UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MOLECULAR EVOLUTION OF OXIDATIVE PHOSPHORYLATION GENES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

FEIFEI ZHANG
Norman, Oklahoma
2015

MOLECULAR EVOLUTION OF OXIDATIVE PHOSPHORYLATION GENES

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF BIOLOGY

BY

_____
Dr. Richard Broughton, Chair

_____
Dr. Rosemary Knapp

_____
Dr. Lawrence Weider

_____
Dr. John Masly

_____
Dr. Cecil Lewis

Dedication

To those who are interested in fish and molecular evolution.

Acknowledgments

I would like to thank the Department of Biology at the University of Oklahoma, who gave me the opportunity to pursuit my Ph.D. degree six years ago. This is an excellent environment full of respect, kindness, and resources, for young scientists (domestic or abroad) to develop their academic careers. I feel lucky to have spent the most important six years of my life studying here.

I would also like to thank my advisor Dr. Richard Broughton. He always gave me professional instructions and provided support during the whole Ph.D. study, from big picture questions to detailed technique support, from presenting in seminars to manuscript editing, etc. In addition, he also gave me the freedom and support to pursue what I like. Without his mentoring, I could not have achieved so much.

I also want to thank my other committee professors. Dr. Rosemary Knapp gave me tremendous help in fish physiology such as metabolism, endothermy, and locomotion mechanisms through the independent study I did with her. I also learned it is important to keep critical thinking and retrospective thinking as a scientist (*e.g.,* reading classic textbook once in a while). As the graduate student liaison of the Department of Biology, she also made my life much easier by helping me out in many situations. She has been my role model as a successful women scientist who can balance research, teaching, mentoring, and family so well. Dr. Lawrence Weider is such a nice and supportive committee member to have. I always gotten help from him quickly, even when he was in the airport on his trip! Dr. John Masly helped me build up my knowledge base of molecular evolution through an independent study, and taught me the R language, which was very helpful. Dr. Cecil Lewis joined my committee in

the late stage but he and his lab gave me tremendous help with the next-generation sequence analysis.

I want to give my thanks to the staff at the Oklahoma Biological Survey and the Department of Biology of the University of Oklahoma, particularly Dan Hough, Trina Steil, Ranell Madding, Carol Baylor, Robbie Stinchcomb, Kaye Carter, and Kyler Baker, for making my life easier.

I also want to show my thanks to my colleagues and friends during my long Ph.D. journey. James Cureton II was such a nice lab mate during the past five years. I enjoyed the time when we talked about science; we gave suggestions on each other's projects. I appreciate his generous help with R and editing my writing. I really appreciate Dr. Krithi Sankaranarayanan's generous help with the next-generation sequence analysis. I also thank my roommate and colleague Rachel Hartnett. She is not only a sweet roommate, but also good company as a great scientist. I thank my friends Dr. Zhaozhe Hao, Dr. Mingzi Xu, Dr. Dahong Chen, Chenmei Xu, Dr. Tingting Gu, Yue Ban, Tian Yuan, Zhicong Chu, Chen Xu, Dr. Andrea Contina, Dr. Claire Curry, Diane Landoll, and Brent Tweedy for the happy hours we spent together.

I can never thank my parents and family, and especially my fiancé, enough for their support all along way!

Table of Contents

List of Tables

List of Figures

Abstract

Physiological processes may serve as mechanistic links between organismal genotypes and phenotypes. Accordingly, adaptations in genes involved in energy metabolism pathways may facilitate the evolution of organismal physiology and behavior with diverse energy requirements. Oxidative phosphorylation (OXPHOS) is one of the most ancient and conserved physiological functions and the major source of ATP in eukaryotic cells. My dissertation research explores how the genes involved in the OXPHOS system evolve from four perspectives: 1) investigating the co-evolution between mitochondrial and nuclear OXPHOS genes in fishes and mammals; 2) how natural selection affects OXPHOS genes in fishes with different swimming performance; 3) how environmental factors (temperature and salinity) and a life history trait (migration) affect the rates of evolution of mitochondrial OXPHOS genes in fishes; and 4) what are the patterns of positive natural selection on mitochondrial OXPHOS genes across bony fishes?

The OXPHOS pathway is composed of protein subunits encoded by both mitochondrial and nuclear genomes. The successful interaction of OXPHOS proteins encoded by both genomes plays a central role in the maintenance of OXPHOS function. Under the compensatory model, deleterious substitutions at one nucleotide site could be compensated by a subsequent (or simultaneous) substitution at an interacting site. Generally, the mitochondrial genome evolves faster than the nuclear genome. If nonsynonymous substitutions of mitochondrial genes drive corresponding nonsynonymous substitutions of nuclear OXPHOS genes, one expects to see the acceleration of d_N (nonsynonymous substitution rate) in nuclear OXPHOS genes

relative to most nuclear genes not involved in OXPHOS. In Chapter 1, I examined the substitution rates of 13 mitochondrial OXPHOS genes, 60 nuclear OXPHOS genes, and 77 nuclear non-OXPHOS genes in 7 fishes and 40 mammals. I found that d_N (mitochondrial OXPHOS genes) > d_N (60 nuclear OXPHOS genes) > d_N (77 nuclear non-OXPHOS genes), which supports the compensatory evolution hypothesis. However, results from two (out of five) OXPHOS complexes did not fit this pattern when analyzed separately. I found that the d_N of nuclear OXPHOS genes for "core" subunits (those involved in the major catalytic activity) was lower than that of "noncore" subunits, whereas there was no significant difference in d_N between genes for nuclear non-OXPHOS and core subunits. This latter finding suggests that compensatory changes play a minor role in the evolution of OXPHOS genes and that the observed accelerated nuclear substitution rates are due largely to reduced functional constraint on noncore subunits.

Fishes exhibit extreme variation in swimming performance, ranging from "high-performance" tunas and billfishes to largely sedentary species like seahorses and flounders. Positive natural selection at the gene level (DNA sequence) favors adaptive substitutions that could benefit the fitness of the whole organism. Evidence of positive selection on OXPHOS genes has been associated with evolution of a variety of energetically demanding characteristics such as the origin of large brains in anthropoid primates, powered flight in bats, and adaptation to cold environment in polar bears. In Chapter 2, I examined all major branches on a phylogeny of fishes with diverse swimming performance, testing whether positive natural selection on mitochondrial and nuclear OXPHOS genes was associated with high-performance fishes. The results were

not as predicted: positive selection was associated with branches leading to fishes with low and moderate performance, while negative (purifying) selection dominated on branches leading to fishes with high-performance. This result indicates a complicated evolutionary scenario of fish swimming and it could be possible that positive selection favors changes leading to low OXPHOS efficiency in low performance fishes.

Understanding the factors that affect the rates of nucleotide substitution is central to evolutionary biology, population genetics, and mutation research. How the evolution of mitochondrial genes is affected by environmental factors or life history traits has not been examined in vertebrates at a broad phylogenetic scale. Fishes exhibit great variation in thermal environments (tropical, subtropical, temperate, and deep cold water), salinity (fresh water, brackish, and salt water), and migration (anadromous, catadromous, amphidromous, oceanodromous, potamodromous, and non-migratory) and the newly published fish tree of life provides a broad evolutionary background for such analyses. In Chapter 3, I investigated how the substitution rates of mitochondrial protein-coding genes are correlated with thermal environment, salinity, and migratory ability in 972 fish species. I found that tropical fishes have the highest d_S and d_N, while deep cold water fishes have higher d_S and d_N than subtropical and temperate fishes. This results suggest that substitution rates may not be only affected by temperature value (high or low), but may also be affected by the stability of the temperature (constant or variable). Similar patterns were also found among fishes from different salinity levels: fishes that can live in both freshwater and salt water (variable salinity environment) have lower d_N and d_S than fishes that live only in either freshwater or salt water (stable salinity environment). Migratory fishes have lower substitution rates than non-

migratory fishes. This is probably because migratory fishes with high energy demands usually have high OXPHOS efficiency, thus OXPHOS genes are under strong constraint with low tolerance for substitutions. Among different types of migratory fishes, amphidromous fishes have the highest d_N and d_S, but the reason for this observation is not clear.

The recently developed codon based branch-site model using maximum likelihood is a powerful tool for detecting positive selection on protein-coding sequences, providing an effective means of identifying the plausible candidate genes or residues for further testing. The OXPHOS pathway is the primary source of ATP in eukaryotic cells and ample evidence has shown positive selection on mitochondrial OXPHOS genes on branches leading to organisms with high energy demands or organisms that are well adapted to cold environments. To test if this is a general pattern, in Chapter 4, I identified 11 groups of bony fishes with some interesting characteristics that incur high energy costs (and two groups are adapted to cold environments), and examined the evidence for positive selection on mitochondrial OXPHOS genes on branches leading to the target group, its sister group, and the most recent common ancestor. In most cases, positive selection was found to be associated with the target group, but various patterns were identified. It appears that the pattern of positive selection is case specific and was determined by the particular evolutionary history of each group.

Though there are many challenges to studying the molecular evolution of genes involved in OXPHOS due to its complexity, much can be learned about the functional importance and unique bi-genome composition of this fundamental system. More

research is needed on nuclear OXHOS gene sequences, protein crystal structures, and direct measurement of OXPHOS efficiency. Studying OXPHOS opens a window to increase our understanding of basic life processes.

Chapter 1: Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes?

This chapter is published, with some modifications, as Zhang F. and Broughton R.E. 2013 Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? *Genome Biology and Evolution* 5(10): 1781-1791.

Abstract

Oxidative phosphorylation (OXPHOS), the major energy-producing pathway in aerobic organisms, includes protein subunits encoded by both mitochondrial (mt) and nuclear (nu) genomes. How these independent genomes have coevolved is a long-standing question in evolutionary biology. Although mt genes evolve faster than most nu genes, maintenance of OXPHOS structural stability and functional efficiency may involve correlated evolution of mt and nu OXPHOS genes. The nu OXPHOS genes might be predicted to exhibit accelerated evolutionary rates to accommodate the elevated substitution rates of mt OXPHOS subunits with which they interact. Evolutionary rates of nu OXPHOS genes should, therefore, be higher than that of nu genes that are not involved in OXPHOS (nu non-OXPHOS). We tested the compensatory evolution hypothesis by comparing the evolutionary rates (synonymous substitution rate d_S and non-synonymous substitution rate d_N) among 13 mt OXPHOS genes, 60 nu OXPHOS genes, and 77 nu non-OXPHOS genes in vertebrates (7 fish and 40 mammal species). The results from a combined analysis of all OXPHOS subunits fit the predictions of the hypothesis. However, results from two OXPHOS complexes did not fit this pattern when analyzed separately. We found that the d_N of nu OXPHOS

genes for "core" subunits (those involved in the major catalytic activity) was lower than that of "non-core" subunits, while there was no significant difference in d_N between genes for nu non-OXPHOS and core subunits. This latter finding suggests that compensatory changes play a minor role in the evolution of OXPHOS genes and that the observed accelerated nu substitution rates are due largely to reduced functional constraint on non-core subunits.

Introduction

The oxidative phosphorylation (OXPHOS) system is unique among fundamental metabolic pathways in animals (aside from maintenance and gene expression of mitochondria) in that functional OXPHOS complexes are composed of subunits encoded by two different genomes. The different hereditary modes of the mitochondrial (mt) and nuclear (nu) genomes substantially increase the complexity of maintaining functional interactions among OXPHOS subunits. The role of OXPHOS in the fundamental process of aerobic ATP production means that slight perturbations to inter-genomic coordination result in a wide range of human pathologies (Wallace 2010). In addition, minimal divergence of OXPHOS subunits may result in reproductive incompatibility among closely related populations (Rand et al. 2004). Consequently, the successful interaction of OXPHOS proteins encoded by mt and nu genomes plays a central role in fundamental processes at the cellular, organismal and population levels of biological organization. The OXPHOS system provides a rare opportunity to investigate the coevolution of subunits that must successfully interact in ATP production, yet are subject to potentially very different evolutionary forces that govern the mt and nu genomes.

OXPHOS is a complex system composed of the electron transport chain and ATP synthase. It produces about 80% of energy in the form of ATP in almost all eukaryote cells and is the major role-player making mitochondria the "powerhouse of the cell". Meanwhile, reactive oxygen species (ROS), the by-product of normal metabolism, can damage DNA, cell membranes and lipids, decrease bioenergetic efficiency, and thus lead to aging, disease and death in humans (Ballard and Melvin 2010; Wallace 2010). The OXPHOS apparatus includes 5 complexes (I-V), each composed of between 4 and 30+ subunits (De Grassi, et al. 2005). Complexes I-IV transfer electrons through a series of redox reactions coupled to the transport of protons into the inter-membrane space. Complex V employs the resulting electro-chemical gradient to phosphorylate ADP to ATP. Like most multimeric protein complexes, successful OXPHOS function requires that subunits maintain structures that allow for specific protein-protein interactions, both within and among complexes. As a result, amino acid residues at many sites are expected to be tightly constrained. An added level of complexity arises from the fact that the composition of 4 of the 5 OXPHOS complexes includes subunits encoded by both mt and nu genomes. Evolution of OXPHOS must therefore be coordinated among a large number of interacting subunits and among genomes, and compensatory amino acid substitutions may, in some cases, be favored to maintain function.

