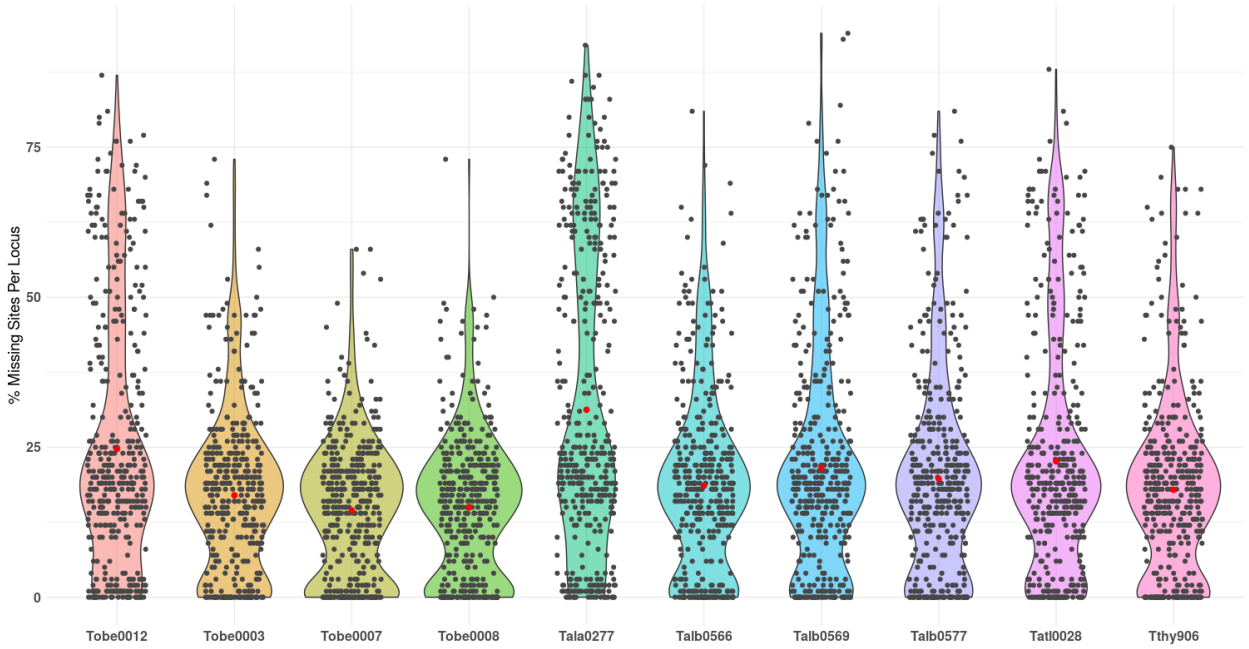

Figure 7: The distribution of the amount of missing data within the remaining 424 loci after removing relatively poorly characterized loci across tuna samples. Each point corresponds to a locus, with the red point representing the average amount of missing data for that sample.

Phylogeny Reconstruction

After the incorporation of the *T. orientalis* sequences, the tree model test revealed the HKY+I model as a high scoring model for both unfiltered and filtered datasets. The phylogeny derived using this model successfully separated the tunas by species for both the unfiltered and filtered loci (**Figure 8**). Both trees resulted in similar topologies, with *T. obesus* as sister species of *T. albacares* (**Figure 8**). Despite the unfiltered dataset containing more loci, node support was greater for the tree constructed using the smaller but cleaner dataset.

Both maximum likelihood and Bayesian methods resulted in relatively similar topologies with the exception of the Albacore Tuna (**Figures 9** & **10**). Branch support was also high for both methods, with the only low bootstrap (< 100) or posterior probability (< 1) values associated with intraspecific nodes (**Figures 9** & **10**). Similar to the previous phylogenies constructed in this study (**Figure 8**), Bigeye Tuna samples were placed within a clade shared by the Yellowfin Tuna. The Atlantic Bluefin Tuna and Pacific Bluefin Tuna are also grouped together in both topologies, with the Albacore Tuna being the most related to this clade; albeit, Albacore Tuna is shown to be grouped with the tropical tunas in the maximum likelihood tree (**Figure 9**), while using a Bayesian approach places this sample in a clade with the other temperate tunas (**Figure 10**).

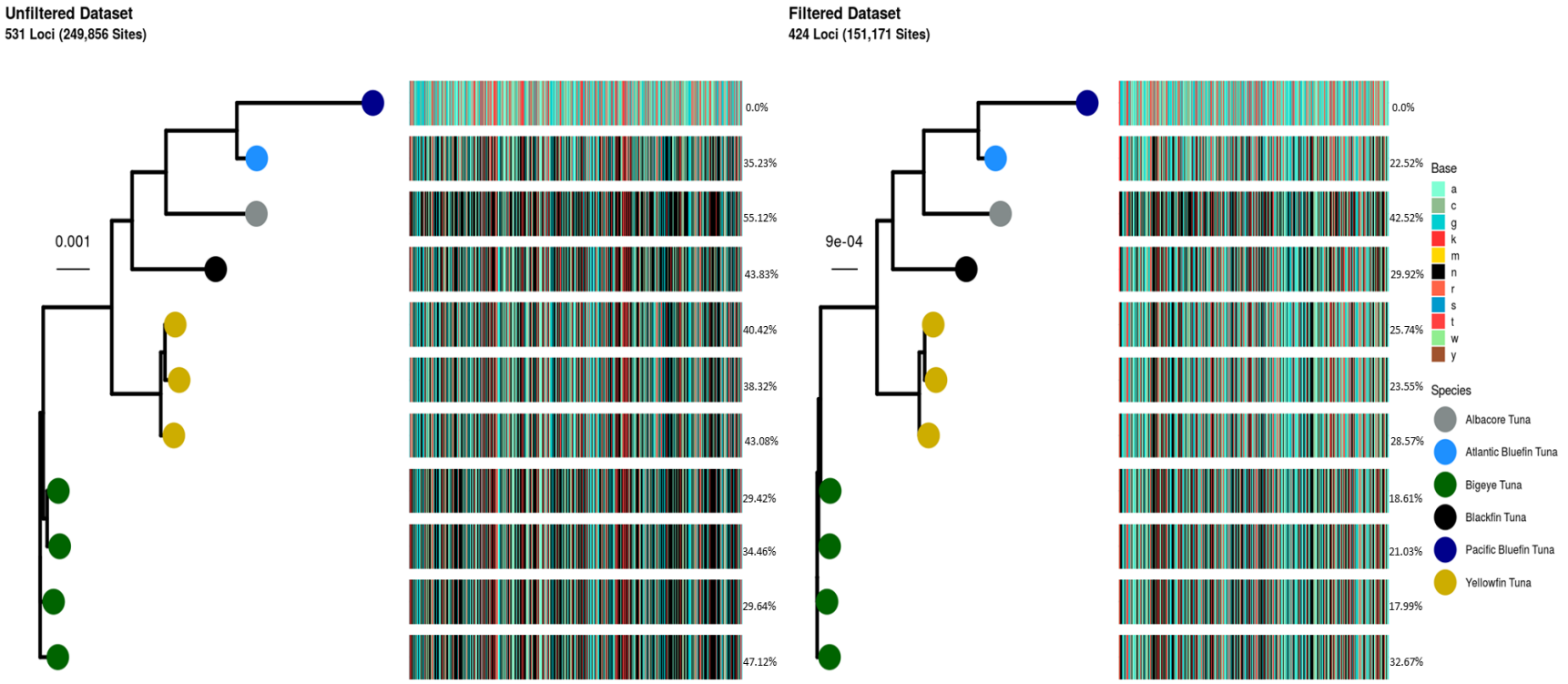

Figure 8: Unrooted phylogenetic trees of *Thunnus* samples using the loci assembled using the ipyrad pipeline for the unfiltered (531 loci) and the filtered assembly (424 loci). The alignment used to construct the trees is placed to the right of the tree, where bases color coded black indicates a missing site. The percentage of missing data for each sequence is shown to the corresponding sequence in the alignment.

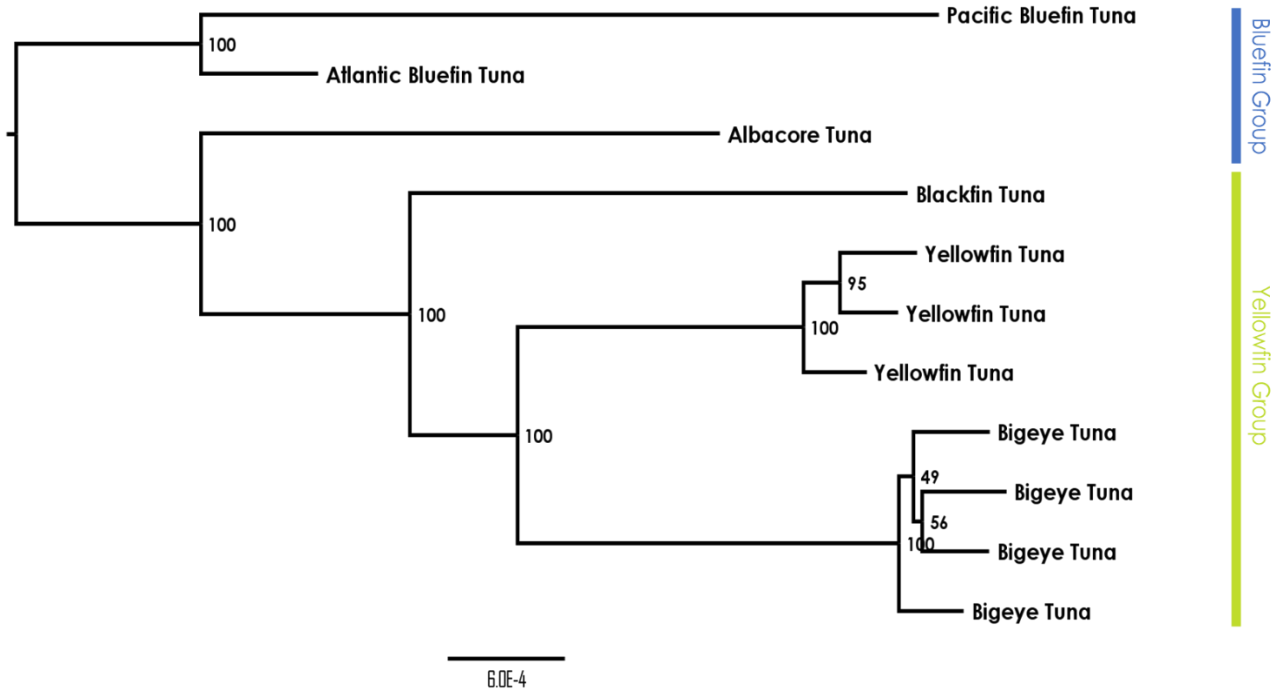

Figure 9: Maximum likelihood tree of *Thunnus spp*. using 424 orthologs. Node values represent bootstrap values.

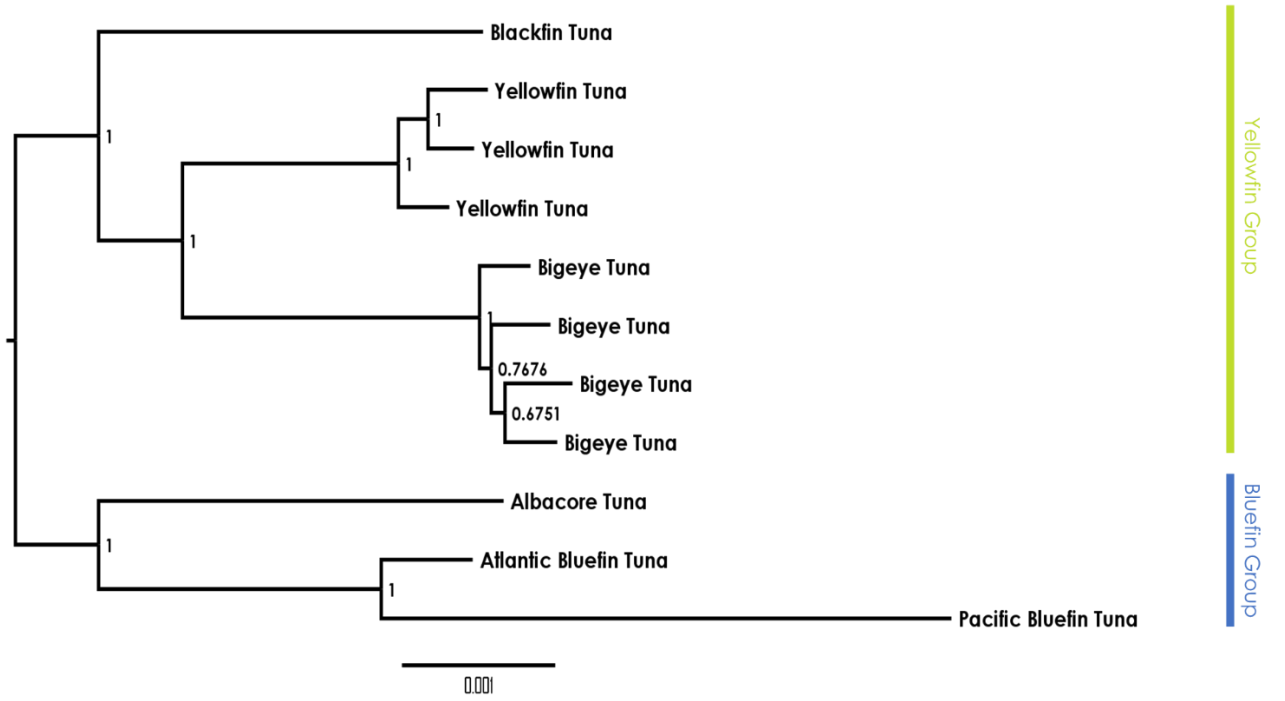

Figure 10: Summarized Bayesian tree of *Thunnus spp*. using 424 orthologs. Node values represent the posterior probability determined from the MCMC chain

Evolutionary Rate Analyses

Assessment of the Effective Sample Size (ESS) values for the posteriors sampled in the MCMC chain were greater than 200, indicating good convergence for the estimated parameters. The "rate.mean" parameter across the entire MCMC chain length with a burnin of 10%, averaged $1.085e^{-04}$ substitutions per site per million year, with a 95% highest posterior density (HPD) interval of $6.3582e^{-05}$ to $1.5427e^{-04}$. Previously, *Thunnus* was split into two groups: the temperate or Bluefin Group, and the tropical or Yellowfin Group (Collette 1979, Collette 1999). The timetree presented in this study separates the tropical tunas from the temperate tunas, with the Bigeye Tuna residing within the Yellowfin Group (**Figure 11**). The branch containing the highest rate is found to lead to the reference sequences for the Pacific Bluefin Tuna, with the highest rate for the sequenced samples leading to the Bigeye Tuna clade (**Figure 11**).

Figure 12 shows the 95% HPD for each node's age in the timetree. With the exception of the nodes containing calibrations and enforced monophyly, the other nodes separating species on the timetree contain relatively high levels of uncertainty (> 20 Ma). The largest 95% HPD for node age is located on the node supporting the Yellowfin Group, extending from the Eocene into the Pliocene. A similar distribution is also associated with the node containing the temperate tunas.