The OXPHOS pathway is localized in the mitochondrial inner membrane. Complex I (NADH: quinone oxidoreductase) and complex II (succinate dehydrogenase) receive electrons from reducing equivalents (NADH and $FADH_2$) produced by the Krebs cycle. The electrons are transferred to ubiquinone, which freely diffuses within

the mitochondrial inter-membrane space and transfers electrons to complex III (cytochrome bc1). Complex III transfers electrons from ubiquinol to cytochrome c. This small protein is localized on the inter-membrane space side of the mitochondrial inner membrane. It transfers electrons to molecular oxygen in complex IV (cytochrome c oxidase), where oxygen is reduced to water. Coupled with the process of electron transport, complexes I, III, and IV transport protons to the inter-membrane space. The generated proton gradient across the mitochondrial inner membrane drives ATP synthesis in complex V (ATP synthase) (Boekema and Braun 2007; Saraste 1999).

The mt genome has a constant gene content composition in all metazoan species (with few exceptions), including 13 protein-coding genes, 2 rRNA genes, and 22 tRNA genes. The rRNA and tRNA genes are involved in mt protein synthesis, whereas the 13 protein-coding genes all encode components of OXPHOS (Nei and Kumar 2000; Shadel and Clayton 1997). At least 60 OXPHOS subunits are encoded in the nu genome. Despite the fundamental importance of OXPHOS to aerobic life, mt protein-coding genes are not highly conserved and evolve at rates that are 5-50 times that of typical nu genes in vetebrates (Lynch 2007). This increased rate has been observed in a range of animal species, including primates (Brown, et al. 1979), Galapagos tortoises (Caccone et al. 2004), lice (Johnson et al. 2003), fly *Drosophila* (Haag-Liautard, et al. 2008), wasps (Kaltenpoth et al. 2012), and nematodes (Denver et al. 2000). The rapid evolutionary rate of mt genomes may be due to the cell cycle-independent replication of mtDNA (Ohta and Kimura 1973), increased exposure to mutagenic oxygen radical species (Beal 1996), lack of protective histones, and limited DNA repair capacity (Croteau and Bohr 1997). Moreover, the mt genome differs from the nu genome in

being effectively haploid, maternally inherited, and exhibiting little recombination (Ballard and Melvin 2010; Lightowlers, et al. 1997; Moritz, et al. 1987). Haploid maternal inheritance means that the effective population size for mt genes is approximately one quarter that of nu autosomal genes, increasing the fixation rate of polymorphic sites (Birky 2001; Gabriel, et al. 1993). In addition, the lack of recombination reduces the ability to purge deleterious mutations (Gabriel, et al. 1993). Incompatibility between mt and nu genomes has been shown to reduce hybrid fitness in copepods (Willett and Burton 2001), *Drosophila* (Sackton, et al. 2003), yeast (Zeyl et al. 2005), and wasps (Niehuis et al. 2008), leading to the suggestion that mt-nu incompatibility may be an important contributor to the process of speciation (Gershoni, et al. 2009).

Among interacting proteins, amino acid substitutions may be deleterious if they affect domains involved in the interaction, yet there is the potential for compensatory changes at interacting amino acid sites to reduce or eliminate any negative effects. Such interacting sites are generally proximal to each other in the three-dimensional protein structure (Pazos and Valencia 2008). Under the compensatory model, substitutions that may destabilize protein structures or inhibit function could be compensated by a subsequent (or simultaneous) substitution at an adjacent site (Pollock et al. 1999). In the OXPHOS system, the elevated evolutionary rate of mtDNA suggests that deleterious substitutions occur more frequently in mt genes and compensatory changes may then occur in proximal sites of nu-encoded proteins. In bacteria (Sharp and Li 1987), *Drosophila* (Comeron and Kreitman 1998; Dunn et al. 2001), and mammals (Wolfe and Sharp 1993), the rate of nonsynonymous substitutions (d_N) is positively correlated with

the synonymous substitution rate (d_S) and both d_N and d_S are higher in mt than nu genes. If nonsynonymous substitutions of mt genes drive corresponding nonsynonymous substitutions of nu OXPHOS genes, we expect to see the acceleration of d_N in nu OXPHOS genes relative to most nu genes not involved in OXPHOS (nu non-OXPHOS).

A few recent investigations have addressed mt-nu coevolution at the molecular level but have been restricted to a few genes or a small number of species [*e.g.*, nu genes *CYC1* and *UQCRFS1*, and mt gene *MT-CYTB* (Willett and Burton 2001), or genes involved in OXPHOS complex IV (Goldberg et al. 2003)]. Here we describe a broader study of 73 OXPHOS genes and a comparison between OXPHOS and 77 non-OXPHOS housekeeping genes. Under a compensatory evolution scenario, we predicted that d_N of mt OXPHOS > nu OXPHOS > nu non-OXPHOS genes. We tested the compensatory substitution model by comparing the evolutionary rates (both d_S and d_N) of 13 mt protein-coding genes (mt OXPHOS genes), 60 nu OXPHOS genes, and 77 non-OXPHOS genes in 47 vertebrate species (7 fishes and 40 mammals). The 7 fish species are phylogenetically disparate, spanning some 250 million years of evolutionary history (Betancur-R et al. 2013) (Figure 1A), whereas most of the mammal lineages emerged within the last 80 million years (Meredith et al. 2011) (Figure 1B).

Methods

This study employed genome sequences of 7 teleost fishes and 40 mammals from which we acquired the coding regions of 13 mt OXPHOS and 60 nu OXPHOS genes. For comparison, we sampled 77 housekeeping genes (Amsterdam et al. 2004; Warrington et al. 2000). These were considered appropriate "control" comparators

because OXPHOS genes contribute to critical function and should be highly expressed in virtually all cell types. All sequences were downloaded from the Ensembl Genome Browser (www.ensembl.org) using a pipeline procedure (Vilella et al. 2009) (see Supplementary Table 1 doi: 10.1093/gbe/evt129). A phylogenetic gene tree was used to identify orthologous and paralogous sequences so that we used only the former. Each gene sequence was then realigned using Geneious Pro v.5.6 (Biomatters, Ltd.).

Phylogenetic relationships of the fish species followed that of Betancur-R et al. (2013) (Figure 1A), while the phylogeny of mammals was from Meredith et al. (2011) (Figure 1B). The program CODEML implemented in the computer software PAML v.4.7 (Yang 2007) was used to estimate d_N and d_S for each gene. The site-model M0 (model: 0, NSsites: 0, and fix-omega: 0) was used when estimating d_N and d_S. For each gene, the estimated substitution rates are the sum of d_N or d_S values from each branch of the tree, so total rate values will vary with the number of taxa in the tree. The intent here was not to compare the absolute values between fishes and mammals, but instead to examine the patterns among mt OXPHOS genes, nu OXPHOS genes, and non-OXPHOS genes within taxonomic groups (where the number of branches was identical). Codon frequencies were estimated from the average nucleotide frequencies at three codon positions. To avoid problems arising from local optima (Yang and Nielsen 1998), replicate runs with three different starting values of ω (0.3, 0.9, and 4.3) were used. Different starting ω yielded similar, and in many cases identical, d_S and d_N estimates for each gene among three runs (see Supplementary Table 2 doi: 10.1093/gbe/evt129). The average d_S and d_N from the three runs was used in the subsequent analysis.

We tested the hypothesis that there are differences in d_S and d_N among three groups of genes: 13 mt OXPHOS genes, 60 nu OXPHOS genes, and 77 non-OXPHOS genes. A Shapiro-Wilks test was performed to assess the normality of d_S and d_N for all genes and Bartlett's test was performed to evaluate homogeneity of variance. Based on these results, a Kruskal-Wallis test followed by post-hoc Tukey HSD tests were used to assess differences among the groups. A Wilcoxon Signed Rank test was performed to evaluate differences between mt and nu OXPHOS genes in each of the five OXPHOS complexes. All statistical analyses were performed in R (Rdevelopment 2008).

Each of the OXPHOS complexes are composed of multiple subunits and each subunit may be more or less important to the core catalytic activity of the complex. We partitioned OXPHOS genes into "core" vs. "non-core", where core genes encode subunits that retain structural, functional and sequence homology with prokaryotic systems. For example, unlike human complex I, which consists of 43 subunits, bacterial complex I consists of 14 subunits. These 14 subunits all have homologs in human complex I, and thus are referred as core subunits. These include 7 nu OXPHOS genes (*NDUFS1-3*, *NDUFS7-8*, and *NDUFV1-2*) and 7 mt OXOPHOS genes (*MT-ND1-6*, *MT-ND4L*) (Janssen, et al. 2006). Similarly, in complex IV, bovine mt OXPHOS genes (*MT-COX1-3*) share high similarity to their bacterial counterparts and can be considered the functional core of the eukaryotic oxidase (Richter and Ludwig 2003a). Complex II is only composed of 4 nuclear subunits (*SDHA-D*), all of which share high sequence similarity with *E. coli* (Cecchini 2003). In complex V, 6 human subunits have bacterial homologs (*ATP5A1*, *ATP5B*, *ATP5C1*, *ATP5D*, *ATP5F1*, and *ATP5G3*) (Yoshida et al. 2001) and were categorized as core subunits. Because *E. coli* does not have a

homologue of complex III (Lenn et al. 2008), none of the nu OXPHOS genes in complex III were categorized as encoding core subunits.

Results

We estimated the total synonymous and nonsynonymous substitution rates on each tree for fishes (Figure 1A) and mammals (Figure 1B) separately using maximum likelihood methods. Because neither d_S nor d_N was normally distributed and each exhibited significant variance heterogeneity, non-parametric statistics were used to compare the different groups of genes. We found significant differences (Kruskal-Wallis) in d_S and d_N among the three groups of genes: 13 mt OXPHOS genes, 60 nu OXPHOS genes, and 77 non-OXPHOS genes (Table 1). The d_S of mt OXPHOS genes was significantly higher than that of nu OXPHOS and non-OXPHOS genes in both fishes and mammals (Figure 2A, C). The d_S of mt and nu genes was 16.02 ± 9.64 and 2.02 ± 0.87, respectively, in fishes and 72.27 ± 26.56 and 8.85 ± 4.15, respectively, in mammals. No significant differences of d_S were detected between nu OXPHOS genes and non-OXPHOS genes (Figure 2A, C). We note that a small number of mt genes had exceptionally high d_S values, which may indicate substitutional saturation, leading to violation of the assumptions of the maximum likelihood estimation method. As predicted, the d_N of mt OXPHOS genes was significantly higher than that of nu OXPHOS genes, which in turn was significantly higher than that of non-OXPHOS genes (Figure 2B, D).

Estimates of d_S and d_N were assessed for each OXPHOS complexes. The d_S of mt genes was higher than that of nu genes in all four complexes (complex II is composed of subunits only encoded by nu genes) (Figure 3A, C). The d_N of mt genes

was not consistently higher than that of nu genes in all complexes. For example, the d_N of mt genes in complex III was lower than that of nu genes in fishes (Figure 3B) while higher than that of nu genes in mammals (Figure 3D); the d_N of mt genes in complex IV was significantly lower than that of nu genes in fishes (Figure 3B) and lower than that of nu genes in mammals (Figure 3D).

Because of their different functions and interactions, OXPHOS genes may be under different levels of functional constraint, so we separated them into "core" and "non-core" groups. Core subunits were defined as those that have bacterial homologs and that participate directly in electron transport or ATP synthesis. After partitioning, d_S of mt OXPHOS genes was found to be significantly higher than that of all nu gene groups (Figure 4A, C), while the d_N of non-core OXPHOS genes was higher than that of core OXPHOS genes (Figure 4B, D). There were no significant differences in d_N between nu non-OXPHOS genes and nu core OXPHOS genes (Figure 4B, D) (Table 2).

Discussion

To infer factors influencing the evolution of mitochondrial-nuclear interactions, we used estimated rates of synonymous and nonsynonymous substitutions among all OXPHOS genes and a broad genomic sample of housekeeping genes. Specifically, we examined the hypothesis of compensatory evolution between mt and nu OXPHOS genes by comparing their evolutionary rates (both d_S and d_N) relative to non-OXPHOS genes in two distinct vertebrate lineages. We found that the mean d_S of mt genes was about 7-9 times higher than that of nu genes. This magnitude is consistent with the well-established pattern of high d_S in animal mt genes (Brown et al. 1979; Caccone et al. 2004; Johnson et al. 2003; Kaltenpoth et al. 2012; Lynch 2007). Although the products

of nu OXHPOS genes are transported into mitochondria where they function, they are expected to show mean d_S similar to the non-OXPHOS genes because the coding sequences reside in the nu genome. Our results show this to be the case and suggest that differences in nonsynonymous rates are due to selective factors rather than different mutation rates.