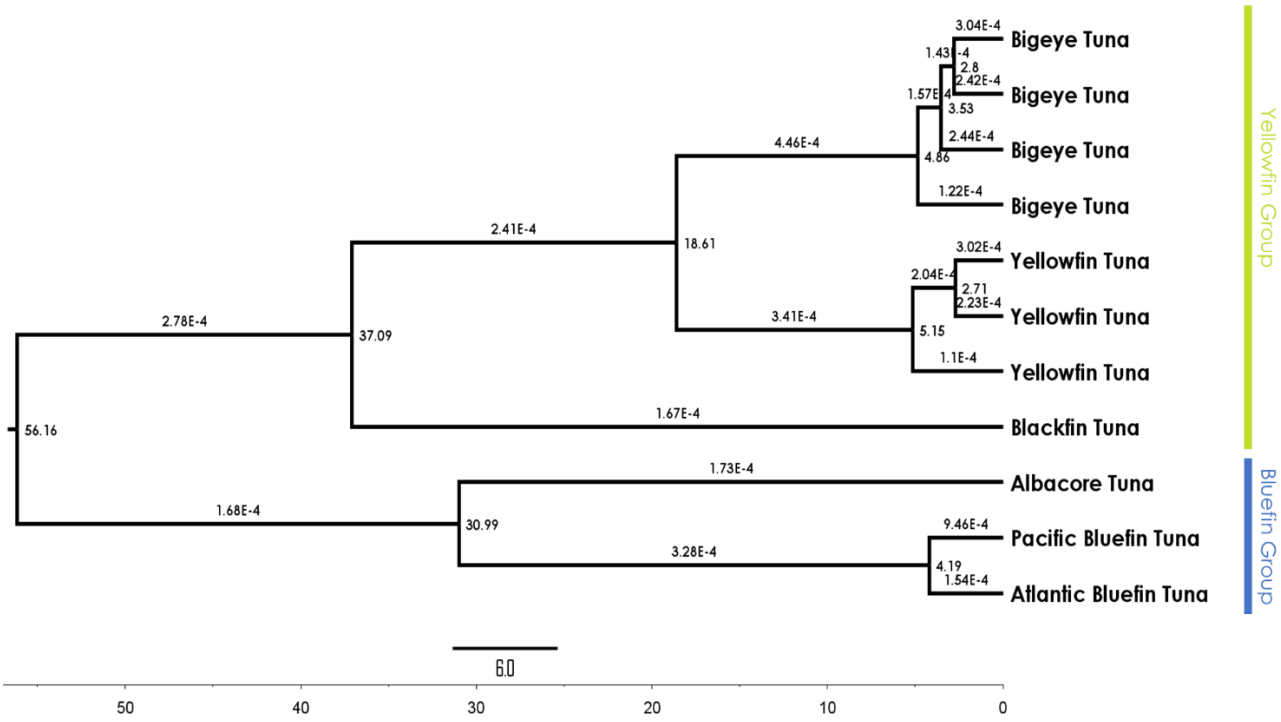

Figure 11: Timetree representing the *Thunnus* samples in this study. Diversification of both the tropical (Yellowfin Group) and temperate (Bluefin Group) tunas is estimated to have occurred between 30-40 Ma, with Bigeye Tuna following into the tropical tuna group. The branch labels correspond to the rate of evolution in per site per million year, while node annotations refer to node age in Ma.

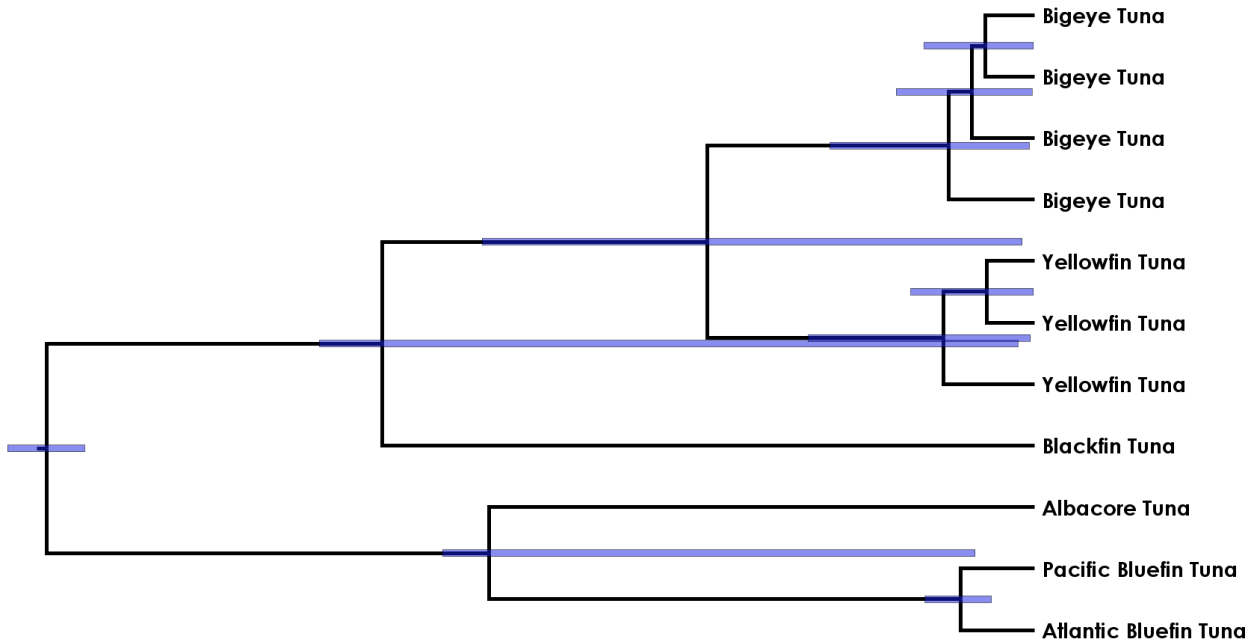

Figure 12: Timetree with blue bars representing the 95% HPD intervals for the age of each node for the genus *Thunnus*. Units on the time scale are in Ma.

The genome-wide assessment for several members of *Thunnus* indicates a mean substitution rate per site per million year of $2.648e^{-04}$. As seen in **Table 3**, the molecular rate of evolution has been documented at faster rates in cephalopods (1.35x), Antarctic fishes (1.58x-2.90x), several species of mammals (3.85x), and birds (2.50x-7.17x), but comparable to rates found in turtles, and slower than what has been reported for crocodilians. The molecular rate of evolution of tunas using mitochondrial data was estimated to be ~1.85x faster than its nuclear counterpart estimated in this study, a smaller magnitude of difference in comparison to other studies using nuclear data (**Table 3**). This mitochondrial mutation is comparable to the rate of molecular evolution estimated on the branch supporting the Bigeye Tunas (**Figure 11**), which is faster than some of the estimates calculated in previous work (**Table 3**). Furthermore, the ratio of nuclear to mitochondrial substitution rate per Ma places tunas on the lower end of nuclear to mitochondrial substitution rates per Ma when compared to several classes of animals (**Figure 13**).

Table 3: Molecular evolutionary rates across several studies and taxonomic groups. With the exception of the study by Ely et al. (2015), investigations documented here used nuclear data. The '4d' acronym is representative of fourfold degenerate sites.

Taxonomic Group	Reported Rate	Study	Data Type
Thunnus; *Thunnus* + *Katsuwonus*	2.648E-04; 4.90E-04	This Study; Ely et al. 2005	ddRAD-Seq; Mitochondrial
Mammals	1.02E-03	Literman et al. 2018	Coding Genes
Squamates	8.79E-04	Literman et al. 2018	Coding Genes
Birds	6.64E-04; 1.90E-03	Literman et al. 2018; Zhang et al. 2014	Coding Genes; 4d Sites in Coding Genes
Crocodilians	1.67E-04	Literman et al. 2018	Coding Genes
Turtles	2.38E-04	Literman et al. 2018	Coding Genes
Dissostichus mawsoni	4.20E-04	Daane et al. 2019	Targeted NGS
Chaenocephalus aceratus	7.70E-04	Daane et al. 2019	Targeted NGS
Octopus bimaculoides	3.60E-04	Albertin et al. 2015	Protein Coding Genes

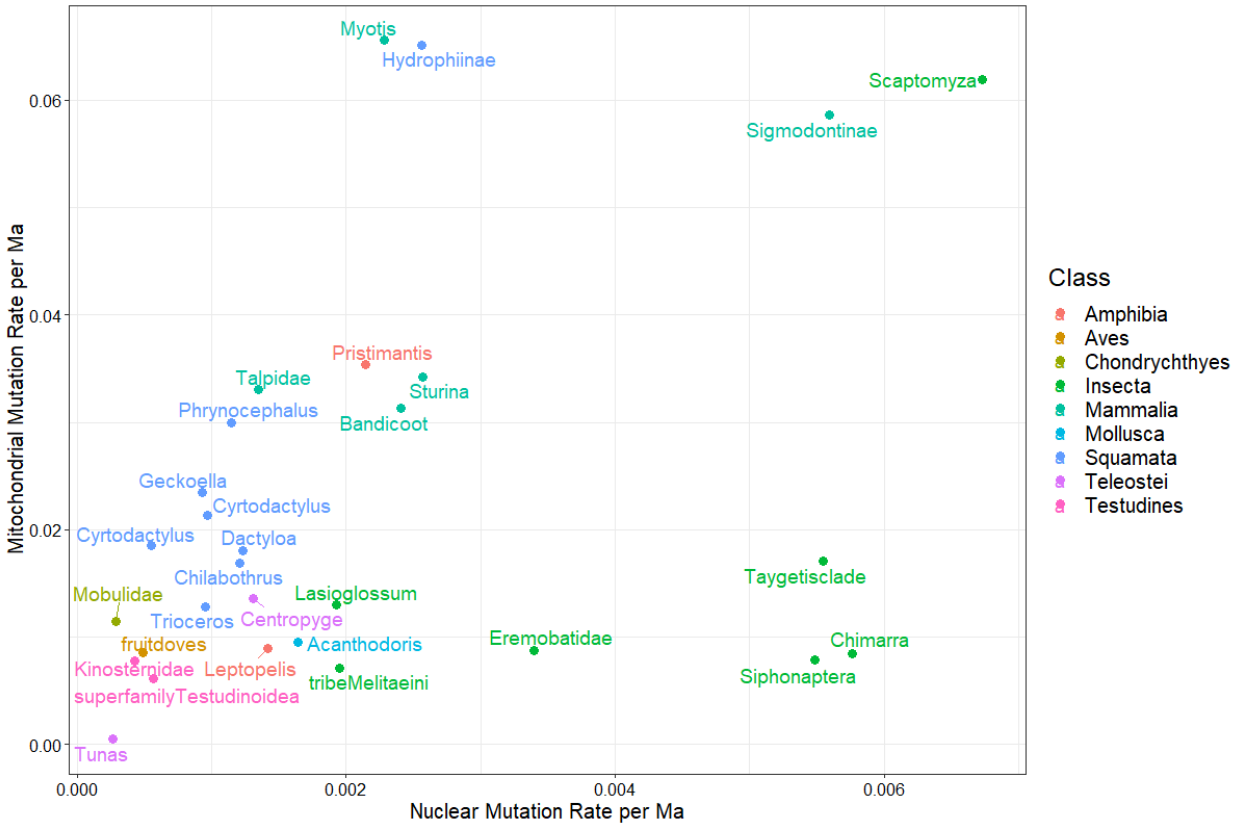

Figure 13: Tunas are estimated to have the lowest nuclear to mitochondrial mutation rate per Ma ratio when compared to estimates used in Allio et al. (2017). Nuclear mutation rate was estimated from this study, while the mitochondrial mutation rate was adapted from Ely et al. (2005).

Gene Ontology

BLAST retrieval resulted in a total of 25,517 BLAST hits. Filtration for a single representative BLAST hit per query sequence with a BLAST hit reduced the number of retrievals to 1,371, representing 159 out of the putative 424 orthologs. Approximately 4% of the 1,371 BLAST hits were found to not contain a RefSeq accession ID, resulting in 1,317 exonic accessions subject to annotation. After retrieving annotations and reducing redundancy, a total of 7,083 annotations were retained. “Regulation of transcription, DNA-templated” was the most prevalent annotation found in the biological process ontology (**Figure 14)**, protein ion binding was the most common annotation found in the molecular function ontology (**Figure 15**). In the cellular

component ontology, the membrane annotation was the most frequent description in the final biomaRt dataset (**Figure 16**). Biological processes such as eye and liver development were found in relatively high frequencies (**Figure 14**), while several annotations showed the binding of metals (**Figure 15**).

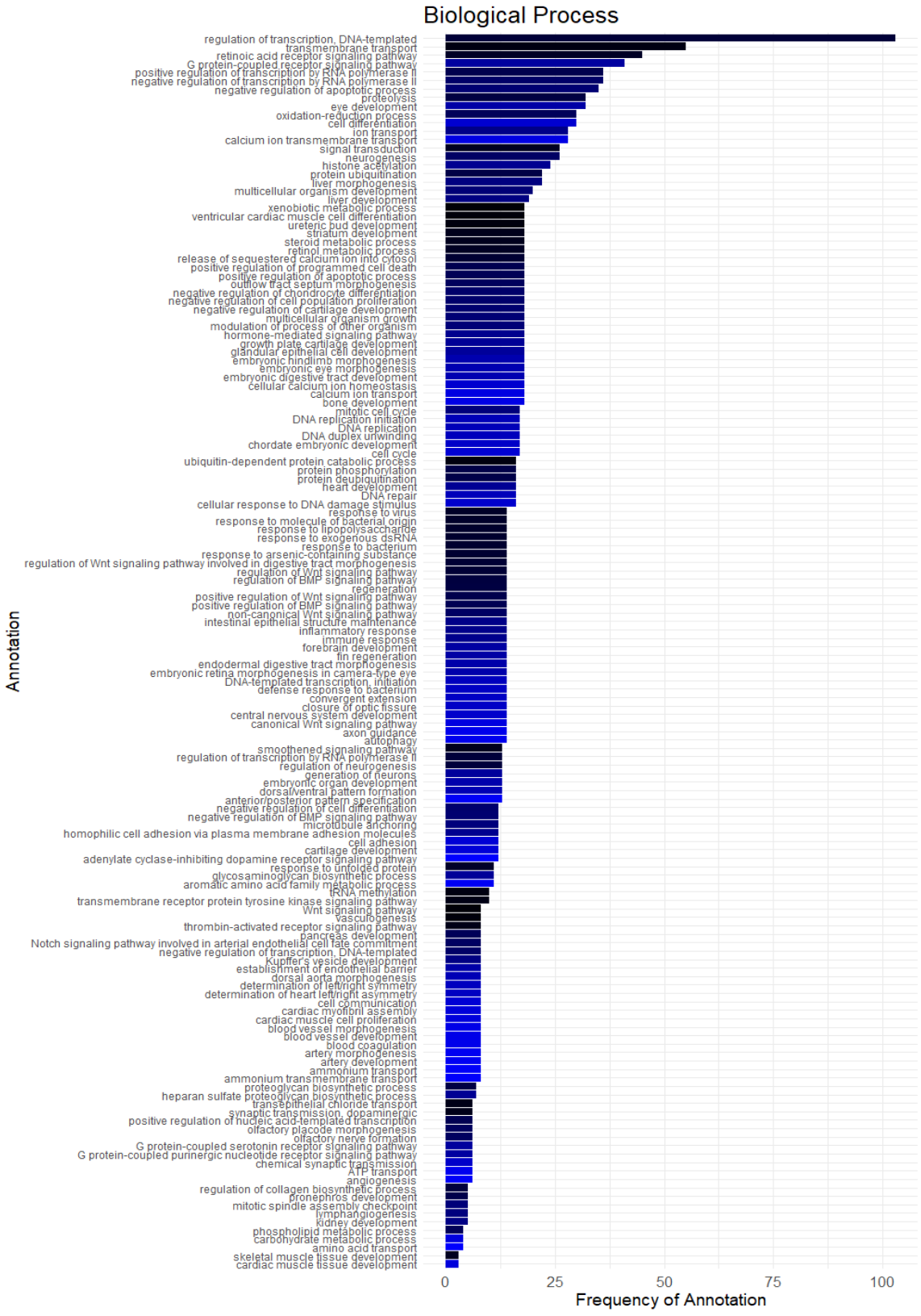

Figure 14: Annotations from the biological process ontology.

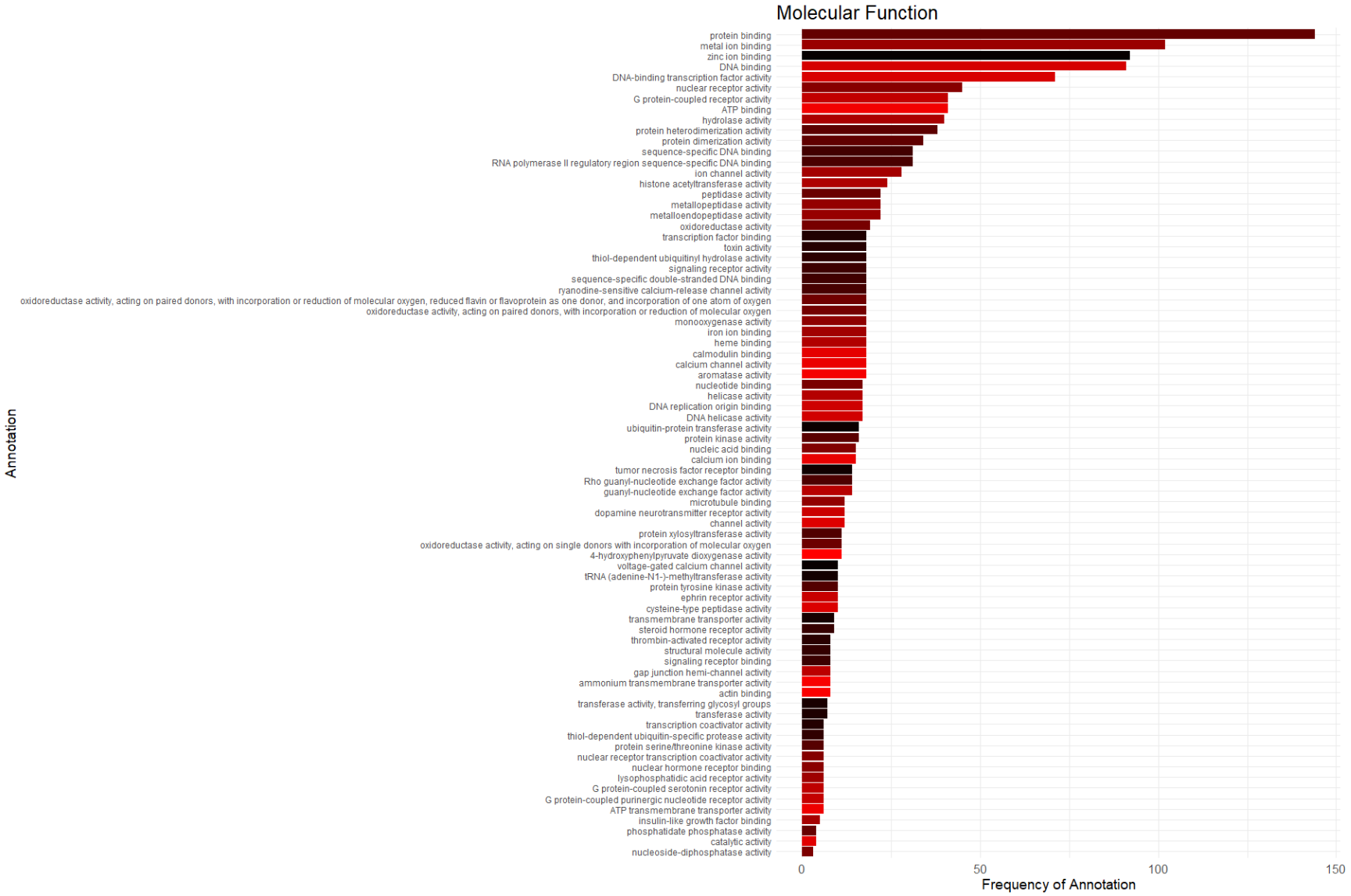

Figure 15: Annotations associated with the molecular function ontology.

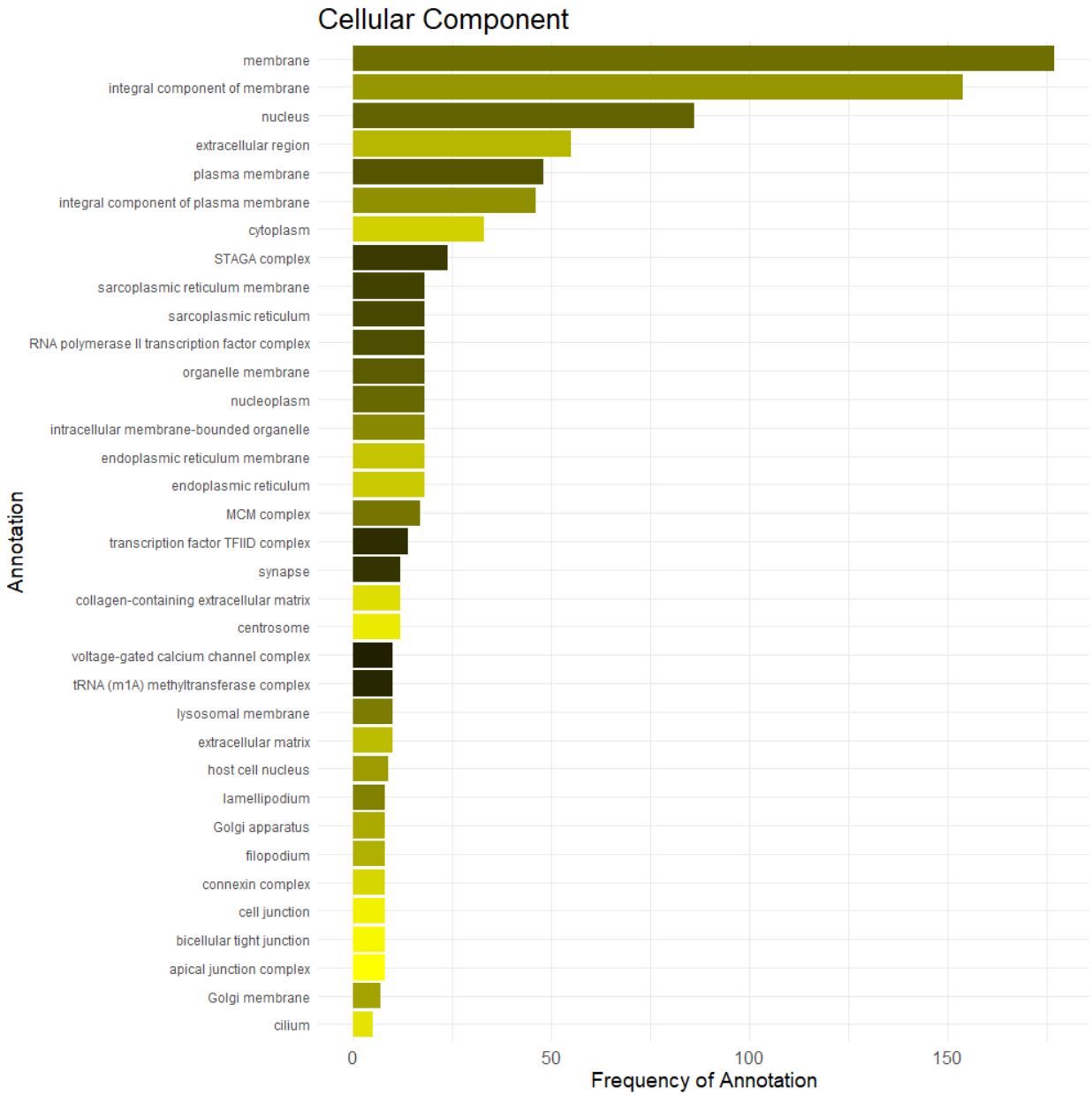

Figure 16: Bar graph of the annotations found for the cellular component ontology.

Discussion

This study aimed to obtain genome-wide loci across several members of *Thunnus* and estimate a molecular rate for these taxa. Low quality samples are often recommended to be removed from a dataset as a means of improving the quality of the overall dataset (Andrews et al.

2016); however, researchers have noted that low quality samples can be rescued and contribute to the focal study (Guo et al. 2014). Albeit there is a large variance in the amount of information initially obtained per sample, 424 loci could be retained and used to accurately delimit the species and produce phylogenies with a topology similar those of other studies using genomic and transcriptomic data (Díaz-Arce et al. 2016, Ciezarek et al. 2019). This congruence of topologies across different studies suggests that a sufficient amount of phylogenetic signal can be detected despite the reduction of data after assembling the loci, and supports Bigeye Tuna as a member of *Neothunnus*.

Tunas are constant swimmers (Boggs 1985) and have been documented swimming at speeds of $15ms^{-1}$ (Magnusson et al. 1978), which requires support by several adaptations for oxygen uptake such as relatively large surface areas in their gills and ram ventilation (Brill & Bushnell 2001, Dickson & Graham 2004). Coupled with regional endothermy, tunas are considered to have high metabolic rates relative to other teleost (Block 1991, Korsmeyer & Dewar 2001). Under the metabolic rate hypothesis, tunas are expected to display a fast rate of molecular evolution due to the constant uptake of oxygen (Shigenaga et al. 1989) and their fast metabolic rates (Martin & Palumbi 1993). Oxygen is known to have genotoxic properties, and can cause indirect or direct mutagenesis at both the mitochondrial and nuclear level (Richter et al. 1988, Joenje 1989, Martin & Palumbi 1993). Given this hypothesis, mutations are predicted to occur at higher rates in DNA receiving constant oxidative damage at higher rates as opposed to base pairs that interact with oxygen radicals at a lower frequency. This increase of mutation rate would therefore increase the molecular clock at which a species evolves.

This study estimated an average of $2.648e^{-04}$ substitution rate per site per million year for the *Thunnus* genus. With a 1:1.85 ratio when compared to the evolutionary rate using

mitochondrial genes (Ely et al. 2005), this ratio is atypically low for teleost and vertebrates in general (Allio et al. 2017). An important factor when considering this ratio is that the authors note that the mutation rate is "very conservative", and that this calculation was estimated using Skipjack Tuna, a tuna with a faster generation time than some of the *Thunnus* members used in this study (Kikawa 1966, Matsumoto et al. 1984, Ely et al. 2005, Rooker et al. 2007). Regardless of these factors, this genus was found to possess a slower molecular rate of evolution when compared to other taxonomic groups at the nuclear and mitochondrial level (**Table 3, Figure 13**). Previous studies have found evidence opposing the metabolic rate hypothesis (Mindell et al. 1996, Lanfear et al. 2007, Galtier et al. 2009), while others have provided support for it (Bleiweiss 1998, Santos 2012). This finding of a slow rate of molecular evolution and high metabolic rate is in line with previous work on scombroids (Qiu et al. 2014). Based on other factors, several alternative hypotheses have been advanced to explain molecular rate variation across taxonomic groups, including the generation time hypothesis (Wu & Li 1985), the longevity hypothesis (Nabholz et al. 2008), and suggestions of variation in DNA replication and repair efficiency (Britten 1986).

Life history traits documented for the species examined within this study vary, with some species spawning seasonally such as the Atlantic Bluefin Tuna (Rooker et al. 2007) and the Albacore Tuna (Nishikawa 1985), with others (e.g., Yellowfin Tuna) spawning year-round (Ueyanagi 1969). Species with a faster generation turnover have been reported to have a faster evolutionary rates per million years in comparison to species with a relatively slower generation rate (Laird et al. 1969, Li et al. 1996, Thomas et al. 2010). Another factor to take into consideration when evaluating possible drivers of the molecular clock, is the effective population size (N_e). Empirical evidence comparing island populations of vertebrates and invertebrates to their mainland counterparts of larger N_e have higher non-synonymous to synonymous substitution ratios

(Woolfit & Bromham 2005). Although the researchers did not find a significant difference between overall substitution rates, the results highlight the possible effects of genetic drift on the molecular clock (Woolfit & Bromham 2005, Ho 2014). Similar to generation time, interspecific differences in N_e among tunas have been described (Ely et al. 2005), and while tunas with smaller effective population sizes would be effected by genetic drift to a greater extent than tunas with larger effective population sizes, the effective population sizes in all tuna species is extremely large, and generation time may be a factor when comparing small tunas (e.g., *Euthynnus* and *Katsuwonus*) against the larger members of the genus *Thunnus*, but cannot explain the substantial differences in mutation rates compared to cephalopods and other vertebrates. Another factor often quoted as an explanation for the disparities in census size and N_e in marine animals is the variance in reproductive success (Hedgecock 1994). However, given the large effective population sizes of tunas (Ely et al. 2005), variance in reproductive success also fails to explain the reduced level of variability reported here for tunas.

While on average the substitution rate for tunas averages $2.648e^{-04}$, a considerable amount of variance is observed, with certain branches suggesting substantially faster rates. The branch with the highest rate found on the timetree is associated with the Pacific Bluefin sample (**Figure 11**). This estimate could be an artifact of the enforcement of monophyly between Pacific and Atlantic Bluefin species, as suggested by the branch's length in the phylogenetic trees in **Figures 8, 9,** and **10**. Unexpectedly, the second largest evolutionary rate detected in the timetree is located on the branch supporting the Bigeye Tuna clade (**Figure 11**). As mentioned above, the phylogenetic position of this species has been subject to debate because of the presence of intermediate physical characteristics also supported by one mitochondrial study, which place it within the Bluefin Group (Chow & Kishino 1995, Collette et al. 2001). By contrast, another

mitochondrial study and two genomic studies and one transcriptomic study support the placement of Bigeye tuna as a member of the Yellowfin Group with Longtail Tuna (*T. tonggol*) and Blackfin Tuna as sister species (Díaz-Arce et al. 2016, Bayona-Vásquez et al. 2017, Ciezarek et al. 2019). Bigeye tuna share similar morphological characteristics with temperate tunas, such as liver shape and striations on the ventral surface of this organ, along with shared vertebrae traits such as their haemal prezygapophyses (Gibbs & Collette 1967). One possible explanation for these overlaps between Bigeye Tuna and the four other members of the Bluefin group is convergent evolution, in which these shared traits are homoplasious.

Several GO annotations point towards mutations in a number of possibly adaptive genes. Liver and eye development annotations (**Figure 14**) are of key interest in regards to the Bigeye Tuna. Several authors have used liver morphology as a justification for the Bigeye Tuna's placement within the Bluefin Group (Gibbs & Collette 1967, Collette et al. 2001). Unlike their sister species Yellowfin Tuna, Bigeye Tuna are known for deep dives below the thermocline for relatively long periods of time (Schaefer et al. 2009, Schaefer & Fuller 2010), where sunlight may be limited. Adaptations towards a more acute visual system than their sister species may benefit these visual hunters (Yang et al. 2019). Future work is required in order to understand the molecular mechanisms associated with Bigeye Tuna's visual system.

The diversification of the tropical and temperate tunas is estimated to have occurred during the late Eocene and early Oligocene periods (**Figure 11**); however, there is a relatively large amount of uncertainty in these regarding these estimates (**Figure 12**). This transition between periods is characterized by a global decrease in temperature (Liu et al. 2009), the formation of Antarctic ice sheets (Ehrmann & Mackensen 1992), the extinction of numerous marine organisms (Keller 1983, Hansen 1987, Gaskell 1991, Aubry 1992, McKinney et al. 1992), and several

terrestrial species (Collinson & Hooker 1987, Prothero 1994). The timetree (**Figure 11**) suggests that Yellowfin Tuna and Bigeye Tuna diverged from a common ancestor during the Miocene, where a number of studies have reported the cooling of ocean cooling during this epoch (Shevenell et al. 2004, Majewski & Bohaty 2010, Tzanova et al. 2015), as well as an alteration to deep ocean circulations (Flower & Kennett 1994, Butzin et al. 2011). Theory predicts that adaptations to distinctive environments may be a driver for speciation (Meszéna & Hendry 2012). For example, resource heterogeneity across depth was associated with divergence between cichlids fishes in the genus *Neochromis* (Magalhaes et al. 2012), whitefish from the genus *Coregonus* (Ingram et al. 2012), and in rockfish within the *Sebastes* genus (Ingram 2011). Bigeye Tunas have developed a series of adaptations that allow them to forage for resources beyond the thermocline for extend periods of time (Holland & Sibert 1994, Dagorn et al. 2000, Schaefer & Fuller 2010), a characteristic not universal with its sister species Yellowfin Tuna (Dagorn et al. 2006). This attribute of Bigeye Tuna allows the species to exploit other resources when living sympatrically with Yellowfin Tuna, a form of niche partitioning (Sardenne et al. 2016). Taking into account the 95% HPD intervals for node ages, these results are incongruent with earlier estimates of speciation for members of *Thunnus* (Graham & Dickson 2004, Santini et al. 2013), with estimates occurring during the last 12 Ma, a time period in which global decreases of temperature have been described (Herbert et al. 2016). Since the timetree does not contain all *Thunnus* species (e.g., Longtail Tuna), and is composed of unequal taxonomic sampling of the genus, these factors are likely causing imprecise estimates of the time of divergence (Marin & Hedges 2018). However, as mentioned above, nuclear data provide strong evidence regarding the sister-species relationship of Longtail Tuna and Blackfin tuna (Díaz-Arce et al. 2016, Ciezarek et al. 2019). Accordingly, despite the biases due to an incomplete representation of the genus, these results support the placement of

Bigeye Tuna within *Neothunnus*, and provide evidence regarding the convergent evolution and homoplasy of morphological adaptations within scombrids.