Mitochondrial-nuclear coevolution could generate predictions at different levels. One prediction at the cellular level is that in experimental "transplants" interacting partners should result in diminished functional performance, and this disruption should increase as the level of evolutionary divergence increases (Rand et al. 2004). In mouse (*Mus musculus domesticus*) cell lines carrying mitochondria from six different murid species spanning 2-12 million years of divergence, a near-linear association between disruption of respiratory chain function and evolutionary distance has been observed (McKenzie et al. 2003). This phenomenon has also been observed in primates (Kenyon and Moraes 1997), copepods (Willett and Burton 2001), *Drosophila* (Sackton et al. 2003), and wasps (Niehuis et al. 2008). However, more limited divergence may not lead to disruption of performance. Introgression experiments of *D. simulans* s*i*II mtDNA type into the sympatric population s*i*III nuclear background did not show a difference in catalytic properties of mitochondria, indicating some naturally occurring mutations in mtDNA can be accommodated by different nuclear background and mitochondrial-nuclear interaction (Pichaud et al. 2012).

At the DNA sequence level, higher evolutionary rates of mt genes could drive accelerated evolutionary rates of nu genes as compensatory response in the nu genes contributes to the maintenance of function. Several studies of the primate complex IV

showed that 7 nu OXPHOS genes: *COX4I1* (Wildman et al. 2002; Wu et al. 1997), *COX7A1H* (Schmidt et al. 1999), *COX5A* (Doan 2004), *COX8AL* (Goldberg et al. 2003; Osada et al. 2002), *COX6B1*, *COX6C*, and *COX7C* (Doan 2004), together with two mt OXPHOS genes: *MT-COX1* (Andrews and Easteal 2000; Keightley 2012) and *MT-COX2* (Adkins and Honeycutt 1994), have shown accelerated d_N in the lineage leading to hominids relative to other primates. However, in these cases it is not clear whether accelerated d_N is due to compensatory changes driven by the higher mt rate, or is due to selection for increased OXPHOS efficiency associated with increased brain size in the hominid lineage. Our current study shows that when all complexes were analyzed together, the d_N of 13 mt OXPHOS genes was significantly higher than that of 60 nu OXPHOS genes, and both of these were significantly higher than d_N of 77 non-OXPHOS genes. The apparent acceleration d_N of nu OXPHOS genes is consistent with a compensatory response to the higher d_N of mt OXPHOS genes.

However, this explanation seems tenuous when considering rates in each complex separately. Contrary to expectations, our results showed that the d_N of mt genes was not always higher than that of nu genes in each complex. Mt genes with higher d_N than nu genes were only observed in complex I and V. In fishes, the d_N of mt genes was lower than that of nu genes in complexes III and IV. In mammals, the d_N of mt genes was lower than that of nu genes in complex IV, but higher in complex III. Goldberg et al. (2003) found that in the anthropoid lineage, the fast-evolving region (12 N-terminal amino acid residues) of *COX8L* encodes amino acid sites contacted with *MT-COX1* sites in three-dimensional structure. This suggested structurally mediated cytonuclear coevolution, which was driven by faster evolving mt OXPHOS genes.

However, we found no evidence of such a pattern in fishes and mammals. Similarly, the only mt gene in complex III (*MT-CYTB*) showed lower d_N than its corresponding nu OXPHOS genes in fishes. Although d_N of *MT-CYTB* was higher than for complex III nu genes in mammals, the difference is not statistically different. These results are clearly counter to expectations under a compensatory model.

One explanation for the lower d_N of mt OXPHOS genes in complexes III and IV could be functional constraints on core subunits. *MT-CYTB* is the only mt gene in complex III and its structure has been shown to be conserved among vertebrates (Kocher et al. 1989). Similarly, *MT-COX1* and *MT-COX2* have been recognized as the core enzyme catalytic subunits and are conserved from bacteria to bovid (Tsukihara, et al. 1996). It therefore seems likely that functional constraints are stronger on core subunits, regardless of which genome encodes them. The higher d_N of nu OXPHOS genes relative to non-OXPHOS genes in general is due mainly to higher rates of non-core proteins alone. The more recently derived non-core subunits (or "accessory OXPHOS families") appear to be important contributors to OXPHOS assembly and/or stabilization (Ugalde et al. 2004), but do not contribute directly to catalytic activity of electron transport and ATP synthesis (De Grassi et al. 2005). Thus, maintenance of primary amino acid sequences of the non-core subunits appears to be less important to OXPHOS function than that of the core subunits and the intensity of selection on these proteins is therefore reduced.

We cannot exclude the possibility of a few compensatory changes in nu OXPHOS genes as they "keep pace" with changes in mt OXPHOS genes because assessment of evolutionary rates does not allow examination of individual sites. Neither

can we reject compensatory changes in mt genes in response to changes in the nu genome; indeed, the rapid rate of mt genome evolution could allow these genes to more readily respond to changes in nu components. Nonetheless, compensatory changes in either direction do not appear to occur with a frequency sufficient to contribute to differences in evolutionary rates among genes or groups of genes. Therefore, despite the necessary interaction of the products of two different genomes, OXPHOS evolution appears to be driven largely by conventional natural selection for functional efficiency acting on individual subunits, regardless of their genome of origin.

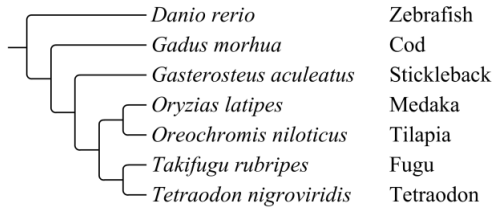
Acknowledgments

This work benefitted from National Science Foundation award DEB-0732988 (to REB). We thank J.C. Cureton II for assistance with statistical analysis and manuscript editing, and L.J. Weider, J.P. Masly and R. Knapp for suggestions on manuscript.

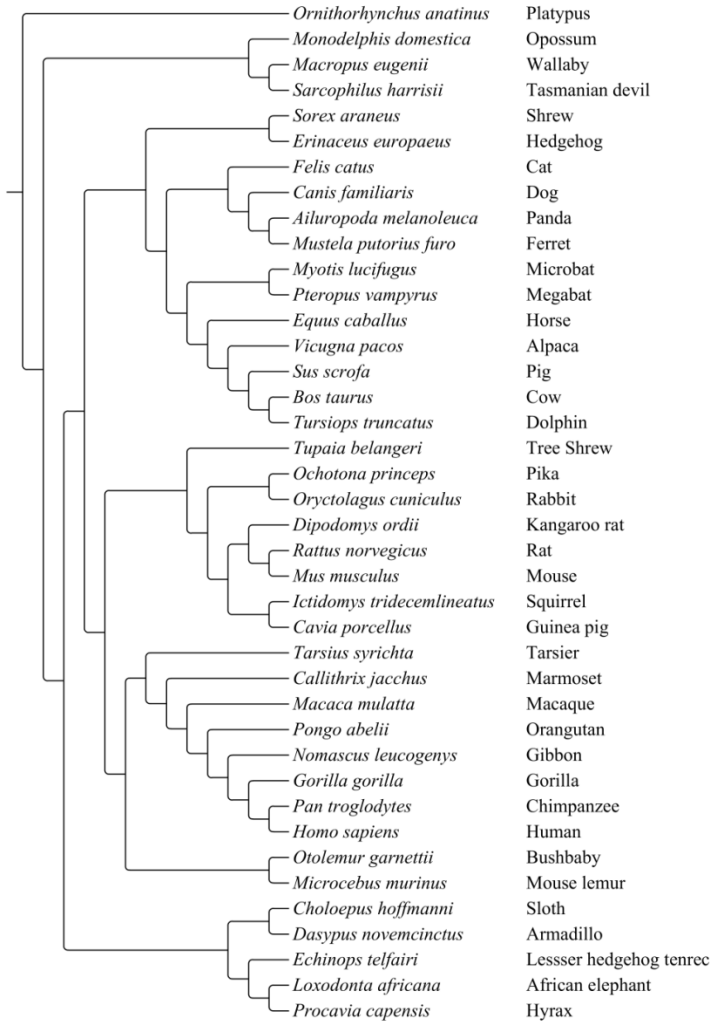
Figures and Tables

Figure 1 Phylogeny of 7 fishes (A) and 40 mammals (B).

A

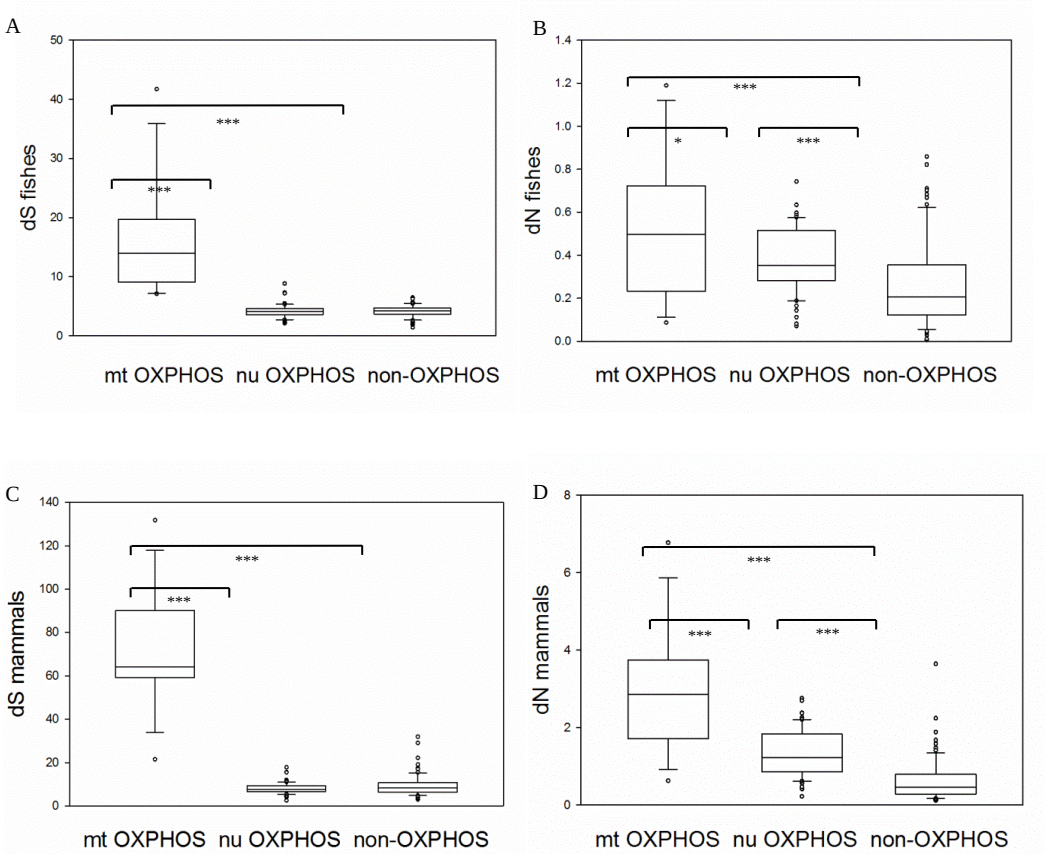


B



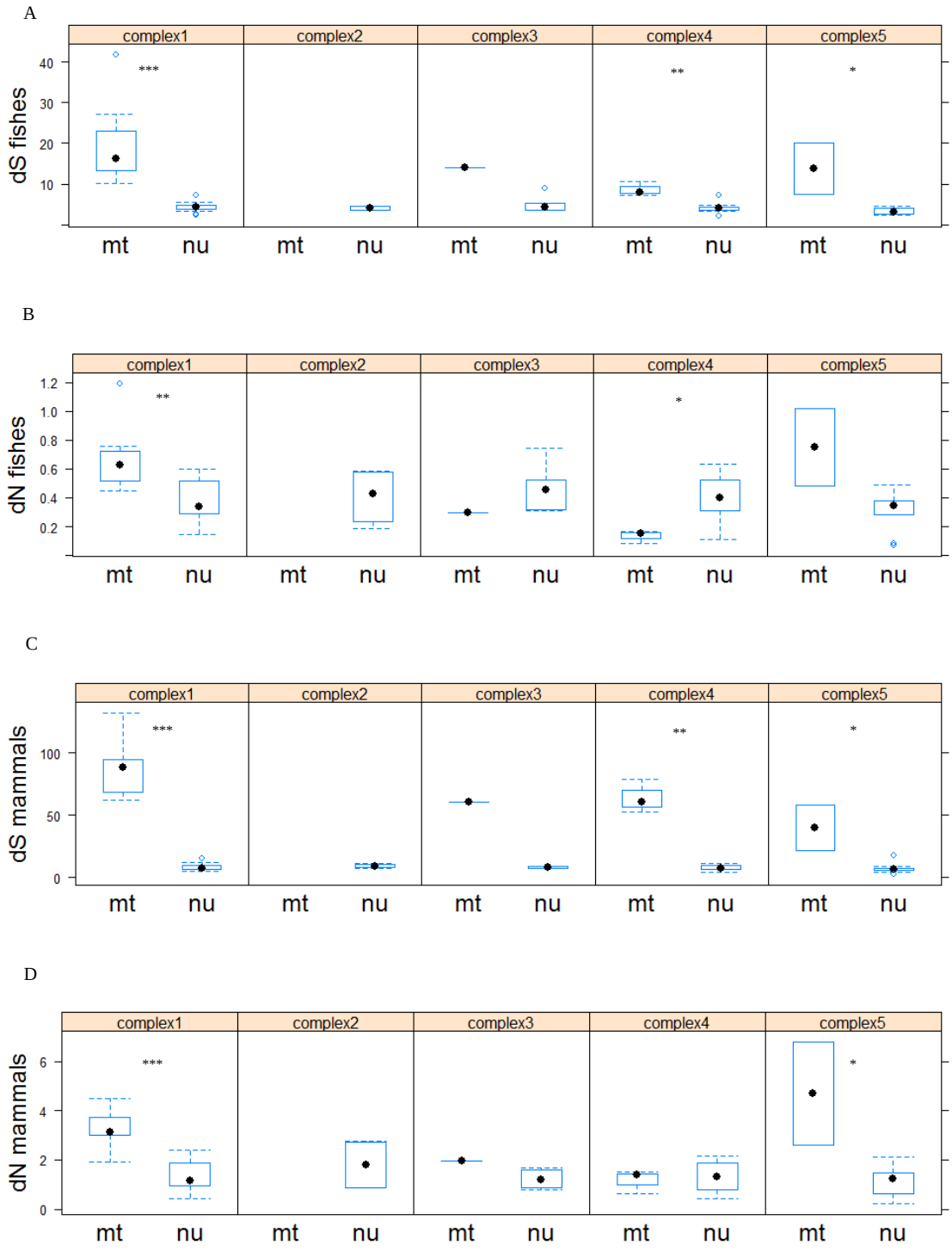
Note: A is from (Betancur-R et al. 2013), and B is from (Meredith et al. 2011).

Figure 2 Comparison of synonymous substitution rate (dS) and nonsynonymous substitution rate (dN) among mitochondrial oxidative phosphorylation (mt OXPHOS), nuclear OXPHOS (nu OXPHOS), and non-OXPHOS genes in 7 fishes (A, B) and 40 mammals (C, D).



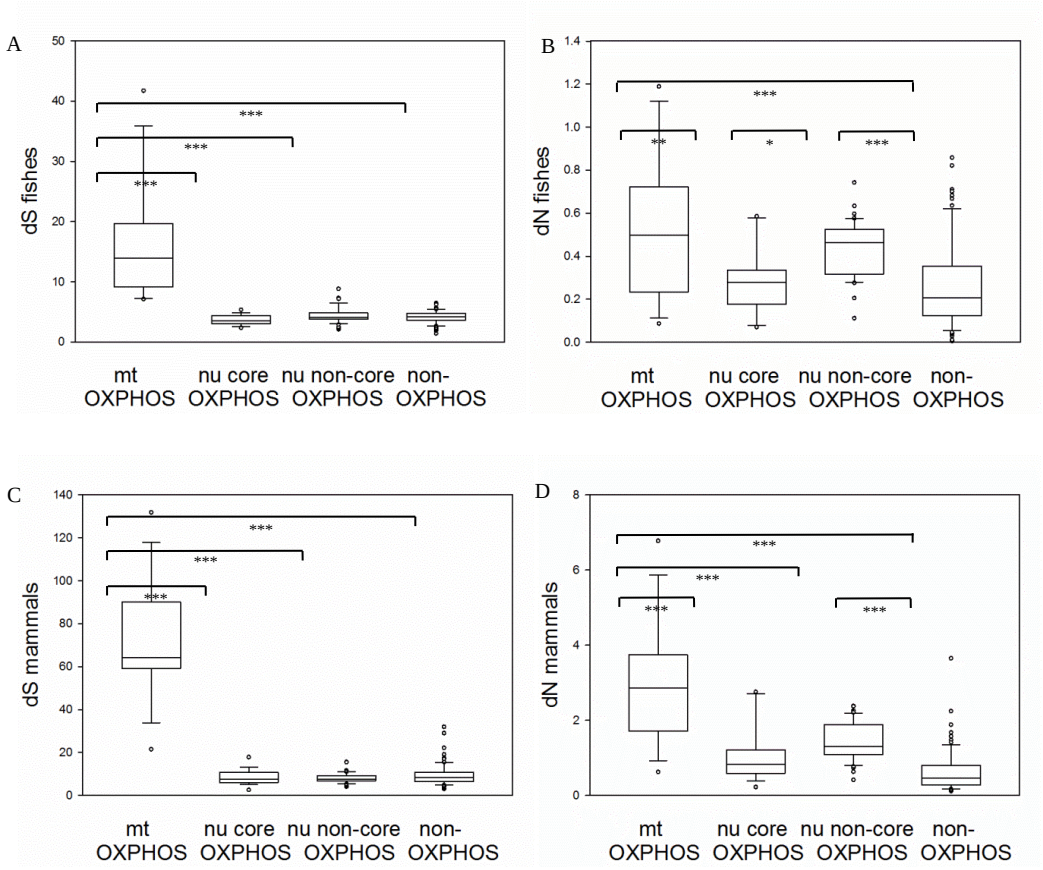
Note: Whisker-ends are at the 5th and 95th percentiles. *P < 0.05, **P < 0.01, ***P <0.001.

Figure 3 Comparison of synonymous substitution rate (dS) and nonsynonymous substitution rate (dN) between mitochondrial (mt) and nuclear (nu) genes in each complex in 7 fishes (A, B) and 40 mammals (C, D).



Note: Whisker-ends are at the 5th and 95th percentiles. *P < 0.05, **P < 0.01, ***P < 0.001. Complex II is composed of subunits encoded only by nu genes.

Figure 4 Comparison of synonymous substitution rate (dS) and nonsynonymous substitution rate (dN) among mitochondrial oxidative phosphorylation (mt OXPHOS), nuclear core OXPHOS (nu core OXPHOS), nuclear non-core OXPHOS (nu non-core OXPHOS), and non-OXPHOS genes in 7 fishes (A, B) and 40 mammals (C, D).



Note: Whisker-ends are at the 5th and 95th percentiles. *P < 0.05, **P < 0.01, ***P < 0.001.

Table 1 Statistical summary for the comparison of synonymous substitution rate (dS) and non-synonymous substitution rate (dN) among mitochondrial oxidative phosphorylation (mt OXPHOS), nuclear OXPHOS (nu OXPHOS), and non-OXPHOS genes in 7 fishes and 40 mammals.

ML Estimate	Taxon	Comparison	*p* value
dS	Fishes	mt OXPHOS > nu OXPHOS	< 0.0001
dS	Mammals	mt OXPHOS > nu OXPHOS	< 0.0001
dS	Fishes	mt OXPHOS > non-OXPHOS	< 0.0001
dS	Mammals	mt OXPHOS > non-OXPHOS	< 0.0001
dS	Fishes	nu OXPHOS > non-OXPHOS	0.987
dS	Mammals	nu OXPHOS > non-OXPHOS	0.670
dN	Fishes	mt OXPHOS > nu OXPHOS	0.037
dN	Mammals	mt OXPHOS > nu OXPHOS	< 0.0001
dN	Fishes	mt OXPHOS > non-OXPHOS	< 0.0001
dN	Mammals	mt OXPHOS > non-OXPHOS	< 0.0001
dN	Fishes	nu OXPHOS > non-OXPHOS	0.0007
dN	Mammals	nu OXPHOS > non-OXPHOS	< 0.0001

Table 2 Statistical summary for the comparison of synonymous substitution rate (dS) and non-synonymous substitution rate (dN) among mitochondrial oxidative phosphorylation (mt OXPHOS), nuclear core OXPHOS (nu core OXPHOS), nuclear non-core OXPHOS (nu non-core OXPHOS), and non-OXPHOS genes in 7 fishes and 40 mammals.

ML Estimate	Taxon	Comparison	*p* value
dS	Fishes	mt OXPHOS > nu core OXPHOS	< 0.0001
dS	Mammals	mt OXPHOS > nu core OXPHOS	< 0.0001
dS	Fishes	mt OXPHOS > nu non-core OXPHOS	< 0.0001
dS	Mammals	mt OXPHOS > nu non-core OXPHOS	< 0.0001
dS	Fishes	mt OXPHOS > non-OXPHOS	< 0.0001
dS	Mammals	mt OXPHOS > non-OXPHOS	< 0.0001
dS	Fishes	nu core OXPHOS < nu non-core OXPHOS	0.881
dS	Mammals	nu core OXPHOS > nu non-core OXPHOS	0.999
dS	Fishes	nu core OXPHOS < non-OXPHOS	0.966
dS	Mammals	nu core OXPHOS < non-OXPHOS	0.967
dS	Fishes	nu non-core OXPHOS > non-OXPHOS	0.969
dS	Mammals	nu non-core OXPHOS < non-OXPHOS	0.847
dN	Fishes	mt OXPHOS > nu core OXPHOS	0.002
dN	Mammals	mt OXPHOS > nu core OXPHOS	< 0.0001
dN	Fishes	mt OXPHOS > nu non-core OXPHOS	< 0.304
dN	Mammals	mt OXPHOS > nu non-core OXPHOS	< 0.0001
dN	Fishes	mt OXPHOS > non-OXPHOS	< 0.0001
dN	Mammals	mt OXPHOS > non-OXPHOS	< 0.0001
dN	Fishes	nu core OXPHOS < nu non-core OXPHOS	0.031
dN	Mammals	nu core OXPHOS < nu non-core OXPHOS	0.193
dN	Fishes	nu core OXPHOS > non-OXPHOS	0.988
dN	Mammals	nu core OXPHOS > non-OXPHOS	0.198
dN	Fishes	nu non-core OXPHOS > non-OXPHOS	< 0.0001
dN	Mammals	nu non-core OXPHOS > non-OXPHOS	< 0.0001

Chapter 2: Heterogeneous natural selection on oxidative phosphorylation genes among fishes with extremes of aerobic performance

This chapter is potentially accepted, pending some modifications, as Zhang F. and Broughton R.E. 2015 Heterogeneous natural selection on oxidative phosphorylation genes among fishes with extremes of aerobic performance. *BMC Evolutionary Biology*

Abstract

Oxidative phosphorylation (OXPHOS) is the primary source of ATP in eukaryotes and serves as a mechanistic link between variation in genotypes and energetic phenotypes. While several physiological and anatomical factors may lead to increased aerobic capacity, variation in OXPHOS proteins may influence OXPHOS efficiency and facilitate adaptation in organisms with varied energy demands. Although there is evidence that natural selection acts on OXPHOS genes, the focus has been on detection of directional (positive) selection on specific phylogenetic branches where traits that increase energetic demands appear to have evolved. We examined patterns of selection in a broader evolutionary context, i.e., on multiple lineages of fishes with extreme differences in aerobic performance. We found that patterns of natural selection on mitochondrial OXPHOS genes are complex among fishes with different swimming performance. Positive selection is not consistently associated with high performance taxa and actually appears to be strongest on lineages containing low performance taxa. In contrast, within high performance lineages, stabilizing (negative) selection appears to predominate. Similar patterns were also found for 15 nuclear OXPHOS genes for a portion of the species. We provide evidence that selection on OXPHOS genes varies in both form and intensity within and among lineages. The type and direction of selection

are heterogeneous through evolutionary time and vary in ways that would not be readily predicted based solely on organismal performance. These results provide empirical evidence for fluctuating selection on OXPHOS genes associated with divergence in aerobic performance. The broader pattern we found indicates a complex interplay between organismal adaptations, ATP demand, and OXPHOS function.

Introduction

Physiological processes may serve as mechanistic links between genotypes and organismal phenotypes. Accordingly, adaptations in genes of energy metabolism pathways may facilitate the evolution of organismal structures and life habits with diverse energy requirements. Proteins encoded by the mitochondrial genome serve as core subunits of the oxidative phosphorylation (OXPHOS) system, the primary source of ATP in eukaryotic cells. Consequently, organismal traits with differing ATP demands may be influenced by adaptation in mitochondrial genomes (Ballard and Melvin 2010; Ballard and Whitlock 2004; da Fonseca et al. 2008; Hill 2014; Rand 1994). Our understanding of patterns and rates of adaptive change in mitochondrial genomes remains limited despite extensive use of mitochondrial genes as molecular markers in evolutionary studies (Ballard and Rand 2005) and the important role of mitochondria in many human pathologies (Wallace 2010). Although there is ample evidence of natural selection acting on mitochondrial genes (Ballard et al. 2007; Bazin et al. 2006; Hassanin et al. 2009; Meiklejohn et al. 2007; Mishmar et al. 2003; Moyer et al. 2005; Rand et al. 2004; Ruiz-Pesini et al. 2004), the functional significance of adaptive mitochondrial change is rarely known.

Evidence of positive selection on OXPHOS genes has been associated with evolution of a variety of energetically demanding characteristics (reviewed in Garvin et al. 2015), including origin of large brains in anthropoid primates (Doan et al. 2004), powered flight in bats (Shen et al. 2010), and adaptation to cold environment in polar bears (Welch et al. 2014). These findings suggest a significant role for OXPHOS in organismal adaptation, and because divergence among lineages often involves traits with different energy usage, OXPHOS evolution may be an important factor in the diversification of life.