Results presented here show that the two Bluefin species under investigation began to diverge from a common ancestor 4.19 Ma (**Figure 11**). This estimate disagrees the time of divergence between these two species in a study conducted by Ciezarek et al. (2019), where the time of divergence is predicted to have occurred ~1 Ma. As current evidence suggests that the formation of the Isthmus of Panama, thus the separation of the Atlantic from the Pacific Ocean, occurred at least over 2 Ma (Jaramillo et al. 2017), divergence between the Atlantic Bluefin and Pacific Bluefin Tuna populations 1 Ma after the Isthmus of Panama's formation is improbable. This study does not argue for the timing of divergence of 4.19 Ma between the Atlantic Bluefin and Pacific Bluefin species, but notes that this estimate may be closer to the precise timing of speciation.

Heavy fishing pressures have drastically reduced tuna populations in several species including the Atlantic Bluefin Tuna (Fromentin & Powers 2005, Boon 2013) and the Bigeye Tuna (Bailey et al. 2013). Overfishing has also been linked to the reduction of N_e in several marine species (Hauser et al. 2002, Hoarau et al. 2005, Chabot et al. 2015, Righi et al. 2020). As the number of individuals and alleles within a population decrease, the potential adaptability of the population is predicted to also decline, making the population more vulnerable to environmental changes (Barton 2010, Jensen & Bachtrog 2011). Climate change is a powerful force altering marine ecosystems (Gille 2002, Pörtner 2008, Lyman et al. 2010), possibly driving species with low adaptability to extinction if no other means of adjustment (e.g., range shift) is carried out (Feeley et al. 2012). A slow pace rate of evolution of members of Tribe Thunnini could represent a major challenge in light of climate as their adaptability could be outpaced by unfavorable

environmental shifts and fishing pressures (Somero 2010). Together, these results provide insight to the evolutionary history of Tribe Thunnini and present valuable information that may be implemented in designing approaches towards more effective conservation and fisheries practices regarding tunas.

CHAPTER IV

CONCLUSIONS

Several competing hypotheses have been advanced regarding the relationship of the members of the genus *Thunnus*. Analysis of conventional morphological data has suggested the division of this genus into two subgenera: the subgenus *Thunnus* (Bluefin Tuna Group) that includes Atlantic Bluefin Tuna, Pacific Bluefin Tuna, Southern Bluefin Tuna, and Albacore Tuna, and the tropical *Neothunnus* (Yellowfin Group) that includes Yellowfin Tuna, Longtail Tuna, and Blackfin Tuna, and which based on morphology are considered to be more primitive than the Bluefin Group, which represent the advanced condition for endothermic characters (Gibbs & Collette 1967, Collette 1978). Accordingly, the entire Bluefin Group is hypothesized to have radiated from a pantropical distribution to a more temperate and subpolar niche (Collette 1978, Sharp & Pirages 1978). This phylogenetic relationship has been tested numerous times, yielding often contradictory results (Chow & Kishino 1995, Block et al. 1997), and has important implications towards niche expansion hypothesis for this genus. According to this view, temperate tunas evolved from a tropical ancestor through a series of adaptations driven by thermal niche expansion and not selected for increased aerobic capacity (Block et al. 1993, but see Block et al. 1997), with the timing of these events coinciding with changes in ocean circulation patterns that took place during the Miocene (Butzin et al. 2011)

This thesis tested these hypotheses using genomic data and focused on answering four main questions: 1) do genome-wide loci data support the reciprocal monophyly of the temperate (Bluefin Tuna Group) and the tropical (Yellowfin Group) tunas within the genus *Thunnus*? 2) Is Bigeye Tuna (*T. obesus*) more closely related to the temperate subgenus *Thunnus* or the subgenus

Neothunnus? 3) What is the average rate of molecular evolution within the genus *Thunnus*, and thus 4) does the timing regarding the origin of the Tropical and Temperate groups yields support to the niche expansion hypothesis? In order to answer these questions, it was important to first address a very important bioinformatics issue: 5) what is the best approach to handle missing data in double digest Restriction-site Associated DNA Sequencing (ddRAD-Seq; Peterson et al. 2012) in phylogenomics. This last question was addressed first (Chapter 2). Accordingly, the general results and conclusions for these five questions are as follows:

Phylogeny Construction of *Thunnus* Using ddRAD-Seq

The phylogeny of *Thunnus* remains unconclusive (Díaz-Arce et al. 2016, Bayona-Vásquez et al. 2017); however, the separation of tunas by their morphological and ecological attributes into two groups (temperate and tropical) is generally accepted (Collette et al. 2001, Bernal et al. 2017, Ciezarek et al. 2019). Genome-wide loci obtained from ddRAD-Seq in Chapter 3 supports the partitioning of *Thunnus spp*., with the proposed phylogeny in congruence with nuclear (Díaz-Arce et al. 2016) and transcriptomic (Ciezarek et al. 2019) data. In here, ultrafast bootstrap support (Hoang et al. 2018) was high for most branches in both the unfiltered and filtered datasets, with both topologies separating the temperate tunas (Atlantic Bluefin Tuna, Pacific Bluefin Tuna, Albacore Tuna) from the tropical tunas (Yellowfin Tuna, Blackfin Tuna). Using a maximum likelihood and Bayesian approach revealed two different topologies; however, Bigeye Tuna remained the most related species to the Yellowfin Tuna. Despite the potential shortcomings associated with RAD-Seq (Puritz et al. 2014), this study adds to the growing literature supporting its applicability in phylogenomics (Hou et al. 2015, Eaton et al. 2017). Genome-wide loci

harboring single nucleotide polymorphisms (SNPs) were able to distinguish tunas by species and provide support to the reciprocal monophyletic origin of the Bluefin and Yellowfin Groups.

The timetree in Chapter 3 suggests that the temperate tunas separated from the tropical tunas sometime between the Paleocene and Eocene. This is in opposition of previous studies that estimate this separation to have occurred in the late Miocene (Graham & Dickson 2004, Santini et al. 2013). This earlier time frame has been characterized by a global decrease in temperatures (Herbert et al. 2016) and would be congruent with the niche transition into temperate waters. The 95% HPD intervals for the node ages on the timetree overlap with the node ages previously reported; however, these estimates did not have the highest likelihood in our results. The transition from the late Paleocene to the initial Eocene is characterized by ocean acidification (Zachos et al. 2005) and global ocean warming (Lu & Keller 1993, Sluijs et al. 2011), even in deep-sea waters (Kennett & Stott 1991). Given this information, the separation of *Thunnus* from *Neothunnus* during the Paleocene-Eocene transition would suggest that the subgenus *Thunnus* would more primitive than *Neothunnus*, as the ancestor to the tropical tunas may have occupied new, warmer waters during this time period, thus supporting the hypothesis of a tropical expansion. However, with relatively large 95% HPD intervals in our estimates (see Chapter 3), our node ages should be interpreted cautiously.

Phylogenetic Placement of Bigeye Tuna

Collette et al. (2001) concluded Bigeye Tuna be placed in the Bluefin Group based on morphology, although, noting that this species is intermediate between the temperate and tropical tunas. Recent molecular phylogenies based on nuclear data have suggested Yellowfin Tuna (*T. albacares*) as the most closely related sister species to Bigeye Tuna (Díaz-Arce et al. 2016,

Ciezarek et al. 2019). Using the Pacific Bluefin Tuna (*T. orientalis*) genome as a reference (Suda et al. 2019), genome-wide SNP data obtained using ddRAD-Seq technology yields support to this conclusion. Despite the uncertainty under the timetree constructed in Chapter 3, it can be hypothesized that speciation between Yellowfin Tuna and Bigeye Tuna was driven by climate change, as ocean temperatures began to decrease during their estimated time of divergence (Shevenell et al. 2004, Tzanova et al. 2015). As speciation has been reported to have occurred across the water column in several fish species (Ingram 2011, Ingram et al. 2012, Magalhaes et al. 2012), it is conceivable that this route of speciation may have occurred within *Thunnus*. Accordingly, the series of adaptations that allow the invasion of colder habitats that Bigeye Tuna shares with members of the Bluefin Group, which include a series of vascular characters (striations of the liver, lateral heat exchangers only, and cranial retia) are likely the result of convergent evolution (i.e., homoplasies). GO annotation of the investigated loci revealed annotations such as liver development and morphogenesis, along with annotations towards the formation of eyes. Liver morphology is used to distinguish temperate tunas from tropical tunas, and was used by Collette et al. (2001) as evidence towards Bigeye Tuna's placement within the Bluefin Group. As Bigeye Tuna's name suggests, this species of tuna possesses relatively large eyes, which may be of high value given that this species searches for prey in light limited environments (Yang et al. 2019). In conclusion, this study supports the notion of Bigeye Tuna as a member of *Neothunnus* and provides evidence of convergent evolution of several adaptive traits that enable this species to enter cooler deeper waters.

Estimating the Average Molecular Evolutionary Rate in *Thunnus*

Tunas share a number of physical traits (Gibbs & Collette 1967), leading to similar physiologies (Sharp & Dizon 1972, Graham & Dickson 2004) and adaptations such as thunniform swimming (Westneat & Wainwright 2001) and ram ventilation (Wegner et al. 2010). Additionally, members of *Thunnus* have been reported to be highly genetically similar (Chow & Kishino 1995). The molecular evolutionary rate estimated in Chapter 3 sought to investigate whether tunas were evolving slowly as inferred by morphological and genetic similarities, or whether their active lifestyle would lead to relatively high rates of molecular evolution as theory would suggest. A literature review to compare our results to other estimates suggest that tunas are a relatively slow evolving taxonomic group, supporting Smith's (1978) Optimization Theory of evolution, as tunas have been reported to have potentially maximized their swimming efficiency (Hertel 1966, Altringham & Shadwick 2001). This rate of molecular evolution along with their fusiform shape may suggest strong positive selection towards loci responsible for their fusiform body shape and thunniform swimming style.

An interesting finding in this study is the rate of molecular evolution associated with Bigeye Tuna. This rate of molecular evolution was relatively high compared to the other rates estimated in the timetree, with the exception of rate for the Pacific Bluefin Tuna, which could possibly be due to biases caused by missing data. An increased rate of molecular evolution would be in agreement with selection for new morphological traits in light of novel ecological opportunities (Thompson 1998), and may be beneficial if Bigeye Tuna diverged from Yellowfin Tuna through sympatric speciation. Although introgression between tunas has been documented (Chow & Kishino 1995, Alvarado Bremer et al. 1997, Rooker et al. 2007), the rate of divergence of genes associated with fertilization is hypothesized to display heterotachy (Doorn et al. 2001),

which may have aided in the rapid speciation of Bigeye Tuna. Altogether, these results suggest a relatively slow rate of molecular evolution within *Thunnus* and provides information necessary towards designing long-term management practices for tunas particularly when considering the effects of global climate change on the horizon (Sunday et al. 2012) and vertical redistribution of marine fauna (Brown & Thatje 2015). The inclusion of additional members of *Thunnus spp.* (i.e., Southern Bluefin Tuna and Longtail Tuna), along with additional members of closely related taxonomic groups (e.g., additional members of *Euthynnus*) would aid in providing a more precise estimate of the rate of molecular evolution in tunas.

Treatment of Missing Data Using ddRAD-Seq in a Phylogenomics Framework

Chapter 2 explored a variety of methods towards inspecting the effect of missing data in topology support using ddRAD-Seq data for tunas of the genera *Euthynnus, Katsuwonus*, and *Thunnus*. As previous studies have found, the inclusion of missing data results in higher branch support during phylogeny reconstruction in comparison to phylogenies constructed with less, but more complete datasets (Eaton et al. 2017, Tripp et al. 2017). Although numerous studies have investigated the effect of missing data in phylogenomics (Hosner et al. 2016, Huang & Knowles 2016, Tripp et al. 2017), this is the first study to our knowledge that minimizes missing data within individual loci. Concatenation methods were more efficient than partitioning approaches in producing similar topologies to topologies generated in Chapter 2; albeit, this may due to the lengths of the fragments used in these analyzes. Using complete gene sequences as opposed to rad-tags may prove to be more useful in partitioning schemes, and the testing of the effects of missing data under this framework would add the growing body of literature in regards to phylogeny construction.

Subsampling the data to only members of *Thunnus* and employing similar strategies used when analyzing all three genera of tunas resulted in dissimilar topologies. This discrepancy may be attributed to the exclusion of an outgroup, as the inclusion of an appropriate outgroup is vital towards differentiating samples within the clade of interest (Stackebrandt & Ludwig 1994). Furthermore, ddRAD-Seq is known for producing relatively short loci, which may not produce a strong phylogenetic signal when analyzed independently. Collectively, our results support the notion of including missing data when applying ddRAD-Seq in phylogenomics, and justifies the use of concatenation over partitioning methods when utilizing relatively short reads.

REFERENCES

Adams JL, Kerstetter DW (2014) Age and growth of three coastal-pelagic tunas (Actinopterygii: Perciformes: Scombridae) in the Florida Straits, USA: blackfin tuna, Thunnus atlanticus, little tunny, Euthynnus alletteratus, and skipjack tuna, Katsuwonus pelamis. Acta Ichthyologica et Piscatoria 44:201

Akaike H Information theory and an extension of the maximum likelihood principle. Proc Second International Symposium on Information Theory Budapest: Akademiai Kado

Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, Rokhsar DS (2015) The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524:220-224

Alexander AM, Su YC, Oliveros CH, Olson KV, Travers SL, Brown RM (2017) Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. Evolution 71:475-488

Allio R, Donega S, Galtier N, Nabholz B (2017) Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. Molecular Biology and Evolution 34:2762-2772

Allman RJ, Grimes CB (1998) Growth and Mortality of Little Tunny (Euthynnus alletteratus) Larvae Off the Mississippi River Plume and Panama City, Florida. Bulletin of Marine Science 62:189-197

Altringham JD, Shadwick RE (2001) Swimming and muscle function. Fish Physiology 19:313-344

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

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513-516

Alvarado Bremer J, Naseri I, Ely B (1997) Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. Journal of Fish Biology 50:540-554

Alvarado Bremer JR, Viñas J, Mejuto J, Ely B, Pla C (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. Molecular Phylogenetics and Evolution 36:169-187

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics 17:81

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G (2014) Trade-offs and utility of alternative RADseq methods: Reply to Puritz et al. Molecular Ecology 23:5943-5946

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom

Arai T, Kotake A, Kayama S, Ogura M, Watanabe Y (2005) Movements and life history patterns of the skipjack tuna Katsuwonus pelamis in the western Pacific, as revealed by otolith Sr: Ca ratios. Marine Biological Association of the United Kingdom Journal of the Marine Biological Association of the United Kingdom 85:1211

Aranda G, Abascal FJ, Varela JL, Medina A (2013) Spawning behaviour and post-spawning migration patterns of Atlantic bluefin tuna (Thunnus thynnus) ascertained from satellite archival tags. PLoS One 8:e76445

Arantes LS, Vilaca ST, Mazzoni CJ, Santos FR (2020) New genetic insights about hybridization and population structure of hawksbill and loggerhead turtles from Brazil. bioRxiv

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RAD seq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology 22:3179-3190

Arrigoni R, Berumen ML, Mariappan KG, Beck PS, Hulver AM, Montano S, Pichon M, Strona G, Terraneo TI, Benzoni F (2020) Towards a rigorous species delimitation framework for scleractinian corals based on RAD sequencing: the case study of Leptastrea from the Indo-Pacific. Coral Reefs:1-25

Atienza D, Sabatés A, Isari S, Saiz E, Calbet A (2016) Environmental boundaries of marine cladoceran distributions in the NW Mediterranean: Implications for their expansion under global warming. Journal of Marine Systems 164:30-41

Aubry M-P (1992) Late Paleogene calcareous nannoplankton evolution: a tale of climatic deterioration. Eocene-Oligocene Climatic and Biotic Evolution:272-309

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences 112:6110-6115

Bailey M, Sumaila UR, Martell SJ (2013) Can cooperative management of tuna fisheries in the western pacific solve the growth overfishing problem. Strategic Behavior and the Environment 3:31-66