OXPHOS functional efficiency may be particularly important to energy intensive processes such as locomotion. Variation in locomotive performance of fishes is among the most extreme among vertebrates, ranging from largely sedentary filter feeders and sit-and-wait predators to highly migratory species and active pelagic foragers. For example, seahorses and flounders spend much of their time nearly motionless, whereas tunas and marlins are "high-performance" swimmers that exhibit high aerobic metabolism and prolonged fast swimming (Beamish 1978; Block et al. 1992). Highly-active fish taxa exhibit many morphological and physiological adaptations that enhance swimming performance (reviewed in Lauder, et al. 2006). Such adaptations include modifications of body shape and hydrodynamics (Blake 1983; Vogel 1994; Webb 1994), swimming form and mechanical kinematics (Videler and Nolet 1990; Wainwright et al. 2002; Walker and Westneat 2002), muscle composition (Graham et al. 1983; Moon et al. 1991; Tang and Wardle 1992), metabolic rates (Blank et al. 2007), heart volume and aerobic capacity (Blank et al. 2007; Brill and Bushnell 1991; Farrell 1991), and mitochondrial structure and concentration (Moyes et al. 1992).

However, much less is known about the molecular adaptations that influence organismal performance.

At the molecular level, expression levels of proteins directly involved in energy metabolism may be increased in highly mobile fishes. Tunas and marlins, which have higher cruising speeds than other active fishes (Block et al. 1992), have been shown to have elevated myoglobin levels (Dickson 1995) and higher concentrations of metabolic enzymes in heart and skeletal muscle (reviewed in Dickson 1995). Elevated activities of citrate synthase (which catalyzes the first reaction of the Krebs cycle), carnitine-palmitoyl transferase, and 3-hydroxyacyl-CoA dehydrogenase (rate-limiting enzymes in fatty acid oxidation) reflect the increased aerobic metabolic potential of scombrid fishes (Moyes et al. 1992). In addition, OXPHOS genes are differentially expressed between morphs of lake whitefish that differ in activity levels in their foraging behavior (Evans and Bernatchez 2012).

Alternatively (or in addition) to variation in gene expression, divergent energy demands may lead to adaptive evolution in the structure of specific OXPHOS proteins. Variation in OXPHOS proteins could influence the efficiency of ATP production by affecting how tightly electron transport and proton pumping are coupled in the electron transport chain. Modifications in the structure of OXPHOS complexes I-IV caused by amino acid substitutions in constituent proteins could affect "slip reactions," resulting in more or fewer protons pumped by the electron-transport-chain for each electron pair transferred (the $H^+/2e$ ratio) (reviewed in Brand 2005). Alternatively, substitutions in proteins of ATP synthase could modify the amount of ATP made by the ATP synthase for each proton driven through that protein (the H^+/ATP ratio) (Brand 2005). These

phenomena are consistent with evidence of positive selection in the *MT-CO2* gene of the high performance billfishes (Dalziel et al. 2006) as well as in the *MT-ND2* and *MT-ND5* genes of some migratory Pacific salmon (Garvin et al. 2011).

Here, we investigated patterns of adaptation on OXPHOS genes in a diverse group of fish taxa with different swimming performances. We hypothesized that positive natural selection would affect OXPHOS efficiency among divergent lineages with long-term differences in ATP demand. Thus, molecular adaptation was predicted to be associated with locomotion intensity (speed, duration and frequency). We examined evidence for positive selection on all mitochondrial OXPHOS genes from six fish groups that can be classified into three different swimming performance categories. Tunas and billfishes represent high performance swimmers, mackerels and jacks represent moderate (or moderate-high) performance swimmers, and flatfishes and seahorses/pipefishes represent low performance swimmers. These categories are based on general locomotion patterns in which pelagic fishes exhibit highly aerobic locomotion with greater endurance than sedentary fishes, and among the pelagic fishes, tunas and billfishes may maintain the highest speeds for the longest duration (Lauder, et al. 2006). Recent phylogenetic analyses (Betancur-R et al. 2013; Miya et al. 2013) indicate that these taxa are arranged into two monophyletic groups each containing representatives of all three performance classes (Figure 1). In contrast to previous studies that focused on a single lineage, this broad phylogenetic sampling allowed us to compare levels of selection across a wide range of locomotion performance. We also inferred functional significance from the position of positively selected amino acid sites in the 3-dimensional structure of specific enzyme complexes. Our results indicate that

selection on OXPHOS genes is indeed associated with divergent swimming habits among these fishes. However, selection is heterogeneous over evolutionary time and positive selection is not strictly associated with high performance taxa. Thus, we provide new insights on the evolution of swimming diversity in fishes and the adaptation of OXPHOS genes relative to organismal energetic performance across a broad phylogenetic scale.

Materials and Methods

Phylogeny reconstruction

We acquired sequences of 13 mitochondrial protein genes (all of them encode subunits in OXPHOS) from the representative species of each fish group from the National Center for Biotechnology Information (NCBI) GenBank (Supplementary material: Table S1). To complement the mitochondrial genes, we added 7 nuclear genes, that are commonly used in fish phylogenies (obtained from NCBI GenBank), or were developed as part of the Fish Tree of Life project (Li, et al. 2007). These genes are *rag1* (recombination activating protein 1), *rag2* (recombination activating protein 2), *rhodopsin*, *tmo4c4* (anonymous, see (Streelman and Karl 1997)), *zic1* (zinc finger protein 1), *myh6* (myosin, heavy chain 6), and *btbd7* (BTB domain containing 7). For species without available sequences, we designed primers (Supplementary material, Table S2) and amplified and sequenced specific genes via standard polymerase chain reaction (PCR). PCR products were sequenced in both forward and reverse directions using ABI BigDye terminator chemistry and an ABI Prism 3130 XL Genetic Analyzer. Sequences have been deposited in GenBank (Supplementary material, table S3). Particular genes for a few species could not be amplified. In such cases, sequences were

obtained from congeners, yielding 19 “chimaeric” individuals. The 20 genes (13 mitochondrial and 7 nuclear) were concatenated for phylogenetic analyses (Gillespie 1991).

Maximum likelihood trees (with 1000 bootstrap replicates) were estimated in the program RAxML v.7.0.4 (Stamatakis 2006). Bayesian phylogenetic analysis was performed with MrBayes v.3.1 (Huelsenbeck et al. 2001; Ronquist and Huelsenbeck 2003). Swimming performance states were reconstructed from extant taxa with parsimony and likelihood using outgroups from (Betancur-R et al. 2013) in Mesquite (Maddison and Maddison 2001).

Analysis of patterns of natural selection

Sequences of 13 mitochondrial OXPHOS genes were used to examine patterns of natural selection. We first ran a series of random sites models (M0, M1a, M2a, M3, M7, M8a and M8) implemented in CODEML of PAML v.4.7 (Yang 2007). Likelihood ratio tests (LRTs) were conducted on the likelihood values produced by specific pairs of models: M3 vs. M0, M2a vs. M1a, M8 vs. M7, and M8 vs. M8a. Next, we ran branch-site models on designated branches with three different starting ω values (0.1, 1.0, and 4.3). We had no prior knowledge of which branch(es) could have experienced positive natural selection (except for branches leading to tunas and billfishes); therefore, we designated one branch (from b1-b25 in Figure 2) as the foreground branch in each test. LRTs were performed to determine if the more complex model A is significantly better than the null model. In the branches that showed evidence of positive selection, we used Bayes empirical Bayes (BEB) to calculate the probability of amino acid sites being under positive selection.

The 13 mitochondrial gene sequences were also analyzed using TreeSAAP (Woolley, et al. 2003), which measures the selective influences on 31 structural and biochemical amino acid properties, and performs goodness-of-fit and categorical statistical tests. We used a sliding window size of 20 amino acids (Supplementary material, Figure S2). Because of the large number of sites under positive selection, we only show sites with significance when its $p < 0.001$ for clarity. Then we used a sliding window size of 1, where one amino acid site is defined as being positively selected as long as one of the 31 properties is showing significance.

Mapping positively selected amino acid sites onto 3-dimensional (3D) crystal protein structure

The *MT-ND* genes that encode subunits in OXPHOS complex I are considered highly conserved between prokaryotes and vertebrates (Walker 1992). We inferred the potential function of positively selected amino acid sites belonging to *MT-ND* subunits (in complex I) by mapping them onto the 3D crystal structure of *Thermus thermophiles* (PDB ID: 4HEA), a Gram negative eubacterium (Baradaran et al. 2013). We mapped the positively selected sites of *MT-CYB* onto chicken bc_1 complex C chain (PDB ID: 1BCC) crystal structure because a study suggests the structure of the catfish *MT-CYB* protein resembles that of chickens (Singh et al. 2012). Similarly, we mapped the positively selected sites onto bovine *MT-CO1-3* subunits 3D structure (PDB ID: 1OCC) because of the highly conserved structure from bacteria to bovine (Schriefer and Hale 2004; Tsukihara et al. 1996). All the mapping was conducted using Geneious pro (v.7.0).

Positive natural selection analysis on nuclear OXPHOS genes

Fresh tissues from *Gnathanodon speciosus* (pilotfish), *Hippocampus zoestrae* (dwarf seahorse), and *Paralichthys lethostigma* (southern flounder) were collected and preserved in RNAlater RNA stabilization reagent (QIAGEN). Although we hoped to include representatives of billfishes and mackerels, fresh tissues from which to extract intact RNA were not available at the time of these analyses. Total RNA from each species was extracted using RNeasy Mini Kit (QIAGEN) and sent to the Oklahoma Medical Research Foundation Genomics facility for Illumina HiSeq 2500 next-generation transcriptome sequencing. We also obtained raw next-generation transcriptome sequences from NCBI Sequence Read Archive (SRA) of *Hippocampus hippocampus* (short-snouted seahorse) (NCBI accession number: SRR1324964), *Hippocampus kuda* (spotted seahorse) (NCBI accession number: SRR1324966), and *Cynoglossus semilaevis* (tongue sole) (NCBI accession number: SRR365042), as well as a *Thunnus orientalis* (Pacific Bluefin tuna) shotgun assembly sequences from 454 GS Titanium (NCBI accession number: BADN01000001-BADN01133062).

An average of 29,000,000 paired end reads (100 base pair) were generated from each species. We first ran the quality control program FastQC (Babraham Bioinformatics) on Illumina raw sequences and then used the read trimming tool Trimmomatic (Bolger et al. 2014) to remove low quality sequence regions. Several programs perform *de novo* transcriptome assembly, although most of them are based on the same computational principle, de Bruijin graphs (Bruijn 1946). We tried three different *de novo* assembly assemblers, including Oases (Schulz et al. 2012), Ray (Boisvert et al. 2012), and Trinity (Grabherr et al. 2013), each under several different k-mer values. We chose to use the results based on Trinity because this assembler

generated relatively more OXPHOS contigs. We included sequences from 15 nuclear OXPHOS genes that were present in all 10 target species (corresponding sequences from *Danio rerio*, *Gasterosteus aculeatus*, and *Takifugu rubripes* as reference sequences). These genes are: *atp5j2, atp5l, atp6ap1a, atp6v0d1, atp6v1d, atp6v1e1b, cox15, cox4i1, cox6a1, cyc1, ndufb5, ndufb8, ndufc2, ndufv1,* and *rieske.* Sequences were aligned and tested for positive selection on branches b1-b9 (Figure 6) using the same random sites models (M0, M1a, M2a, M3, M7, M8a and M8) and branch-site models as for the mitochondrial genes.

Results

Phylogeny and ancestral state reconstruction

Both maximum likelihood and Bayesian analyses generated the same topology for the species with mitochondrial genes (Figure 2) regardless of partition scheme with relatively high bootstrap support and posterior probabilities (see Supplementary material, Figure S1). The recent comprehensive phylogenetic analysis of Betancur-R et al. (2013) resolved the phylogeny of bony fishes to the level of taxonomic family, while a study by Miya et al. (2013) focused on relationships of several pelagic fish groups. Our results are consistent with both of these studies for the taxa in common to both. Specifically, we recovered two monophyletic groups, each containing taxa with low, moderate, and high swimming performance. In one group, the tunas and mackerels (Scombriformes) are sister to the seahorses and pipefishes (Syngnathiformes), while the series Carangaria contains the billfishes (Istiophoriformes), jacks (Carangiformes), and flatfishes (Pleuronectiformes) (Betancur-R et al. 2013) see also Phylogenetic Classification of Bony Fishes-Version 3: www.deepfin.org).

Swimming performance states were reconstructed from extant taxa with parsimony and likelihood using outgroups from (Betancur-R et al. 2013). These results of reconstructed swimming performance suggest that the ancestors of each major group were low performance swimmers (Figure 1), although incomplete taxon sampling of related groups could mislead the reconstructions. Even if the ancestors were moderate performance swimmers, it is clear that high performance swimming evolved independently in the tunas and billfishes.

Analysis of positive selection

Positive selection under site models cannot identify particular branches where positive selection has occurred, but the models can detect positive selection among sites where it occurs as long-term trends or among multiple branches separated on the tree. For the 13 mitochondrial genes, a significant difference was detected between model M0 and M3, which suggests ω is variable among sites (Table 1a). We used three model pairs to test for positive selection: M1a vs. M2a, M7 vs. M8, and M8a vs. M8. Neither M2a-M1a nor M8a-M8 showed significant differences. However, M8-M7 showed a significant difference, suggesting some positive selection signal on certain sites somewhere on the tree. For the 15 nuclear OXPHOS genes, a significant difference was also detected between model M0 and M3 (Table 1b). None of the three model pairs showed significant differences.

Table 2a lists the resulting likelihood values, likelihood ratio tests (LRTs), and estimated model parameters for each branch examined under branch-site models on 13 mitochondrial genes. We performed a false discovery rate analysis (Benjamini 2010) on these results where we performed 20 tests and recovered 12 positive results. All 12

positive results had a q value of 0.03 or less, and the probability of a false positive among them is 0.36. Because this is less than 1.0, no false positives are expected. The number of amino acid sites on each branch inferred to be under positive selection via Bayes Empirical Bayes is provided in the Supplementary material Table S4 (the identity of these sites and more details are provided in Supplementary material, Table S5). These results showed substantial differences in natural selection on lineages leading to taxa with different swimming performance. We found significant evidence of positive selection on the ancestral lineages of all three performance classes (branches b2, b17, b3, and b4 in Figure 2), lineages leading to moderate-performance swimmers (b6 and b13), and low-performance swimmers (b16, b23, b24, b25, and b11). Conversely, strong purifying (negative) selection was identified on lineages of high-performance swimmers (b7, b8, b9, b10, b14, and b18). Fourteen sites that appeared to be positively selected occur on multiple branches; however, they are not associated with branches leading to a particular performance group (see details in Supplementary material: Table S5).

Similar patterns of selection were found on the 15 nuclear OXPHOS genes: there was no evidence of positive selection on high-performance tunas (b5, Figure 6) and there was evidence of positive selection on the branch (b4, Figure 6) leading to seahorses and the branch (b6, Figure 6) leading to flatfishes and jacks. Table 2b lists the resulting likelihood values, likelihood ratio tests (LRTs), and estimated model parameters for each branch examined under branch-site models on the 15 nuclear OXPHOS genes.

The above tests assess the strength of the evidence for positive selection rather than the strength of selection itself. An indicator of the strength of positive selection on a branch can be obtained from the product of the proportion of sites having $\omega > 1$ and the estimated ω value for those sites. In Figures 2 and 6, the branch width is shown proportional to the strength of positive selection based on this measure (except for b12, b23 and b24 in Figure 2 which had extremely high ω and a cap of $\omega = 50$ is used). Among those branches harboring sites under positive selection, branches ancestral to more than one swimming class (b2, b3, b4, b12, b17) and branches leading to low-performance fishes (b11 and b23) exhibited greater selection than branches leading to moderate-performance fishes (b6 and b13), and substantially more than any branches associated with high-performance fishes. This is counter to the notion that high-performance swimmers have been most strongly influenced by positive selection for enhanced OXPHOS performance assuming enhanced OXPHOS efficiency is the major contributor to high performance. However, there are alternative contributors to high performance such as increased mitochondrial density per tissue mass, more closely packed inner mitochondrial membrane cristae, increased metabolic enzyme activity, and increased expression of genes involved in a number of biological pathways such as glycolysis, protein biosynthesis, and cytoskeletal structure (*e.g.*, tuna tissues had as much as 30–80% more mitochondrial protein per gram of tissue than carp; cristae packing value of 63-70 m^2/cm^3 in tuna red muscle is greater than the range reported for skeletal muscles of Antarctic fish, and a wide variety of mammals, hummingbird flight muscle, and mammalian and reptilian hearts; activities of carnitine palmitoyl trasferase per milligram of protein were 2-2.5 times higher in tuna red muscle and ventricle

mitochondria than in carp mitochondria from the same tissue; Moyes et al. 1992). Under any of the latter scenarios, increase in OXPHOS efficiency might not be responsible for high performance.

The physico-chemical properties analysis as implemented by TreeSAAP (Woolley et al. 2003) does not provide inferences about particular branches on the phylogeny, but does provide information about changes of particular amino acid sites across the whole tree. A large number of amino acid sites (in mitochondrial OXPHOS genes) were identified as having substitutions with significant physico-chemical differences; these data are presented in Supplementary material: Figure S2 and Table S4. A mitochondrial genome-wide (only the protein-coding genes) sliding window analysis (window size = 20 amino acids) revealed that the vast majority of windows contained substitutions that exhibited between 0 and 15 properties with significant differences (Supplementary material: Figure S2). Windows with substitutions exhibiting between 15 and 30 significant properties occurred at much lower frequency. Although windows with the greatest number of significant property differences were observed in the genes for *MT-ND1, MT-CO1*, *MT-ND4* and *MT-CYB*, there were no other apparent patterns of variation in the distribution of such sites across the genome (Supplementary material: Figure S2). Positively-selected sites identified by both methods are summarized in Supplementary material: Tables S4 and S5.

The two lineages, seahorses and flatfishes, on which positive selection was most pronounced, have also experienced extraordinary morphological evolution (Cairns et al. 1988; Thaler 1994). Thus, it is possible that the positive selection signal detected in mitochondrial OXPHOS genes in these lineages is unrelated to energy demands, but is

simply a consequence of rapid genomic evolution in these groups. The probability of nucleotide change is indeed higher in these two lineages than for the other major groups as indicated by five nuclear genes that are not directly involved in OXPHOS system (*rag1*, *rhodopsin*, *tmo4c4*, *mhy6*, and *zic1*). However, a positive selection signal was found for only two genes, *rag1* and *rhodopsin*, on one branch leading to flatfishes (LRTs = 72.53, ω = 50.88, three sites 199, 425, and 605 under selection with Bayes Empirical Bayes (BEB) probability of 0.684, 0.501, and 0.958, respectively). Because there was no consistent pattern of selection on these non-OXPHOS genes, the positive selection detected on mitochondrial genes likely does not reflect a genome-wide pattern of divergence but may be related to adaptation of OXPHOS efficiency.

Structural position of positively selected sites

The position of particular amino acids in the tertiary and quaternary structure of a protein may allow inferences about the function of individual residues. In particular, those sites near the catalytic core or other functionally important regions, or those in physical proximity (likely to interact with) other amino acids, would seem most likely to influence protein function. The analysis presented here was conducted on mitochondrial OXPHOS genes (similar analysis will be conducted after adding more species for nuclear genes).

Complex I performs the first step, and is the largest and most complicated enzyme complex, in the OXPHOS pathway. It catalyzes the transfer of two electrons from NADH to ubiquinone (Q), coupled to the translocation of four protons across the inner mitochondrial membrane. It is also a major source of reactive oxygen species in mitochondria. Complex I exhibits an L-shaped architecture with a membrane arm and a

hydrophilic peripheral arm that protrudes into the mitochondrial matrix (Figure 3a). The membrane arm consists of 7 core mitochondrial NADH dehydrogenase (*MT-ND*) gene encoded subunits.

Site 12, one of the sites identified in subunit ND1 as being positively selected (Figure 3b), is included in the region associated with MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes)/DEAF enhancer/hypertension (Campos et al. 1997) and sudden infant death (Opdal et al. 2002). The other positively selected sites, 172 and 192, are in the alpha helix close to a critical site, 184, that is associated with adult onset dystonia (Simon et al. 2003) (in yellow in Figure 3b).

Subunit ND5 contains an unusual structural element, the helix HL (see enlarged Figure 3c, indicated in yellow), that extends nearly the entire length of the membrane domain and coordinates conformational changes. On the opposite side of the membrane domain, a series of β-hairpins (βH element) (indicated in white in Figure 3c, d) from neighboring subunits contribute to conformational changes and stability of the complex. Positively-selected sites include 515 [ND5], which is adjacent to helix HL; 53 [ND5] and 54 [ND4], which are in βH elements; 63 [ND5], which is in the seven-residue loop connecting two βH elements; and 86 [ND2] (not shown here) and 71 [ND5], which are in regions (including residues 83 [ND2] and 88 [ND5]) demonstrated to have significant negative effects on function (Efremov and Sazanov 2011).

In addition to the sites listed above, there are several sites that appear to be under positive selection, yet their location in the molecular structure does not provide any clear suggestion of functional significance. These residues could be involved in

stabilizing the tertiary or quaternary structure of the various multi-subunit complexes or facilitate their assembly. These include residues 199 [ND5], 269 [ND5], 425 [ND5], 515 [ND5], 600 [ND5], 20 [ND4], 86 [ND2], 168 [ND2], 359 [ND2], 360 [ND2], 408 [ND2], 23 [ND4L], 4 [ND6], 8 [ND6], 11 [ND6], 90 [ND6], 138 [ND6], 94 [ND3] and 98 [ND3] (not shown in figures for clarity).

Complex III is an intermediate component of the respiratory chain, which transfers electrons from reduced ubiquinone to cytochrome c, coupled to proton translocation across the mitochondrial membrane (Scheffler 1999). The *MT-CYB* gene encoded cytochrome b forms the active redox center: a cavity surrounded by the transmembrane helices A (residues 33-54), D (172-204), and E (221-245), and the amphipathic surface helix a (65-72) (indicated in white in Figure 4). Residues 221 and 194 are also close enough to contact the active site inhibitor (Zhang et al. 1998). The positively selected sites 194 and 235 (indicated in cyan in Figure 4) are within these regions, suggesting important functional effects.

The pocket bound by stigmatellin, one OXPHOS inhibitor, is formed by the end of helix C, the helix cd1, the helix ef linker, and the end of helix F. Specific residues of known importance include 271, 275, 125-129, 138-153 (indicated in green in Figure 4). The positively selected sites 126 and 297 are close to this area.

Complex IV is the terminal component of the respiratory chain, in which electrons received from cytochrome c reduce molecular oxygen to water and protons are pumped into the intermembrane space. Six positively selected sites were identified in complex IV: 178 in subunit COX1, 54 and 187 in subunit COX2, and 55, 155, and 171 in subunit COX3 (Figure 5a).

Subunit I (encoded by *MT-CO1* gene) is largely embedded in the membrane with three redox centers: heme a, heme a_3, and Cu_B. One of the Cu_B liganded residues is 240. It has been suggested that residue 244 and 240 are close enough for binding interaction, thereby forming an unusual cross-link structure (Richter and Ludwig 2003b). One positively selected site 178 is very close to 244 and 240 (Figure 5b). Two proton translocation pathways, D-pathway and K-pathway, are identified in subunit I. Three key residues of D-pathway 91, 98, and 242, and one key residue of K-pathway 319 are well defined from mutagenesis studies (Richter and Ludwig 2003b) (key residues are indicated in white in Figure 5b). The positively selected site 178 (indicated in cyan in Figure 5b) is close to (within 30 Å) of the four key residues and could potentially affect proton translocation.

Subunit II (encoded by *MT-CO2* gene) has two trans-membrane helices and a hydrophilic beta strand extending to the extra-membrane domain, housing the Cu_A center. The first residue crucial for electrons to enter the oxidase complex cytochrome c is 106 (indicated in magenta in Figure 5c). The other Cu_A liganded residues identified through mutagenesis include 196, 200, 198, 161, 204, and 207 (indicated in magenta in Figure 5c). One positively selected site 187 is very close to the Cu_A center and the other positively selected site 54 is in a linking strand between two helices of the entry site which includes critical residue 62 (indicated in cyan in Figure 5c).

Subunit III (encoded by *MT-CO3* gene) is fully embedded in the membrane domain. No key residue of this subunit has been identified through mutagenesis. However, it has been shown that this subunit stabilizes the integrity of the binuclear

center in subunit I; for example, when this gene is deleted, only a partially assembled complex results (Richter and Ludwig 2003b).

The sites described above as having potential functional significance were more or less uniformly distributed among branches of the tree and showed no association with particular swimming performance groups. Many sites exhibited evidence of being under positive selection yet did not appear in structural locations that would suggest special functional importance. However, the interaction between amino acids among different proteins or among sites in the same protein is known to have significant functional effects (*e.g.,* Altringham and Johnston 1990; Willett and Burton 2003). Such interactions could affect protein assembly, stability or specific function. We used 4 Å as the nominal upper limit for weak interactions between amino acid sites as described in Schmidt et al. (2001). None of the sites identified here as under positive selection were found to be within 4 Å of any other amino acid sites.

Discussion

We examined evidence for adaptation of mitochondrial and nuclear OXPHOS genes in fishes with different swimming performance. Selection was investigated for specific amino acids sites across the whole phylogenetic tree for these species, as well as for amino acid sites on individual branches of the tree. The results show a strong signal of positive selection on branches in the ancestral parts of the tree, and branches leading to low- and moderate-performance swimmers. However, no evidence of positive selection was observed within clades of the high-performance tunas or billfishes. The same patterns were also observed on 15 nuclear OXPHOS genes on a smaller set of taxa (tissues or sequences from only 10 species were available). We did

not observe a disproportional effect of selection on any particular gene, as all genes exhibited some positively selected sites, but this varied across branches of the tree. We also showed that many of the sites identified as being under positive selection, occurred in structural regions where they were likely to have effects on OXPHOS function. Positively selected sites in other regions may also have functional significance, but their potential effects are less clear.