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Barton N (2010) Understanding adaptation in large populations. PLoS Genetics 6:e1000987

Baxter I (2020) We aren't good at picking candidate genes, and it's slowing us down. Current Opinion in Plant Biology 54:57-60

Baylis A, Page B, Goldsworthy S (2008) Effect of seasonal changes in upwelling activity on the foraging locations of a wide-ranging central-place forager, the New Zealand fur seal. Canadian Journal of Zoology 86:774-789

Bayona-Vásquez N, Glenn T, Uribe-Alcocer M, Pecoraro C, Díaz-Jaimes P (2017) Complete mitochondrial genome of the yellowfin tuna (Thunnus albacares) and the blackfin tuna (Thunnus atlanticus): notes on mtDNA introgression and paraphyly on tunas. Conservation Genetics Resources 10:697-699

Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, Rochette R, Bernatchez L (2016) Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). Molecular Ecology 25:5073-5092

Bernal D, Brill RW, Dickson KA, Shiels HA (2017) Sharing the water column: physiological mechanisms underlying species-specific habitat use in tunas. Reviews in Fish Biology and Fisheries 27:843-880

Blanco-Bercial L, Bucklin A (2016) New view of population genetics of zooplankton: RAD-seq analysis reveals population structure of the North Atlantic planktonic copepod Centropages typicus. Molecular Ecology 25:1566-1580

Bleiweiss R (1998) Relative-rate tests and biological causes of molecular evolution in hummingbirds. Molecular Biology and Evolution 15:481-491

Block B, Keen J, Castillo B, Dewar H, Freund E, Marcinek D, Brill R, Farwell C (1997) Environmental preferences of yellowfin tuna (Thunnus albacares) at the northern extent of its range. Marine Biology 130:119-132

Block BA (1991) Endothermy in fish: thermogenesis, ecology and evolution. Biochemistry and Molecular Biology of Fishes, Book 1. Elsevier

Block BA, Finnerty JR (1994) Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. Environmental Biology of Fishes 40:283-302

Block BA, Stewart A, Kidd J (1993) Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. Science 260:210-214

Boehlert GW, Mundy BC (1994) Vertical and onshore-offshore distributional patterns of tuna larvae in relation to physical habitat features. Marine Ecology Progress Series:1-13

Boggs CH (1985) Tuna Bioenergetics and Hydrodynamics. Ph.D., The University of Wisconsin-Madison,

Bohling J (2020) Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets. Ecology and Evolution 10:7585-7601

Bonfil R, Clarke S, Nakano H (2008) The biology and ecology of the oceanic whitetip shark, Carcharhinus longimanus. Blackwell Publishing

Boon KE (2013) Overfishing of Bluefin Tuna: Incentivizing Inclusive Solutions. U Louisville L Rev 52:1

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10

Boyce D (2004) Effects of water temperature on the global distribution of tuna and billfish. B.S., Dalhousie University, Halifax, Nova Scotia

Brill RW, Bigelow KA, Musyl MK, Fritsches KA, Warrant EJ (2005) Bigeye tuna (Thunnus obesus) behavior and physiology and their relevance to stock assessments and fishery biology. Col Vol Sci Pap ICCAT 57:142-161

Brill RW, Bushnell PG (2001) The cardiovascular system of tunas. Fish Physiology 19:79-120

Brink K (1989) Observations of the response of thermocline currents to a hurricane. Journal of Physical Oceanography 19:1017-1022

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393-1398

Brown A, Thatje S (2015) The effects of changing climate on faunal depth distributions determine winners and losers. Global Change Biology 21:173-180

Brownstein CA, Beggs AH, Homer N, Merriman B, Timothy WY, Flannery KC, DeChene ET, Towne MC, Savage SK, Price EN (2014) An international effort towards developing standards for best practices in analysis, interpretation and reporting of clinical genome sequencing results in the CLARITY Challenge. Genome Biology 15:1-18

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59

Burr B, Evola S, Burr F, Beckmann J (1983) The application of restriction fragment length polymorphism to plant breeding. Genetic Engineering. Springer, Boston, MA

Bushnell PG, Brill RW (1992) Oxygen transport and cardiovascular responses in skipjack tuna (Katsuwonus pelamis) and yellowfin tuna (Thunnus albacares) exposed to acute hypoxia. Journal of Comparative Physiology B 162:131-143

Butzin M, Lohmann G, Bickert T (2011) Miocene ocean circulation inferred from marine carbon cycle modeling combined with benthic isotope records. Paleoceanography 26:1-19

Cantalupo PG, Pipas JM (2019) Detecting viral sequences in NGS data. Current Opinion in Virology 39:41-48

Carey FG, Teal JM (1969) Regulation of body temperature by the bluefin tuna. Comparative Biochemistry and Physiology 28:205-213

Carey FG, Teal JM, Kanwisher JW, Lawson KD, Beckett JS (1971) Warm-bodied fish. American Zoologist 11:137-143

Carroll RL (1988) Vertebrate paleontology and evolution. W.H. Freeman and Company, New York, NY

Cartamil DP, Sepulveda CA, Wegner NC, Aalbers SA, Baquero A, Graham JB (2011) Archival tagging of subadult and adult common thresher sharks (Alopias vulpinus) off the coast of southern California. Marine Biology 158:935-944

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124-3140

Cerca J, Maurstad MF, Rochette N, Rivera-Colón A, Rayamajhi N, Catchen J, Struck T (2020) Removing the bad apples: a simple bioinformatic method to improve loci-recovery in de novo RADseq data for non-model organisms. ecoevorxiv.org/47tka.

Chabot CL, Hawk HA, Allen LG (2015) Low contemporary effective population size detected in the Critically Endangered giant sea bass, Stereolepis gigas, due to fisheries overexploitation. Fisheries Research 172:71-78

Chen C, Khaleel SS, Huang H, Wu CH (2014) Software for pre-processing Illumina next-generation sequencing short read sequences. Source Code for Biology and Medicine 9:8

Cho M, Ahn S, Hong M, Bang H, Van Vrancken M, Kim S, Lee J, Park SH, Park JO, Park YS (2017) Tissue recommendations for precision cancer therapy using next generation sequencing: a comprehensive single cancer center's experiences. Oncotarget 8:42478

Chow S, Inoue S (1993) Intra-and interspecific restriction fragment length polymorphism in mitochondrial genes of Thunnus tuna species. Bull Natl Res Inst Far Seas Fish 30:207-225

Chow S, Kishino H (1995) Phylogenetic relationships between tuna species of the genus Thunnus (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. Journal of Molecular Evolution 41:741-748

Ciccarese S, Carlucci R, Ciani E, Corcella E, Cosentino A, Fanizza C, Linguiti G, Antonacci R (2019) Cytochrome b marker reveals an independent lineage of Stenella coeruleoalba in the Gulf of Taranto. PLoS One 14:e0213826

Ciezarek AG, Osborne OG, Shipley ON, Brooks EJ, Tracey SR, McAllister JD, Gardner LD, Sternberg MJ, Block B, Savolainen V (2019) Phylotranscriptomic insights into the diversification of endothermic Thunnus tunas. Molecular Biology and Evolution 36:84-96

Clucas GV, Younger JL, Kao D, Rogers AD, Handley J, Miller GD, Jouventin P, Nolan P, Gharbi K, Miller KJ (2016) Dispersal in the sub-Antarctic: king penguins show remarkably little population genetic differentiation across their range. BMC Evolutionary Biology 16:211

Collette B (1979) Adaptations and systematics of the mackerels and tunas. In “The Physiological Ecology of Tunas”(Sharp, GD, and Dizon, AE, Eds.). Academic Press, New York

Collette BB (1978) II. Adaptations and systematics of the mackerels and tunas. The Physiological Ecology of Tunas:7-39

Collette BB Mackerels, molecules, and morphology. Proc Proceedings of the 5th Indo-Pacific Fish Conference, Nouméa

Collette BB, Aadland C (1996) Revision of the frigate tunas (Scombridae, Auxis), with descriptions of two new subspecies from the eastern Pacific. Fishery Bulletin

Collette BB, Nauen CE (1983) Scombrids of the world: an annotated and illustrated catalogue of tunas, mackerels, bonitos, and related species known to date., Vol 2. Food and Agriculture Organization of the United Nations, Rome, Italy

Collette BB, Reeb C, Block BA (2001) Systematics of the tunas and mackerels (Scombridae). Fish Physiology 19

Collinson M, Hooker J Vegetational and mammalian faunal changes in the Early Tertiary of southern England. Proc International Congress of Systematics an Evolutionary Biology 3

Conkright M, Levitus S, O'Brien T, Boyer T, Antonov J, Stephens C (1998) World Ocean Atlas 1998 CD-ROM data set documentation. National Oceanographic Data Center (NODC) Internal Report, Silver Spring, Maryland

Corbett-Detig RB, Russell SL, Nielsen R, Losos J (2020) Phenotypic convergence is not mirrored at the protein level in a lizard adaptive radiation. Molecular Biology and Evolution 37:1604-1614

Cornic M, Smith BL, Kitchens LL, Bremer JRA, Rooker JR (2018) Abundance and habitat associations of tuna larvae in the surface water of the Gulf of Mexico. Hydrobiologia 806:29-46

Coyne JA (1992) Genetics and speciation. Nature 355:511-515

Crist DT, Scowcroft G, Harding JM (2009) World Ocean census: a global survey of marine life. Firefly, Buffalo, New York

Daane JM, Dornburg A, Smits P, MacGuigan DJ, Hawkins MB, Near TJ, Detrich III HW, Harris MP (2019) Historical contingency shapes adaptive radiation in Antarctic fishes. Nature Ecology & Evolution 3:1102-1109

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. PLoS One 9:e106713

Dagorn L, Bach P, Josse E (2000) Movement patterns of large bigeye tuna (Thunnus obesus) in the open ocean, determined using ultrasonic telemetry. Marine Biology 136:361-371

Dagorn L, Holland KN, Hallier J-P, Taquet M, Moreno G, Sancho G, Itano DG, Aumeeruddy R, Girard C, Million J (2006) Deep diving behavior observed in yellowfin tuna (Thunnus albacares). Aquatic Living Resources 19:85-88

Dann TH, Habicht C, Baker TT, Seeb JE (2013) Exploiting genetic diversity to balance conservation and harvest of migratory salmon. Canadian Journal of Fisheries and Aquatic Sciences 70:785-793

Darwell CT, Rivers DM, Althoff DM (2016) RAD-seq phylogenomics recovers a well-resolved phylogeny of a rapid radiation of mutualistic and antagonistic yucca moths. Systematic Entomology 41:672-682

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. Briefings in Functional Genomics 9:416-423

Davis-Turak J, Courtney SM, Hazard ES, Glen Jr WB, da Silveira WA, Wesselman T, Harbin LP, Wolf BJ, Chung D, Hardiman G (2017) Genomics pipelines and data integration: challenges and opportunities in the research setting. Expert Review of Molecular Diagnostics 17:225-237

de Oca AN-M, Barley AJ, Meza-Lázaro RN, García-Vázquez UO, Zamora-Abrego JG, Thomson RC, Leaché AD (2017) Phylogenomics and species delimitation in the knob-scaled lizards of the genus Xenosaurus (Squamata: Xenosauridae) using ddRADseq data reveal a substantial underestimation of diversity. Molecular Phylogenetics and Evolution 106:241-253

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24:332-340

Dell'Ampio E, Meusemann K, Szucsich NU, Peters RS, Meyer B, Borner J, Petersen M, Aberer AJ, Stamatakis A, Walzl MG (2014) Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Molecular Biology and Evolution 31:239-249

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics 6:361-375

Díaz-Arce N, Arrizabalaga H, Murua H, Irigoien X, Rodríguez-Ezpeleta N (2016) RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. Molecular Phylogenetics and Evolution 102:202-207

Díaz-Arce N, Rodríguez-Ezpeleta N (2019) Selecting RAD-seq data analysis parameters for population genetics: the more the better? Frontiers in Genetics 10:533

Dickson K (1994) Tunas as small as 207 mm fork length can elevate muscle temperatures significantly above ambient water temperature. Journal of Experimental Biology 190:79-93

Dickson KA (1996) Locomotor muscle of high-performance fishes: what do comparisons of tunas with ectothermic sister taxa reveal? Comparative Biochemistry and Physiology Part A: Physiology 113:39-49

Dickson KA, Graham JB (2004) Evolution and consequences of endothermy in fishes. Physiological and Biochemical Zoology 77:998-1018

Dierickx EG, Shultz AJ, Sato F, Hiraoka T, Edwards SV (2015) Morphological and genomic comparisons of H awaiian and J apanese B lack-footed A lbatrosses (P hoebastria nigripes) using double digest RAD seq: implications for conservation. Evolutionary Applications 8:662-678

Doorn GSV, Luttikhuizen PC, Weissing FJ (2001) Sexual selection at the protein level drives the extraordinary divergence of sex–related genes during sympatric speciation. Proceedings of the Royal Society of London Series B: Biological Sciences 268:2155-2161

Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols 4:1184

Eaton DA, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36:2592-2594

Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Systematic Biology 62:689-706

Eaton DA, Spriggs EL, Park B, Donoghue MJ (2017) Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66:399-412

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797

Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31:3476-3482

Editorial NM (2009) Metagenomics versus Moore's law. Nat Methods 6:623-623

Ehrmann WU, Mackensen A (1992) Sedimentological evidence for the formation of an East Antarctic ice sheet in Eocene/Oligocene time. Palaeogeography, Palaeoclimatology, Palaeoecology 93:85-112

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1-15

Elliott NG, Ward RD (1995) Genetic relationships of eight species of Pacific tunas (Teleostei: Scombridae) inferred from allozyme analysis. Marine and Freshwater Research 46:1021-1032

Ely B, Viñas J, Bremer JRA, Black D, Lucas L, Covello K, Labrie AV, Thelen E (2005) Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (Thunnus albacares) and the skipjack tuna (Katsuwonus pelamis). BMC Evolutionary Biology 5:19

Esquerré D, Ramírez-Álvarez D, Pavón-Vázquez CJ, Troncoso-Palacios J, Garín CF, Keogh JS, Leaché AD (2019) Speciation across mountains: phylogenomics, species delimitation and taxonomy of the Liolaemus leopardinus clade (Squamata, Liolaemidae). Molecular Phylogenetics and Evolution 139:106524

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Molecular Biology and Evolution 16:1357-1368

Farrell A, Davie P, Franklin C, Johansen J, Brill R (1992) Cardiac physiology in tunas. I. In vitro perfused heart preparations from yellowfin and skipjack tunas. Canadian Journal of Zoology 70:1200-1210

Feeley KJ, Rehm EM, Machovina B (2012) Perspective: the responses of tropical forest species to global climate change: acclimate, adapt, migrate, or go extinct? Frontiers of Biogeography 4:69-84

Finnerty J, Block BA (1995) Evolution of cytochrome b in Scombroidei (Teleostei): molecular insight into billfish (Istiophoridae and Xiphiidae) relationships. Fishery Bulletin 93:78-96

Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork VL (2017) Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (Quercus section Quercus). Genome 60:743-755

Flower BP, Kennett JP (1994) The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. Palaeogeography, Palaeoclimatology, Palaeoecology 108:537-555

Friedman M (2009) Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. Proceedings of the National Academy of Sciences:PNAS. 0808468106

Friedman M (2010) Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. Proceedings of the Royal Society of London B: Biological Sciences:rspb20092177

Fromentin JM, Powers JE (2005) Atlantic bluefin tuna: population dynamics, ecology, fisheries and management. Fish and Fisheries 6:281-306

Gadagkar SR, Rosenberg MS, Kumar S (2005) Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 304:64-74

Galla SJ, Forsdick NJ, Brown L, Hoeppner MP, Knapp M, Maloney RF, Moraga R, Santure AW, Steeves TE (2019) Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. Genes 10:9

Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU (2009) Mitochondrial whims: metabolic rate, longevity and the rate of molecular evolution. Biology Letters 5:413-416

García N, Folk RA, Meerow AW, Chamala S, Gitzendanner MA, de Oliveira RS, Soltis DE, Soltis PS (2017) Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). Molecular Phylogenetics and Evolution 111:231-247

Gaskell BA (1991) Extinction patterns in Paleogene benthic foraminiferal faunas; relationship to climate and sea level. Palaios 6:2-16

Gatesy J, Springer MS (2014) Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution 80:231-266