Our results show the strongest signal of positive selection on branches leading to the lowest performance fishes, while purifying selection was identified on the branches of high performance fishes. This suggests that a fairly efficient OXPHOS system had evolved under positive selection in the ancestors of the two major groups. Selective modification of OXPHOS on branches leading directly to tunas and billfishes may have further facilitated the evolution of high performance swimming. But once high performance swimming had evolved in these lineages, purifying (negative) selection appears to have predominated in the OXPHOS system as it existed at that time. Conversely, substantial modification due to positive selection occurred in the lower performance lineages of flatfishes and the seahorse/pipefish group. The moderate-high performance jacks and mackerels exhibited moderate-high conservation, with limited positive selection. Thus, the strength of the positive selection signal is inversely proportional to swimming performance among taxonomic groups.

A common expectation is that positive selection should lead to enhanced organismal performance. Moreover, it might be expected that positive selection will lead to increased functional efficiency of OXPHOS in response to the increased ATP demands associated with enhanced performance. In the absence of information on the

exact functional significance of individual substitutions, the effects of positive selection to increase or decrease OXPHOS efficiency in these taxa remains unknown. However, our results do not match simple expectations. Positive selection appears to be heterogeneous, fluctuating over phylogenetic time scales, with no simple relationship between the strength of selection and organismal performance. However, one clear pattern is that there is strong functional constraint (negative selection) in high performance systems.

We speculate that once a reasonably efficient OXPHOS system evolved it may have become difficult to change in the high performance groups. The high performance system might be expected to have a much lower tolerance for non-synonymous substitutions as most (even slight) changes would be likely to have negative functional effects. Conversely, lower performance swimmers may have a much broader tolerance for non-synonymous substitutions because the OXPHOS system is under much lower performance demands. We envision the high performance swimmers as occupying a local optimum on a fitness landscape (Block et al. 1993), but their OXPHOS system is so fine-tuned that substitutions that would allow them to cross fitness valleys and reach higher peaks could be strongly deleterious. On the other hand, lower performance fishes might readily cross such valleys without significant fitness costs because OXPHOS efficiency will be less critical and they may then climb higher peaks due to positive selection.

It is also possible that organismal fitness in taxa with low energy demands may be increased by a modified OXPHOS regulatory system or even a reduction in OXPHOS efficiency. In such cases, we would expect to see evidence of positive

selection on OXPHOS genes in these taxa. Because there may be trade-offs between OXPHOS rate/efficiency and deleterious effects, reducing OXPHOS efficiency may be adaptive in systems where ATP demand is chronically low. For example, maintenance of a strong chemi-osmotic gradient in organisms with low ATP demand may cause increased production of reactive oxygen species (ROS) (Pegram et al. 2010) leading to oxidative stress (Block and Finnerty 1994). Therefore, reduction of the $H^+/2e$ ratio (increased slippage) due to altered protein structures could be adaptive in low performance species. OXPHOS regulation is highly complex and involves mechanisms independent of structural OXPHOS proteins. Elevated expression of OXPHOS genes and others involved in aerobic respiration could clearly increase aerobic capacity in the absence of selection on specific OXPHOS variants. However, it is clear that positive selection acts on OXPHOS proteins and the effects of positive selection are frequently associated with the evolution of differences in ATP demand.

Our results are consistent with previous studies, where positive selection or evolutionary rate variation was found on lineages leading to organisms with high ATP demands. For example, *MT-CO1* (Andrews and Easteal 2000) and *MT-CO2* (Adkins and Honeycutt 1994) exhibited accelerated d_N in the lineage leading to hominids that appears to be associated with increased brain size. Grossman et al. (2001) found accelerated rate variation for *COX4*, a nuclear gene in OXPHOS complex IV in catarrhine ancestors of hominids in the period between 18 and 40 Mya and then decelerated rate variation in the descendant hominid lineages. On the lineage leading to bats, the only mammals capable of powered flight, eight OXPHOS genes were found to have undergone positive selection (Shen et al. 2010). Foote et al. (2011) found two

positively selected amino acid sites, which could influence overall metabolic performance, in the mitochondrial genes of killer whales (*Orcinus orca*). In addition, the *MT-ND2* and *MT-ND5* genes of highly migratory Pacific salmon exhibit evidence of positive selection (Garvin et al. 2011).

Dalziel et al. (2006) examined *MT-CO2* gene among high-performance fishes, including billfishes and tunas. They found ω was not increased in lineages leading to the tunas, but was significantly increased in the lineage preceding the billfish (including several amino acid sites). However, the phylogeny used in Dalziel et al. (2006) does not include the flatfish clade, which is now recognized as a close relative to the billfishes. Little et al. (2012) found positive selection along the branch leading to the billfish clade, when flatfishes were excluded, but no selection was detected when flatfishes were included. The latter result is consistent with our findings and inclusion of flatfishes appears to provide a more realistic phylogenetic context for the analysis of positive selection.

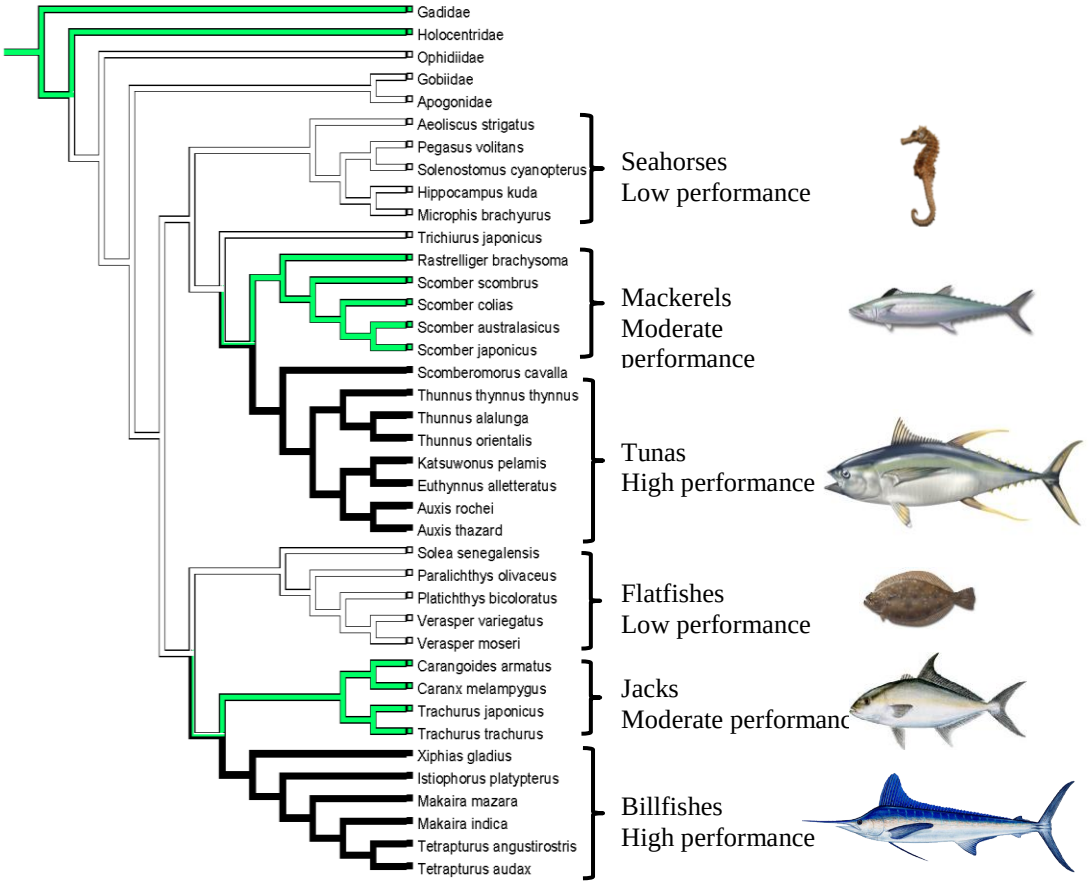
In conclusion, we found that patterns of natural selection on mitochondrial OXPHOS genes are complex among fishes with different swimming performance. The type and direction of selection are heterogeneous through evolutionary time and vary in ways that would not be readily predicted based solely on organismal performance. Although examination of the most recent lineages indicates that positive selection is inversely proportional to organismal performance, the broader pattern indicates a complex interplay between organismal adaptations, ATP demand and OXPHOS function.

Acknowledgments

This work was supported by NSF award DEB-0732988 (to REB). We thank KUNHM Ichthyology library for some of the fish tissues. We thank Z. Hao and T.E. Dowling for helpful discussions, and J.C. Cureton II, L.J. Weider, J.P. Masly, R. Knapp, and C.M. Lewis for comments on the manuscript.

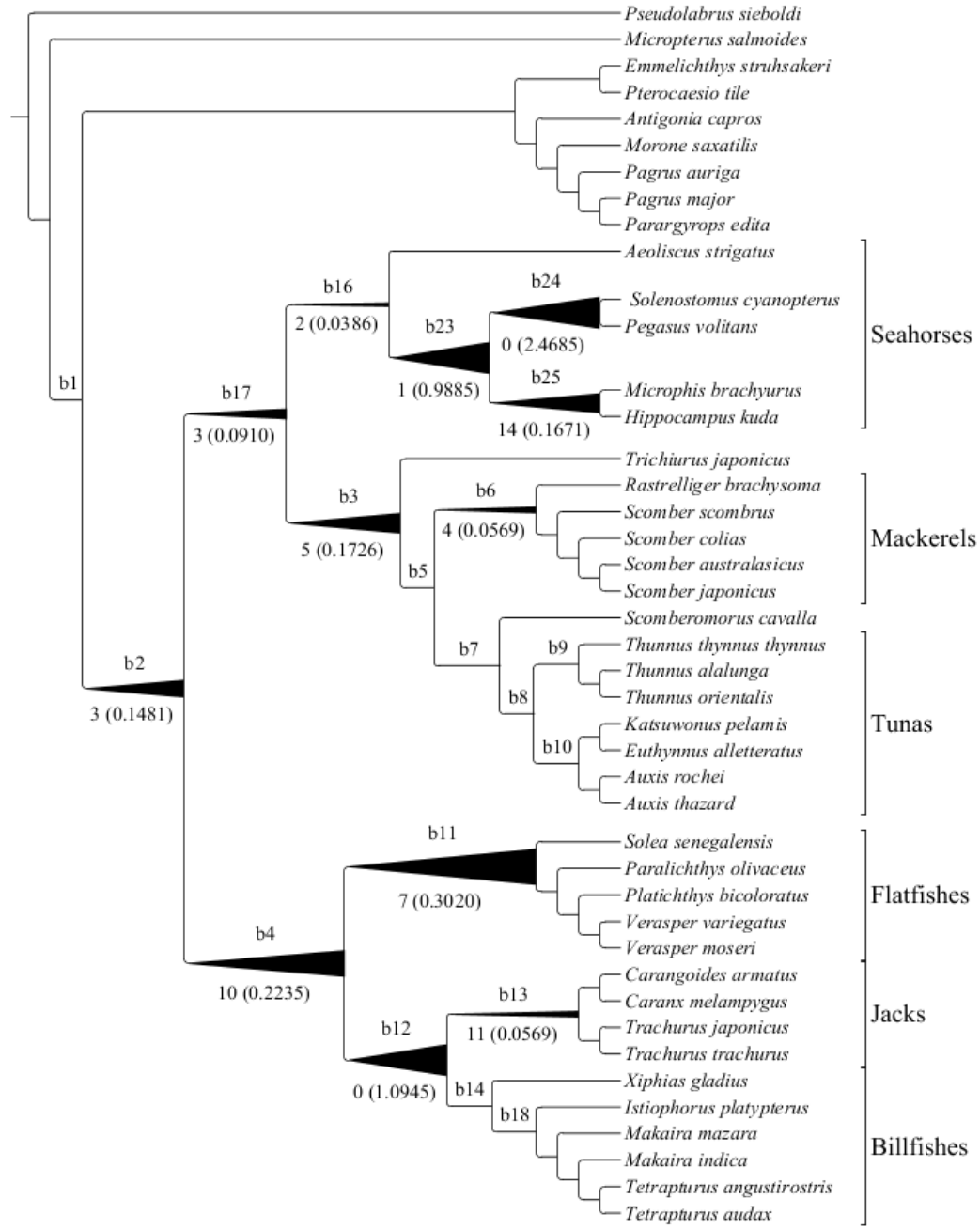
Figures and Tables

Figure 1 Phylogeny and swimming performance states of interested fish groups in this study.