Gibbs RH, Collette BB (1967) Comparative Anatomy and Systematics of the Tunas. Fishery Bulletin of the Fish and Wildlife Service 66:65

Gille ST (2002) Warming of the Southern Ocean since the 1950s. Science 295:1275-1277

Glenn TC, Bayona-Vasquez NJ, Kieran TJ, Pierson TW, Hoffberg SL, Scott PA, Bentley KE, Finger JW, Watson PR, Louha S (2017) Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1 USD per Sample (3RAD). BioRxiv:205799

Godsil HC (1954) A descriptive study of certain tuna-like fishes. Department of fish game

Goldman KJ (1997) Regulation of body temperature in the white shark, Carcharodon carcharias. Journal of Comparative Physiology B 167:423-429

Goldman KJ, Anderson SD, Latour RJ, Musick JA (2004) Homeothermy in adult salmon sharks, Lamna ditropis. Environmental Biology of Fishes 71:403-411

Graham J (1975) Heat exchange in the yellowfin tuna, Thunnus albacares, and skipjack tuna, Katsuwonus pelamis, and the adaptive significance of elevated body temperatures in scombrid fishes. Fishery Bulletin 73:219-229

Graham J, Koehrn F, Dickson K (1983) Distribution and relative proportions of red muscle in scombrid fishes: consequences of body size and relationships to locomotion and endothermy. Canadian Journal of Zoology 61:2087-2096

Graham JB (1973) Heat Exchnage in the Black Skipjack, and the Blood-Gas Relationship of Warm-Bodied Fishes. Proceedings of the National Academy of Sciences 70:1964-1967

Graham JB, Dickson KA (2000) The evolution of thunniform locomotion and heat conservation in scombrid fishes: new insights based on the morphology of Allothunnus fallai. Zoological Journal of the Linnean Society 129:419-466

Graham JB, Dickson KA (2001) Anatomical and physiological specializations for endothermy. Fish Physiology 19:121-165

Graham JB, Dickson KA (2004) Tuna comparative physiology. Journal of Experimental Biology 207:4015-4024

Grant W, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. Journal of Heredity 89:415-426

Grant WS (1987) Genetic divergence between congeneric Atlantic and Pacific Ocean fishes. Population Genetics and Fishery Management:225-246

Guevara EE, Steiper ME (2014) Molecular phylogenetic analysis of the Papionina using concatenation and species tree methods. Journal of human evolution 66:18-28

Guinet C, Domenici P, De Stephanis R, Barrett-Lennard L, Ford J, Verborgh P (2007) Killer whale predation on bluefin tuna: exploring the hypothesis of the endurance-exhaustion technique. Marine Ecology Progress Series 347:111-119

Gunn JS, Clear NP, Carter TI, Rees AJ, Stanley CA, Farley JH, Kalish JM (2008) Age and growth in southern bluefin tuna, Thunnus maccoyii (Castelnau): direct estimation from otoliths, scales and vertebrae. Fisheries Research 92:207-220

Günther T, Nettelblad C (2019) The presence and impact of reference bias on population genomic studies of prehistoric human populations. PLoS Genetics 15:e1008302

Guo C, Guo Z-H, Li D-Z (2019) Phylogenomic analyses reveal intractable evolutionary history of a temperate bamboo genus (Poaceae: Bambusoideae). Plant Diversity 41:213-219

Guo Y, Ye F, Sheng Q, Clark T, Samuels DC (2014) Three-stage quality control strategies for DNA re-sequencing data. Briefings in Bioinformatics 15:879-889

Hansen TA (1987) Extinction of late Eocene to Oligocene molluscs; relationship to shelf area, temperature changes, and impact events. Palaios 2:69-75

Hasegawa M, Kishino H, Yano T-a (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160-174

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). Proceedings of the National Academy of Sciences 99:11742-11747

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proceedings of the National Academy of Sciences 101:14812-14817

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms. Genetics and Evolution of Aquatic Organisms 122:122-134

Heger A, Ponting CP (2007) Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Research 17:1837-1849

Heino M, Godø OR (2002) Fisheries-induced selection pressures in the context of sustainable fisheries. Bulletin of Marine Science 70:639-656

Heled J, Drummond AJ (2012) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic Biology 61:138-149

Herbert TD, Lawrence KT, Tzanova A, Peterson LC, Caballero-Gill R, Kelly CS (2016) Late Miocene global cooling and the rise of modern ecosystems. Nature Geoscience 9:843-847

Herrera S, Reyes-Herrera PH, Shank TM (2015) Predicting RAD-seq marker numbers across the eukaryotic tree of life. Genome Biology and Evolution 7:3207-3225

Herrera S, Watanabe H, Shank TM (2014) Evolutionary and biogeographical patterns of barnacles from deep-sea hydrothermal vents. Molecular Ecology 24:673-689

Hertel H (1966) Structure, form, movement. Reinhold Publishing Corporation, New York, New York

Heywood VH, Watson RT (1995) Global biodiversity assessment, Vol 1140. Cambridge University Press Cambridge

Hicks SC, Townes FW, Teng M, Irizarry RA (2018) Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics 19:562-578

Hillis DM, Moritz C, Mable BK, Olmstead RG (1996) Molecular systematics, Vol 23. Sinauer Associates Sunderland, MA

Hinke J, Kaplan I, Aydin K, Watters G, Olson R, Kitchell JF (2004) Visualizing the food-web effects of fishing for tunas in the Pacific Ocean. Ecology and Society 9

Ho SY (2014) The changing face of the molecular evolutionary clock. Trends in Ecology & Evolution 29:496-503

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35:518-522

Hoarau G, Boon E, Jongma DN, Ferber S, Palsson J, Van der Veer HW, Rijnsdorp AD, Stam WT, Olsen JL (2005) Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (Pleuronectes platessa L.). Proceedings of the Royal Society B: Biological Sciences 272:497-503

Hodel RG, Chen S, Payton AC, McDaniel SF, Soltis P, Soltis DE (2017) Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Scientific Reports 7:1-14

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. Nature 470:479

Holland KN, Sibert JR (1994) Physiological thermoregulation in bigeye tuna, Thunnus obesus. Environmental Biology of Fishes 40:319-327

Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT (2016) Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Molecular Biology and Evolution 33:1110-1125

Hou Y, Nowak MD, Mirre V, Bjorå CS, Brochmann C, Popp M (2015) Thousands of RAD-seq loci fully resolve the phylogeny of the highly disjunct arctic-alpine genus Diapensia (Diapensiaceae). PLoS One 10

Houghton S (1988) Thermocline control on coccolith diversity and abundance in Recent sediments from the Celtic Sea and English Channel. Marine Geology 83:313-319

Huang H, Knowles LL (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic Biology 65:357-365

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T (2002) The Ensembl genome database project. Nucleic Acids Research 30:38-41

Hunsicker ME, Olson RJ, Essington TE, Maunder MN, Duffy LM, Kitchell JF (2012) Potential for top-down control on tropical tunas based on size structure of predator− prey interactions. Marine Ecology Progress Series 445:263-277

Hurvich CM, Tsai C-L (1989) Regression and time series model selection in small samples. Biometrika 76:297-307

Inagake D, Hirano T (1983) Vertical distribution of the Japanese sardine in relation to temperature and thermocline at the purse seine fishing grounds east of Japan. Bull Japan Soc Sci Fish 49:1533-1539

Ingram T (2011) Speciation along a depth gradient in a marine adaptive radiation. Proceedings of the Royal Society B: Biological Sciences 278:613-618

Ingram T, Hudson AG, Vonlanthen P, Seehausen O (2012) Does water depth or diet divergence predict progress toward ecological speciation in whitefish radiations? Evolutionary Ecology Research:487-502

ISSF ITSS (2018) Status of the world fisheries for tuna: February 2018. ISSF Technical Report 2018–02. Washington DC, US, Washington, D.C., USA

Ito H, Nakajima N, Onuma M, Murayama M (2020) Genetic Diversity and Genetic Structure of the Wild Tsushima Leopard Cat from Genome-Wide Analysis. Animals 10:1375

Jaramillo C, Montes C, Cardona A, Silvestro D, Antonelli A, Bacon CD (2017) Comment (1) on "Formation of the Isthmus of Panama" by O'Dea et al. Science Advances 3:e1602321

Jensen JD, Bachtrog D (2011) Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. Genome Biology and Evolution 3:687-701

Jeong S, Kim J-Y, Jeong S-C, Kang S-T, Moon J-K, Kim N (2017) GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. PLoS One 12:e0181420

Jiang Z, Wang H, Michal JJ, Zhou X, Liu B, Woods LCS, Fuchs RA (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges and future development. International Journal of Biological Sciences 12:100

Joenje H (1989) Genetic toxicology of oxygen. Mutation Research/DNAging 219:193-208

Johnson PL, Slatkin M (2008) Accounting for bias from sequencing error in population genetic estimates. Molecular Biology and Evolution 25:199-206

Jumper Jr GY, Baird RC (1991) Location by olfaction: a model and application to the mating problem in the deep-sea hatchetfish Argyropelecus hemigymnus. The American Naturalist 138:1431-1458

Keller G (1983) Paleoclimatic analyses of middle Eocene through Oligocene planktic foraminiferal faunas. Palaeogeography, Palaeoclimatology, Palaeoecology 43:73-94

Kennett JP, Stott L (1991) Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. Nature 353:225-229

Kikawa S (1966) The distribution of maturing bigeye and yellowfin and an evaluation of their spawning potential in different areas in the tuna longline grounds in the Pacific. Bull Nankai Regional Fish Res Lab 23:131-208

Kishinouye K (1915) A study of the mackerels, cybiids and tunas. US Fish and Wildlife Service

Kitagawa T, Kimura S, Nakata H, Yamada H (2004) Diving behavior of immature, feeding Pacific bluefin tuna (Thunnus thynnus orientalis) in relation to season and area: the East China Sea and the Kuroshio–Oyashio transition region. Fisheries Oceanography 13:161-180

Kitagawa T, Kimura S, Nakata H, Yamada H (2007) Why do young Pacific bluefin tuna repeatedly dive to depths through the thermocline? Fisheries Science 73:98-106

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. Cell 155:27-38

Koched W, Hattour A, Alemany F, Garcia A, Said K (2013) Spatial distribution of tuna larvae in the Gulf of Gabes (Eastern Mediterranean) in relation with environmental parameters. Mediterranean Marine Science 14:5-14

Kohn MH, Wayne RK (1997) Facts from feces revisited. Trends in Ecology & Evolution 12:223-227

Korsmeyer KE, Dewar H (2001) Tuna metabolism and energetics. Fish Physiology 19:35-78

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453-4455

Kumar G, Kocour M (2015) Population genetic structure of tunas inferred from molecular markers: a review. Reviews in Fisheries Science & Aquaculture 23:72-89

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549

Laehnemann D, Borkhardt A, McHardy AC (2016) Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. Briefings in Bioinformatics 17:154-179

Laidre KL, Heide-Jørgensen MP, Dietz R, Hobbs RC, Jørgensen OA (2003) Deep-diving by narwhals Monodon monoceros: differences in foraging behavior between wintering areas? Marine Ecology Progress Series 261:269-281

Laird CD, McConaughy BL, McCarthy BJ (1969) Rate of fixation of nucleotide substitutions in evolution. Nature 224:149-154

Lalu XC, Kosen JD, Tjakrawidjaja AH, Kusumah RV, Sadhotomo B, Pouyaud L, Slembrouck J, Paradis E (2010) Mitochondrial genomic divergence in coelacanths (Latimeria): slow rate of evolution or recent speciation? Marine Biology 157:2253-2262

Lanfear R, Thomas JA, Welch JJ, Brey T, Bromham L (2007) Metabolic rate does not calibrate the molecular clock. Proceedings of the National Academy of Sciences 104:15388-15393

Lanier HC, Knowles LL (2012) Is recombination a problem for species-tree analyses? Systematic Biology 61:691-701

Le Gall J-Y, Laurec A, Chardy P (1976) Mise en évidence des relations phénotypiques et phylogénétiques à l'intérieur du genre Thunnus par une analyse multicritère. Bull Mus Nat Hist Nat Sci 3:1349-1368

Lea AJ, Tung J, Archie EA, Alberts SC (2017) Developmental plasticity: bridging research in evolution and human health. Evolution, Medicine, and Public Health 2017:162-175

Leache AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW (2015) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biology and Evolution 7:706-719

Lehodey P (2008) Tunas in hot water: forecasts of population trends for two species of tuna under a scenario of Climate Change. Effects of Climate Change on the World's Oceans. Effects of Climate Change on the World's Oceans. Gijon, Spain, Gijon, Spain

Leifer I, Judd A (2002) Oceanic methane layers: the hydrocarbon seep bubble deposition hypothesis. Terra Nova 14:417-424

Leis J, Trnski T, Harmelin-Vivien M, Renon J-P, Dufour V, El Moudni M, Galzin R (1991) High concentrations of tuna larvae (Pisces: Scombridae) in near-reef waters of French Polynesia (Society and Tuamotu Islands). Bulletin of Marine Science 48:150-158

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99-121

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26:589-595

Li W-H, Ellsworth DL, Krushkal J, Chang BH-J, Hewett-Emmett D (1996) Rates of nucleotide substitution in primates and rodents and the generation–time effect hypothesis. Molecular Phylogenetics and Evolution 5:182-187

Lighthill MJ (1970) Aquatic animal propulsion of high hydromechanical efficiency. Journal of Fluid Mechanics 44:265-301

Literman R, Burrett A, Bista B, Valenzuela N (2018) Putative independent evolutionary reversals from genotypic to temperature-dependent sex determination are associated with accelerated evolution of sex-determining genes in turtles. Journal of Molecular Evolution 86:11-26

Liu Z, Pagani M, Zinniker D, DeConto R, Huber M, Brinkhuis H, Shah SR, Leckie RM, Pearson A (2009) Global cooling during the Eocene-Oligocene climate transition. Science 323:1187-1190

Llopiz JK, Richardson DE, Shiroza A, Smith SL, Cowen RK (2010) Distinctions in the diets and distributions of larval tunas and the important role of appendicularians. Limnology and Oceanography 55:983-996

Lloret J (2010) Human health benefits supplied by Mediterranean marine biodiversity. Marine Pollution Bulletin 60:1640-1646

Lowe T, Brill R, Cousins K (2000) Blood oxygen-binding characteristics of bigeye tuna (Thunnus obesus), a high-energy-demand teleost that is tolerant of low ambient oxygen. Marine Biology 136:1087-1098

Lu G, Keller G (1993) The Paleocene-Eocene transition in the Antarctic Indian Ocean: inference from planktic foraminifera. Marine Micropaleontology 21:101-142

Lyman JM, Good SA, Gouretski VV, Ishii M, Johnson GC, Palmer MD, Smith DM, Willis JK (2010) Robust warming of the global upper ocean. Nature 465:334

Lythgoe JN (1988) Light and vision in the aquatic environment. Sensory Biology of Aquatic Animals. Springer

Macfadyen G, Defaux V (2016) Estimate of Global Sales Values From Tuna Fisheries - Phase 2 Report. Estimate of Global Sales Values From Tuna Fisheries, UK

Maddison WP (1997) Gene trees in species trees. Systematic Biology 46:523-536

Magalhaes I, Lundsgaard-Hansen B, Mwaiko S, Seehausen O (2012) Evolutionary divergence in replicate pairs of ecotypes of Lake Victoria cichlid fish. Evolutionary Ecology Research 14:381-401

Magnusson J, Hoar W, Randall D (1978) Locomotion by Scombroid fishes: hy-dromechanics, morphology, and behaviour. Fish Physiology 7:240-315

Majewski W, Bohaty SM (2010) Surface-water cooling and salinity decrease during the Middle Miocene climate transition at Southern Ocean ODP Site 747 (Kerguelen Plateau). Marine Micropaleontology 74:1-14

Manthey JD, Campillo LC, Burns KJ, Moyle RG (2016) Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: Piranga). Systematic Biology 65:640-650

Margulies D (1997) Development of the visual system and inferred performance capabilities of larval and early juvenile scombrids. Marine and Freshwater Behaviour and Physiology 30:75-98

Marin J, Hedges SB (2018) Undersampling genomes has biased time and rate estimates throughout the tree of life. Molecular Biology and Evolution 35:2077-2084

Marroni F, Pinosio S, Di Centa E, Jurman I, Boerjan W, Felice N, Cattonaro F, Morgante M (2011) Large-scale detection of rare variants via pooled multiplexed next-generation sequencing: towards next-generation Ecotilling. The Plant Journal 67:736-745

Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences 90:4087-4091

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17:10-12

Masta SE (2000) Phylogeography of the jumping spider Habronattus pugillis (Araneae: Salticidae): recent vicariance of sky island populations? Evolution 54:1699-1711

Matsumoto WM, Ahlstrom EH, Jones S (1972) On the clarification of larval tuna identification. Fish B-NOAA 70:1-12

Matsumoto WM, Skillman RA, Dizon AE (1984) Synopsis of biological data on skipjack tuna, Katsuwonus pelamis. National Oceanic and Atmospheric Administration, National Marine Fisheries Services

McFall-Ngai M, Montgomery MK (1990) The anatomy and morphology of the adult bacterial light organ of Euprymna scolopes Berry (Cephalopoda: Sepiolidae). The Biological Bulletin 179:332-339

McKinney M, McNamara K, Carter B, Donovan S (1992) Evolution of Paleogene echinoids: a global and regional view. Eocene-Oligocene Climatic and Biotic Evolution:349-367

McLellan T (1977) Feeding strategies of the macrourids. Deep Sea Research 24:1019-1036

McVay JD, Carstens BC (2013) Phylogenetic model choice: justifying a species tree or concatenation analysis. Journal of Phylogenetics & Evolutionary Biology 1

McVay JD, Hipp AL, Manos PS (2017) A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proceedings of the Royal Society B: Biological Sciences 284:20170300

Ménard F, Labrune C, Shin Y-J, Asine A-S, Bard F-X (2006) Opportunistic predation in tuna: a size-based approach. Marine Ecology Progress Series 323:223-231

Meszéna G, Hendry AP (2012) Introduction to niche theory and speciation. Evolutionary Ecology Research 14:361-363

Meyer A (1993) Trophic polymorphisms in cichlid fish: do they represent intermediate steps during sympatric speciation and explain their rapid adaptive radiation. Trends in Ichthyology 7:257-266

Meyer BS, Matschiner M, Salzburger W (2016) Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Systematic Biology 66:531-550

Mihaescu R, Levy D, Pachter L (2009) Why neighbor-joining works. Algorithmica 54:1-24

Miller MA, Pfeiffer W, Schwartz T Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proc Gateway Computing Environments Workshop (GCE), 2010. Ieee

Mindell DP, Knight A, Baer C, Huddleston CJ (1996) Slow rates of molecular evolution in birds and the metabolic rate and body temperature hypotheses. Molecular Biology and Evolution 13:422-426

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530-1534

Momigliano P, Jokinen H, Fraimout A, Florin A-B, Norkko A, Merilä J (2017) Extraordinarily rapid speciation in a marine fish. Proceedings of the National Academy of Sciences 114:6074-6079

Monsch KA (2000) The phylogeny of the scombroid fishes. University of Bristol,

Monsch KA (2006) A revision of scombrid fishes (Scombroidei, Perciformes) from the Middle Eocene of Monte Bolca, Italy. Palaeontology 49:873-888

Montes C, Cardona A, Jaramillo C, Pardo A, Silva J, Valencia V, Ayala C, Pérez-Angel L, Rodriguez-Parra L, Ramirez V (2015) Middle Miocene closure of the Central American seaway. Science 348:226-229

Moore TA (2014) Trophic Dynamics and Feeding Ecology of the Southeast Florida Coastal Pelagic Fish Community.

Mora C, Tittensor DP, Adl S, Simpson AG, Worm B (2011) How many species are there on Earth and in the ocean? PLoS Biology 9:e1001127

Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J (2016) The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biology 17:1-9

Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. Molecular Biology and Evolution 25:120-130

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22:2841-2847

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG (2017) Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources 17:1136-1147

Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbrink FT, Friedman M (2013) Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proceedings of the National Academy of Sciences 110:12738-12743

Near TJ, MacGuigan DJ, Parker E, Struthers CD, Jones CD, Dornburg A (2018) Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. Molecular Phylogenetics and Evolution 129:268-279

Nesse RM (2008) The importance of evolution for medicine. Evolutionary Medicine:416-432

Ng EY, Garg KM, Low GW, Chattopadhyay B, Oh RR, Lee JG, Rheindt FE (2017) Conservation genomics identifies impact of trade in a threatened songbird. Biological Conservation 214:101-108

Nilsson D-E, Warrant EJ, Johnsen S, Hanlon R, Shashar N (2012) A unique advantage for giant eyes in giant squid. Current Biology 22:683-688

Nishikawa Y (1985) Average distribution of larvae of oceanic species of scombroid fishes, 1956-1981. NRIFSF S-Ser 12:1-99

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44:D733-D745

O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, Collins LS, De Queiroz A, Farris DW, Norris RD (2016) Formation of the Isthmus of Panama. Science Advances 2:e1600883

Olson ND, Lund SP, Colman RE, Foster JT, Sahl JW, Schupp JM, Keim P, Morrow JB, Salit ML, Zook JM (2015) Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Frontiers in Genetics 6:235

Pagel M, Meade A (2008) Modelling heterotachy in phylogenetic inference by reversible-jump Markov chain Monte Carlo. Philosophical Transactions of the Royal Society B: Biological Sciences 363:3955-3964

Pagès M, Bazin E, Galan M, Chaval Y, Claude J, Herbreteau V, Michaux J, Piry S, Morand S, Cosson JF (2013) Cytonuclear discordance among Southeast Asian black rats (Rattus rattus complex). Molecular Ecology 22:1019-1034

Palumbi S (2009) Speciation and the evolution of gamete recognition genes: pattern and process. Heredity 102:66

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289-290

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528

Patwardhan A, Ray S, Roy A (2014) Molecular markers in phylogenetic studies-a review. Journal of Phylogenetics & Evolutionary Biology 2014

Pauers MJ, Mckinnon JS (2012) Sexual selection on color and behavior within and between cichlid populations: Implications for speciation. Current Zoology 58:475-483

Pavlov D, Sadkovskii R, Kostin V, Lupandin A (2000) Experimental study of young fish distribution and behaviour under combined influence of baro-, photo-and thermo-gradients. Journal of Fish Biology 57:69-81

Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. PeerJ 4:e1813

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:e37135

Pethybridge H, Choy CA, Logan JM, Allain V, Lorrain A, Bodin N, Somes CJ, Young J, Ménard F, Langlais C (2018) A global meta-analysis of marine predator nitrogen stable isotopes: Relationships between trophic structure and environmental conditions. Global Ecology and Biogeography

Pillai N (2009) An Overview of World tuna fisheries. Seafood Export Journal 41:21-29

Piovesan A, Antonaros F, Vitale L, Strippoli P, Pelleri MC, Caracausi M (2019) Human protein-coding genes and gene feature statistics in 2019. BMC Research Notes 12:315

Pörtner H-O (2008) Ecosystem effects of ocean acidification in times of ocean warming: a physiologist's view. Marine Ecology Progress Series 373:203-217

Posada D (2008) jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25:1253-1256

Prothero DR (1994) The late Eocene-Oligocene extinctions. Annual Review of Earth and Planetary Sciences 22:145-165

Pujolar J, Roldán M, Pla C (2003) Genetic analysis of tuna populations, Thunnus thynnus thynnus and T. alalunga. Marine Biology 143:613-621

Purcell JE, Sturdevant MV, Galt CP (2005) A review of appendicularians as prey of invertebrate and fish predators. Response of Marine Ecosystems to Global Changes: Ecological Impact of Appendicularians:359-435

Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014) Demystifying the RAD fad. Molecular Ecology 23:5937-5942

Qiu F, Kitchen A, Burleigh JG, Miyamoto MM (2014) Scombroid fishes provide novel insights into the trait/rate associations of molecular evolution. Journal of Molecular Evolution 78:338-348

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842

Rancilhac L, Goudarzi F, Gehara M, Hemami M-R, Elmer KR, Vences M, Steinfarz S (2019) Phylogeny and species delimitation of near Eastern Neurergus newts (Salamandridae) based on genome-wide RADseq data analysis. Molecular Phylogenetics and Evolution 133:189-197

Razkin O, Sonet G, Breugelmans K, Madeira MJ, Gómez-Moliner BJ, Backeljau T (2016) Species limits, interspecific hybridization and phylogeny in the cryptic land snail

complex Pyramidula: the power of RADseq data. Molecular Phylogenetics and Evolution 101:267-278

Reglero P, Tittensor D, Álvarez-Berastegui D, Aparicio-González A, Worm B (2014) Worldwide distributions of tuna larvae: revisiting hypotheses on environmental requirements for spawning habitats. Marine Ecology Progress Series 501:207-224

Reglero P, Urtizberea A, Torres AP, Alemany F, Fiksen Ø (2011) Cannibalism among size classes of larvae may be a substantial mortality component in tuna. Marine Ecology Progress Series 433:205-219

Richards WJ, Dove GR (1971) Internal development of young tunas of the genera Katsuwonus, Euthynnus, Auxis, and Thunnus (Pisces, Scombridae). Copeia:72-78

Richter C, Park J-W, Ames BN (1988) Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proceedings of the National Academy of Sciences 85:6465-6467

Righi T, Splendiani A, Fioravanti T, Casoni E, Gioacchini G, Carnevali O, Caputo Barucchi V (2020) Loss of Mitochondrial Genetic Diversity in Overexploited Mediterranean Swordfish (Xiphias gladius, 1759) Population. Diversity 12:170

Rivera-Colón AG, Rochette NC, Catchen JM (2020) Simulation with RADinitio improves RADseq experimental design and sheds light on sources of missing data. Molecular Ecology Resources

Rivers DM, Darwell CT, Althoff DM (2016) Phylogenetic analysis of RAD-seq data: examining the influence of gene genealogy conflict on analysis of concatenated data. Cladistics 32:672-681

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical Methods for Paired-end Sequencing Improve RADseq-based Population Genomics. Molecular Ecology:615385

Roden C, Raine R (1994) Phytoplankton blooms and a coastal thermocline boundary along the west coast of Ireland. Estuarine, Coastal and Shelf Science 39:511-526

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539-542

Rooker JR, Alvarado Bremer JR, Block BA, Dewar H, De Metrio G, Corriero A, Kraus RT, Prince ED, Rodríguez-Marín E, Secor DH (2007) Life history and stock structure of Atlantic bluefin tuna (Thunnus thynnus). Reviews in Fisheries Science 15:265-310

Ropert-Coudert Y, Chiaradia A, Kato A (2006) An exceptionally deep dive by a Little Penguin Eudyptula minor. Marine Ornithology 34:71-74

Rubin BE, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. PLoS One 7

Ruffalo M, LaFramboise T, Koyutürk M (2011) Comparative analysis of algorithms for next-generation sequencing read alignment. Bioinformatics 27:2790-2796

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491

Saikia U, Sharma N, Das A (2008) What is a species? An endless debate. Resonance 13:1049-1064

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425

Sambrook J, Fritsch E, Maniatis T (1989) Extraction with Phenol: Chloroform. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press

Sanderson MJ, Hufford L (1996) Homoplasy: the recurrence of similarity in evolution. Elsevier

Santini F, Carnevale G, Sorenson L (2013) First molecular scombrid timetree (Percomorpha: Scombridae) shows recent radiation of tunas following invasion of pelagic habitat. Italian Journal of Zoology 80:210-221

Santos JC (2012) Fast molecular evolution associated with high active metabolic rates in poison frogs. Molecular Biology and Evolution 29:2001-2018

Sardenne F, Bodin N, Chassot E, Amiel A, Fouche E, Degroote M, Hollanda S, Pethybridge H, Lebreton B, Guillou G (2016) Trophic niches of sympatric tropical tuna in the Western Indian Ocean inferred by stable isotopes and neutral fatty acids. Progress in Oceanography 146:75-88

Satoh K, Tanaka Y, Iwahashi M (2008) Variations in the instantaneous mortality rate between larval patches of Pacific bluefin tuna Thunnus orientalis in the northwestern Pacific Ocean. Fisheries Research 89:248-256

Schaefer KM (2001) Reproductive biology of tunas. Fish Physiology 19:225-270

Schaefer KM, Fuller DW (2010) Vertical movements, behavior, and habitat of bigeye tuna (Thunnus obesus) in the equatorial eastern Pacific Ocean, ascertained from archival tag data. Marine Biology 157:2625-2642

Schaefer KM, Fuller DW, Block BA (2009) Vertical movements and habitat utilization of skipjack (Katsuwonus pelamis), yellowfin (Thunnus albacares), and bigeye (Thunnus obesus) tunas in the equatorial eastern Pacific Ocean, ascertained through archival tag data. Tagging and Tracking of Marine Animals with Electronic Devices. Springer

Schaefer MB, Marr JC (1948) Juvenile Euthynnus lineatus and Auxis thazard from the Pacific Ocean off Central America. Pacific Sci 2:262-271

Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnology Journal 15:149-161

Schliep K, Potts AA, Morrison DA, Grimm GW (2016) Intertwining phylogenetic trees and networks. PeerJ Preprints

Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27:592-593

Schuster SC (2007) Next-generation sequencing transforms today's biology. Nature Methods 5:16

Shapiro BJ, Leducq J-B, Mallet J (2016) What is speciation? PLoS Genetics 12:e1005860

Sharp G, Dizon AE (1972) The physiological ecology of tunas. Elsevier

Sharp GD, Pirages S (1978) The distribution of red and white swimming muscles, their biochemistry, and the biochemical phylogeny of selected scombrid fishes. The Physiological Ecology of Tunas:41-78

Shevenell AE, Kennett JP, Lea DW (2004) Middle Miocene southern ocean cooling and Antarctic cryosphere expansion. Science 305:1766-1770

Shigenaga MK, Gimeno CJ, Ames BN (1989) Urinary 8-hydroxy-2'-deoxyguanosine as a biological marker of in vivo oxidative DNA damage. Proceedings of the National Academy of Sciences 86:9697-9701

Shimose T, Tanabe T, Chen K-S, Hsu C-C (2009) Age determination and growth of Pacific bluefin tuna, Thunnus orientalis, off Japan and Taiwan. Fisheries Research 100:134-139

Slatko BE, Gardner AF, Ausubel FM (2018) Overview of next-generation sequencing technologies. Current Protocols in Molecular Biology 122:e59

Sluijs A, Bijl PK, Schouten S, Röhl U, Reichart G-J, Brinkhuis H (2011) Southern ocean warming, sea level and hydrological change during the Paleocene-Eocene thermal maximum. Climate of the Past 7:47-61

Smith JM (1978) Optimization theory in evolution. Annual Review of Ecology and Systematics 9:31-56

Smith KL, Harmon LJ, Shoo LP, Melville J (2011) Evidence of constrained phenotypic evolution in a cryptic species complex of agamid lizards. Evolution: International Journal of Organic Evolution 65:976-992

Somero G (2010) The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine ‘winners’ and ‘losers’. Journal of Experimental Biology 213:912-920

Somiya H, Takei S, Mitani I (2000) Guanine and its retinal distribution in the tapetum of the bigeye tuna, Thunnus obesus. Ichthyological Research 47:367-372

Spear LB, Ballance LT, Ainley DG (2001) Response of seabirds to thermal boundaries in the tropical Pacific: the thermocline versus the Equatorial Front. Marine Ecology Progress Series 219:275-289

Stackebrandt E, Ludwig W (1994) The importance of using outgroup reference organisms in phylogenetic studies: the Atopobium case. Systematic and Applied Microbiology 17:39-43

Strasburg DW (1959) An instance of natural mass mortality of larval frigate mackerel in the Hawaiian Islands. ICES Journal of Marine Science 24:255-263

Sturmbauer C, Meyer A (1992) Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes. Nature 358:578-581

Suda A, Nishiki I, Iwasaki Y, Matsuura A, Akita T, Suzuki N, Fujiwara A (2019) Improvement of the Pacific bluefin tuna (Thunnus orientalis) reference genome and development of male-specific DNA markers. Scientific Reports 9:1-12

Sugiura N (1978) Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. Communications in Statistics-Theory and Methods 7:13-26

Sunday JM, Bates AE, Dulvy NK (2012) Thermal tolerance and the global redistribution of animals. Nature Climate Change 2:686-690

Suzuki Z (1994) A review of the biology and fisheries for yellowfin tuna, Thunnus albacares, in the western and central Pacific Ocean. Interactions of Pacific Tuna Fisheries Edited by RS Shomura, J Majkowski, and S Langi FAO Fish Tech Pap 336:108-137

Takagi M, Okamura T, Chow S, Taniguchi N (1999) PCR primers for microsatellite loci in tuna species of the genus Thunnus and its application for population genetic study. Fisheries Science 65:571-576

Takahashi T, Nagano AJ, Kawaguchi L, Onikura N, Nakajima J, Miyake T, Suzuki N, Kanoh Y, Tsuruta T, Tanimoto T (2020) A ddRAD-based population genetics and phylogenetics of an endangered freshwater fish from Japan. Conservation Genetics:1-12

Tang Q, Liu H, Mayden R, Xiong B (2006) Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 39:347-357

Taub MA, Bravo HC, Irizarry RA (2010) Overcoming bias and systematic errors in next generation sequencing data. Genome Medicine 2:87

Taylor RE (2008) Tedanolide and the evolution of polyketide inhibitors of eukaryotic protein synthesis. Natural Product Reports 25:854-861

Team RC (2013) R: A language and environment for statistical computing.

Thiergart T, Landan G, Martin WF (2014) Concatenated alignments and the case of the disappearing tree. BMC Evolutionary Biology 14:1-12

Thomas JA, Welch JJ, Lanfear R, Bromham L (2010) A generation time effect on the rate of molecular evolution in invertebrates. Molecular Biology and Evolution 27:1173-1180

Thompson AW, Betancur-R R, López-Fernández H, Ortí G (2014) A time-calibrated, multi-locus phylogeny of piranhas and pacus (Characiformes: Serrasalmidae) and a comparison of species tree methods. Molecular Phylogenetics and Evolution 81:242-257

Thompson JN (1998) Rapid evolution as an ecological process. Trends in Ecology & Evolution 13:329-332

Tidd A, Blanchard JL, Kell L, Watson RA (2018) Predicting global tuna vulnerabilities with spatial, economic, biological and climatic considerations. Scientific Reports 8:10572

Tim L, Davis GPJ, Jock WY (1990) Diel patterns of vertical distribution in larvae of southern bluefin Thunnus maccoyii, and other tuna in the East Indian Ocean. Marine Ecology Progress Series 59:63-74

Tollis M, Schiffman JD, Boddy AM (2017) Evolution of cancer suppression as revealed by mammalian comparative genomics. Current Opinion in Genetics & Development 42:40-47

Tonini J, Moore A, Stern D, Shcheglovitova M, Ortí G (2015) Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Currents 7

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, Bird CE (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1:e203

Torkamaneh D, Laroche J, Belzile F (2016) Genome-wide SNP calling from genotyping by sequencing (GBS) data: a comparison of seven pipelines and two sequencing technologies. PLoS One 11:e0161333

Townsend TM, Mulcahy DG, Noonan BP, Sites Jr JW, Kuczynski CA, Wiens JJ, Reeder TW (2011) Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. Molecular Phylogenetics and Evolution 61:363-380

Tripp EA, Tsai YHE, Zhuang Y, Dexter KG (2017) RAD seq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. Ecology and Evolution 7:7920-7936

Tseng M-C, Jean C-T, Smith PJ, Hung Y-H (2012) Interspecific and intraspecific genetic diversity of Thunnus species. Analysis of Genetic Variation in Animals. InTech

Tzanova A, Herbert TD, Peterson L (2015) Cooling Mediterranean Sea surface temperatures during the Late Miocene provide a climate context for evolutionary transitions in Africa and Eurasia. Earth and Planetary Science Letters 419:71-80

Ueyanagi S (1969) Observations on the distribution of tuna larva in the Indo-Pacific Ocean with emphasis on the delineation of spawning areas of albacore, Thunnus alalunga. Bull Far Seas Fish Res Lab 2:177-219

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics 43:11.10. 11-11.10. 33

van Eeden R, Reid T, Ryan PG, Pichegru L (2016) Fine-scale foraging cues for African penguins in a highly variable marine environment. Marine Ecology Progress Series 543:257-271

van Oppen MJ, Oliver JK, Putnam HM, Gates RD (2015) Building coral reef resilience through assisted evolution. Proceedings of the National Academy of Sciences 112:2307-2313

Varela JL, Larrañaga A, Medina A (2011) Prey-muscle carbon and nitrogen stable-isotope discrimination factors in Atlantic bluefin tuna (Thunnus thynnus). Journal of Experimental Marine Biology and Ecology 406:21-28

Vendrami DL, Telesca L, Weigand H, Weiss M, Fawcett K, Lehman K, Clark MS, Leese F, McMinn C, Moore H (2017) RAD sequencing resolves fine-scale population structure in a benthic invertebrate: implications for understanding phenotypic plasticity. Royal Society Open Science 4:160548

Wachowiak W, Zaborowska J, Łabiszak B, Perry A, Zucca GM, González-Martínez SC, Cavers S (2018) Molecular signatures of divergence and selection in closely related pine taxa. Tree Genetics & Genomes 14:83

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the L ake V ictoria cichlid adaptive radiation. Molecular Ecology 22:787-798

Wang H-P, Guo L, Shepherd B, Sepulveda Villet OJ, Zhang D (2019) Development of a genomic resource and identification of nucleotide diversity of yellow perch by rad sequencing. Frontiers in Genetics 10:992

Wang J-T, Lin Y-Y, Chang S-Y, Yeh S-H, Hu B-H, Chen P-J, Chang S-C (2020a) The role of phylogenetic analysis in clarifying the infection source of a COVID-19 patient. Journal of Infection 81:147-178

Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T (2020b) treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Molecular Biology and Evolution 37:599-603

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods 9:808-810

Wang X, Zhao L, Eaton D, Li D, Guo Z (2013) Identification of SNP markers for inferring phylogeny in temperate bamboos (P oaceae: B ambusoideae) using RAD sequencing. Molecular Ecology Resources 13:938-945

Ward R (1995) Population genetics of tunas. Journal of Fish Biology 47:259-280

Ward RD, Elliott NG, Grewe PM (1995) Allozyme and mitochondrial DNA separation of Pacific northern bluefin tuna, Thunnus thynnus orientalis (Temminck and Schlegel), from southern bluefin tuna, Thunnus maccoyii (Castelnau). Marine and Freshwater Research 46:921-930

Wardle C, Videler J, Arimoto T, Franco J, He P (1989) The muscle twitch and the maximum swimming speed of giant bluefin tuna, Thunnus thynnus L. Journal of Fish Biology 35:129-137

Warnow T (2015) Concatenation analyses in the presence of incomplete lineage sorting. PLoS Currents 7

Wegner NC, Sepulveda CA, Bull KB, Graham JB (2010) Gill morphometrics in relation to gas transfer and ram ventilation in high-energy demand teleosts: Scombrids and billfishes. Journal of Morphology 271:36-49

Wenne R, Boudry P, Hemmer-Hansen J, Lubieniecki KP, Was A, Kause A (2007) What role for genomics in fisheries management and aquaculture? Aquatic Living Resources 20:241-255

Werner T, Buchholz F (2013) Diel vertical migration behaviour in Euphausiids of the northern Benguela current: seasonal adaptations to food availability and strong gradients of temperature and oxygen. Journal of Plankton Research 35:792-812

Westneat MW, Wainwright SA (2001) Mechanical design for swimming: muscle, tendon, and bone. Fish Physiology 19:271-311

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer

Willing E-M, Dreyer C, Van Oosterhout C (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. PLoS One 7:e42649

Woolfit M, Bromham L (2005) Population size and molecular evolution on islands. Proceedings of the Royal Society B: Biological Sciences 272:2277-2282

Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proceedings of the National Academy of Sciences 82:1741-1745

Xi Z, Liu L, Davis CC (2015) Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Molecular Phylogenetics and Evolution 92:63-71

Xi Z, Liu L, Rest JS, Davis CC (2014) Coalescent versus concatenation methods and the placement of Amborella as sister to water lilies. Systematic Biology 63:919-932

Yang S, Zhang B, Zhang H, Zhang S, Fan X, Hua C, Fan W (2019) Bigeye tuna fishing ground in relation to thermocline in the Western and Central Pacific Ocean using Argo data. Indian Journal of Geo Marine Sciences

Yoder JB, Briskine R, Mudge J, Farmer A, Paape T, Steele K, Weiblen GD, Bharti AK, Zhou P, May GD (2013) Phylogenetic signal variation in the genomes of Medicago (Fabaceae). Systematic Biology 62:424-438

Young JW, Davis TL (1990) Feeding ecology of larvae of southern bluefin, albacore and skipjack tunas (Pisces: Scombridae) in the eastern Indian Ocean. Marine Ecology Progress Series:17-29

Yu G, Lam TT-Y, Zhu H, Guan Y (2018) Two methods for mapping and visualizing associated data on phylogeny using ggtree. Molecular Biology and Evolution 35:3041-3043

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28-36

Zachos JC, Röhl U, Schellenberg SA, Sluijs A, Hodell DA, Kelly DC, Thomas E, Nicolo M, Raffi I, Lourens LJ (2005) Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. Science 308:1611-1615

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311-1320

Zhang J, Chiodini R, Badr A, Zhang G (2011) The impact of next-generation sequencing on genomics. Journal of Genetics and Genomics 38:95-109

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTERS II AND III

Methodology for general ddRAD methods from the sequencing center

One hundred micrograms of DNA per sample in 96 well plates were digested in a final volume of 25µl in 1X NEB Cut Smart Buffer and 100 U each ENZYME 1 and ENZYME 2 (NEB) at 37°C for 4 hours. Following a 20 min 80°C enzyme inactivation, samples were held at 12°C until ligation. To each 25µl digest was added 3.5 µl 10X Ligase buffer (NEB), 0.5 µl T4 DNA Ligase (NEB) and adapters containing 1 of 48 unique barcodes and Illumina-compatible P5 sequences coupled to an ENZYME 1 overhang and Illumina-compatible P7 sequences coupled to the ENZYME 2 overhang. Plates were incubated 8 hours at 16°C and heat inactivated at 80°C for 20 min. Pools of no more than 48 samples were combined and EDTA was added to a final concentration of 25µM. One tenth volume of 3M NaAc, pH 5.2 and two volumes of 100% ethanol were added and pools were placed at -20°C for 1 hour before spinning at high speed for 10 min in a bench top microfuge. Pellets were washed twice in 1 ml freshly made 70% ethanol and resuspended in 200µl of EB. Samples were purified with Qiagen PCR Purification columns and eluted in 2 X 50 µl EB for a total of 100 µl. One volume of AMPure XP beads were added to the elutant and DNA purified as per the manufacturers protocol and eluted in 35 µl EB. Thirty µl of each pool containing between 1.9 – 2.2 µg DNA was subjected to Pippin Prep size selection on a 2% dye-free agarose gel with internal size markers aiming for 300-600 bp inserts. Recovered samples were cleaned with 1X AMPure XP beads and quantified on a DeNovix spectrophotometer. One hundred and fifty ng of each pool was then subjected to a pre-selection PCR (PreCR) in which a biotinylated forward primer and unique indexed reverse primers were used to amplify and tag desired DNA fragments. Reactions (200µl total) contained 200 nM dNTPs, biotinylated forward and two P7-index primers per pool and 4 units Q5 Hi-Fidelity Taq (NEB) and were split into 2 X 100 µl volumes for thermocycling. Following an initial denaturation at 98°C for 30 sec, samples were subjected to 15 cycles of 98°C for 10 sec, 72°C for 30 sec then a final elongation at 72°C for 5 min and held at 4°C. PCR products were cleaned up in QIagen PCR purification columns then 1X AMPure XP beads and quantified as before. Removal of nondesirable fragments (P5 to P5 and P7 to P7 ligated products) was achieved with Dynabeads M-270 Streptavidin coupled magnetic beads (ThermoFisher). Briefly, 50 µl of beads per sample were captured and washed twice with 1X Bead Washing Buffer (1X BWB, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl). Beads were resuspended in 100µl 2X BWB and mixed with 2000 ng of PreCR product in 100 µl EB. After 20 minutes at RT, beads were captured and washed three times in 200 µl 1X BWB, twice in 200 µl water and once in 100 µl 1X SSC. Beads were then resuspended in 50 µl 1X SSC and heated at 98°C for 5 min and placed on a magnet and supernatant removed as soon as possible. This elution was repeated and the final supernatants were cleaned up with Qiagen PCR columns. The eluted ssDNA was DeNovix quantified, and diluted to 1 ng/µl with EB. A final PCR was performed on 10 ng of input DNA using P5 and P7 primers in a 50 µl reaction as described above but with only 8 cycles. Final PCR products were purified with 1X AMPure XP beads, quantified and assessed for quality on a Fragment Analyzer (Advanced Analytics).

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER III

Table B-1: DNA extraction size and concentration parameters for sample Talb066 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.3207	1725.716	186.91
TIC (ng/uL)	0			
TIM (nmole/L)	0			
Total Concentration (ng/uL)	3.0641			
Final Concentration (ng/uL)	41.667			
GQN	0			

Table B-2: DNA extraction size and concentration parameters for sample Talb069 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.3319	1740.411	1556.7
TIC (ng/uL)	0			
TIM (nmole/L)	0			
Total Concentration (ng/uL)	200.52			
Final Concentration (ng/uL)	41.667			
GQN	0.3			

Table B-3: DNA extraction size and concentration parameters for sample Talb077 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.2554	1640.428	290.47
TIC (ng/uL)	0			
TIM (nmole/L)	0			
Total Concentration (ng/uL)	47.329			
Final Concentration (ng/uL)	41.667			
GQN	1.7			

Table B-4: DNA extraction size and concentration parameters for sample Tobe0003 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.8638	2435.379	315.96
Peak 2	11418	71.4974	10.309	70.72
TIC (ng/uL)	71.497			
TIM (nmole/L)	10.309			
Total Concentration (ng/uL)	90.152			
Final Concentration (ng/uL)	41.667			
GQN	4.4			

Table B-5: DNA extraction size and concentration parameters for sample Tobe0007 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.8638	2435.379	315.96
Peak 2	11418	71.4974	10.309	70.72
TIC (ng/uL)	71.497			
TIM (nmole/L)	10.309			
Total Concentration (ng/uL)	90.152			
Final Concentration (ng/uL)	41.667			
GQN	4.4			

Table B-6: DNA extraction size and concentration parameters for sample Tobe0008 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	2.8141	3677.162	361.53
Peak 2	5781	31.4906	8.968	65.72
TIC (ng/uL)	31.491			
TIM (nmole/L)	8.968			
Total Concentration (ng/uL)	41.944			
Final Concentration (ng/uL)	41.667			
GQN	1.4			

Table B-7: DNA extraction size and concentration parameters for sample Tobe0012 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	0.865	1130.313	391.87
Peak 2	7640	140.8642	30.354	61.41
TIC (ng/uL)	140.86			
TIM (nmole/L)	30.354			
Total Concentration (ng/uL)	154.08			
Final Concentration (ng/uL)	41.667			
GQN	2.3			

Table B-8: DNA extraction size and concentration parameters for sample Tatl0028 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.2554	1640.428	290.47
TIC (ng/uL)	0			
TIM (nmole/L)	0			
Total Concentration (ng/uL)	47.329			
Final Concentration (ng/uL)	41.667			
GQN	1.7			

Table B-9: DNA extraction size and concentration parameters for sample Tala0277 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	5.9247	7741.731	353.44
TIC (ng/uL)	0			
TIM (nmole/L)	0			
Total Concentration (ng/uL)	31.05			
Final Concentration (ng/uL)	41.667			
GQN	2.7			

Table B-10: DNA extraction size and concentration parameters for sample Tthy906 provided by the sequencing center. Parameters include Total Integrated Concentration (TIC), Total Integrated Molarity (TIM), Genomic Quality Number (GQN), and Coefficient of Variance (CV).

	Size (bp)	Concentration (ng/uL)	Molarity (nmole/L)	CV%
Peak 1	1 (LM)	1.4882	1944.566	254.39
Peak 2	61.43	32.8533	8.804	78.64
TIC (ng/uL)	32.853			
TIM (nmole/L)	8.804			
Total Concentration (ng/uL)	45.378			
Final Concentration (ng/uL)	41.667			
GQN	2.3			