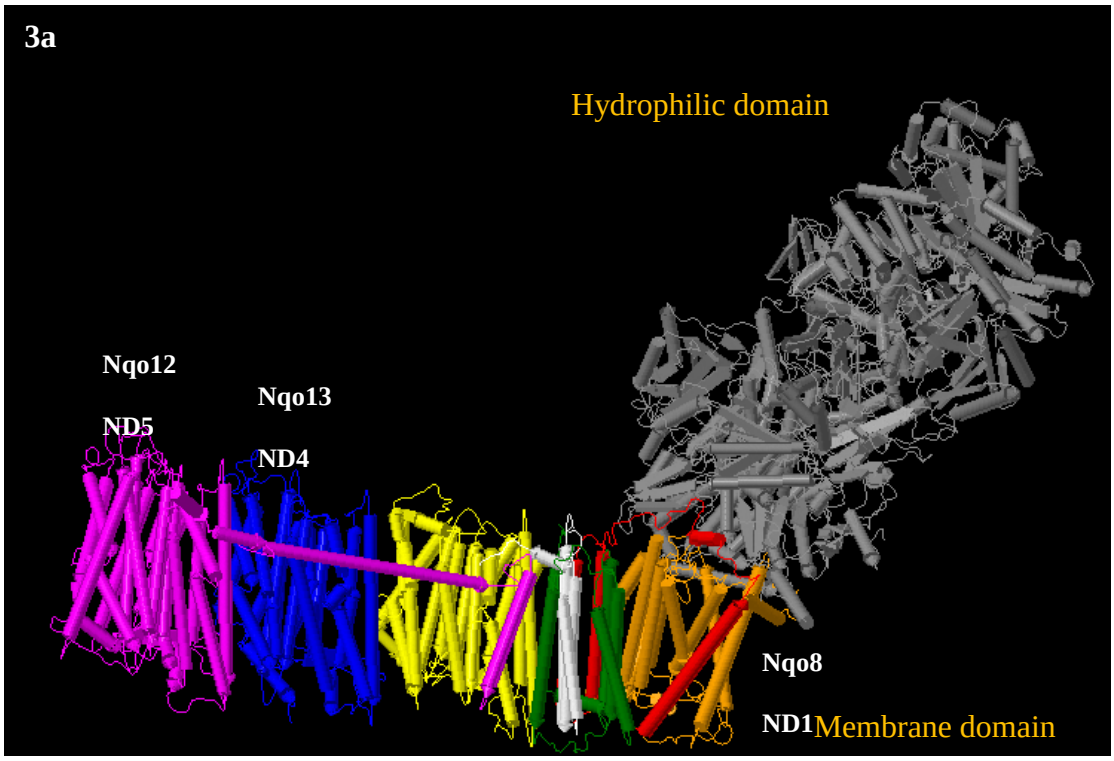
Note: Swimming performance was reconstructed with parsimony (shown), as well as maximum likelihood, which was qualititatively similar, with outgroups from Betancur-R et al. (2013). Branch colors represent performance levels as follows: white, low performance; green, moderate performance; black, high performance.

Figure 2 Positive selection analysis on mitochondrial genes on each branch (arbitrary labels, b1-b24, appear above each branch).

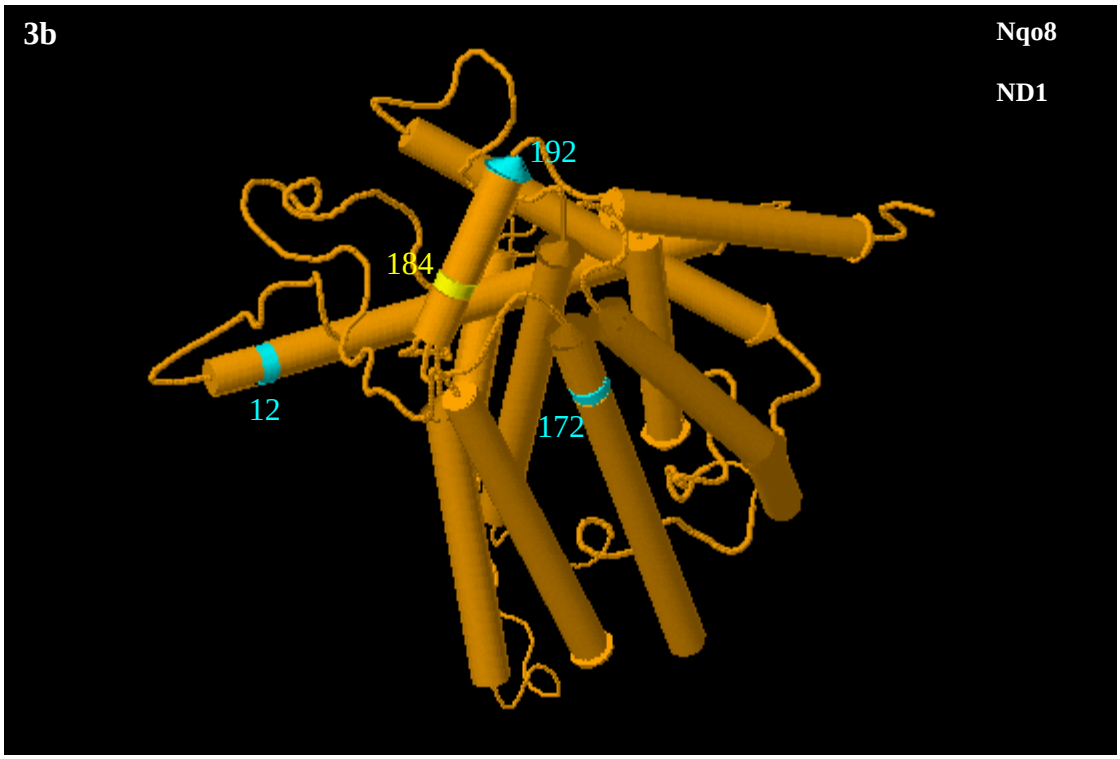


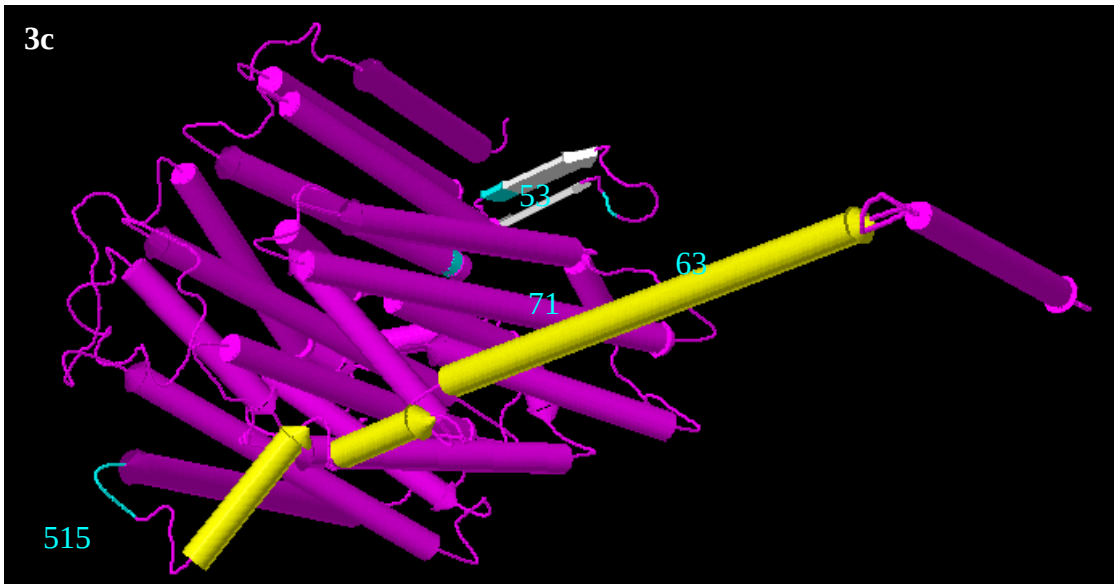
Note: The number below each branch represents the number of positively selected sites with posterior probability higher than 0.8. The number in parentheses is the product of proportion of sites having $\omega > 1$ and ω_2 for that branch, as an indicator of the strength of positive selection. The width of the branches is proportional to the strength of positive selection (because ω_2 for b12 and b24 is extremely high, the width was capped for ease of visualization). Labled branches without values listed have no evidence of positive selection.

Figure 3 Crystal structure of the entire respiratory complex I at 3.3 Å (PDB: 4HEA) from *Thermus thermophiles* front view (3a). Mitochondrial gene encoded subunits include Nqo8 (MT-ND1) (orange) (3b), Nqo14 (MT-ND2) (yellow), Nqo7 (MT-ND3) (red), Nqo13 (MT-ND4) (blue) (3d), Nqo11 (MT-ND4L) (white), Nqo12 (MT-ND5) (magenta) (3c), and Nqo10 (MT-ND6) (green) subunits.



Note: Amino acid sites in cyan are those with evidence of positive selection in this study. Structure in gray indicates the nuclear-gene encoded hydrophilic domain.


3b
Nqo8
ND1
192
184
12
172


3c
53
63
71
515

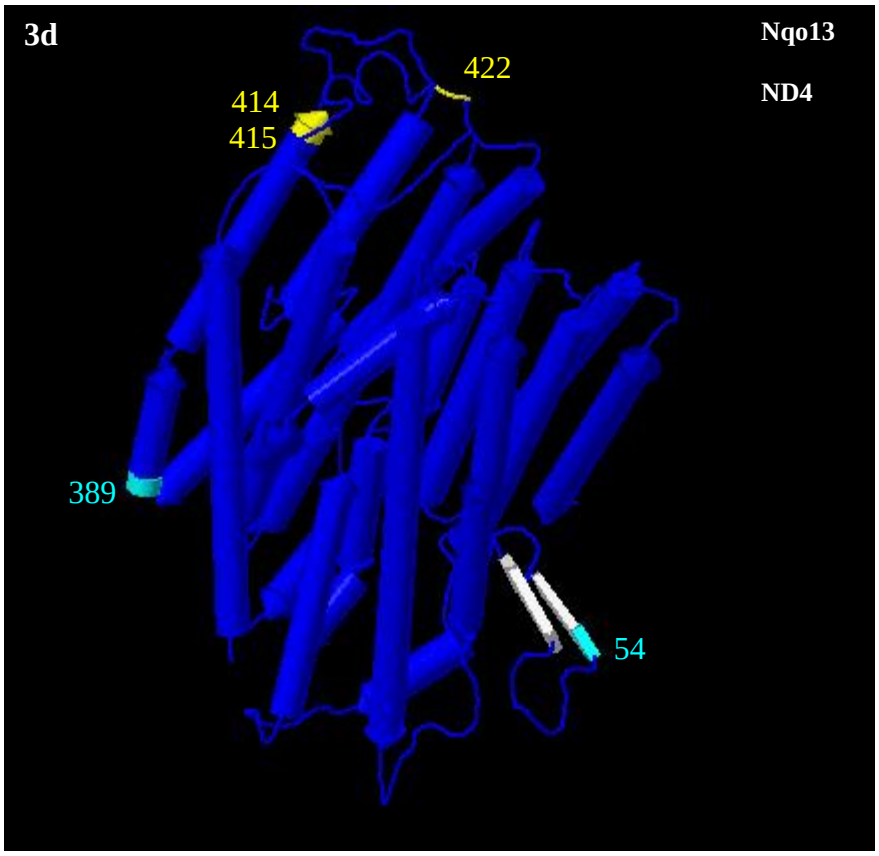
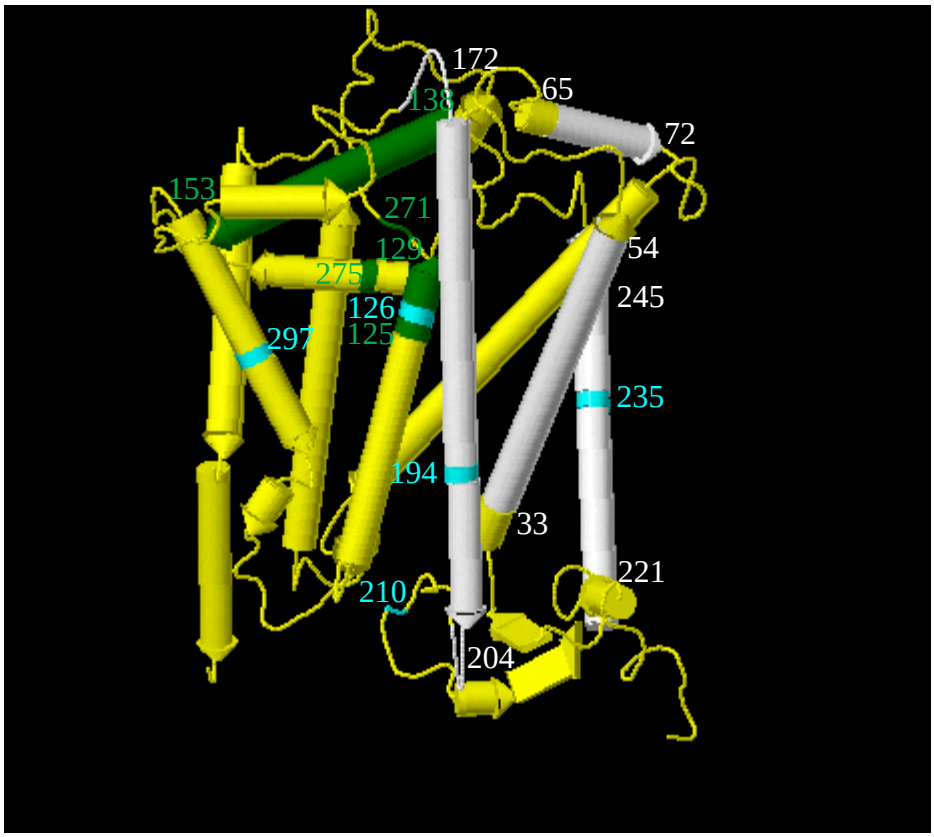
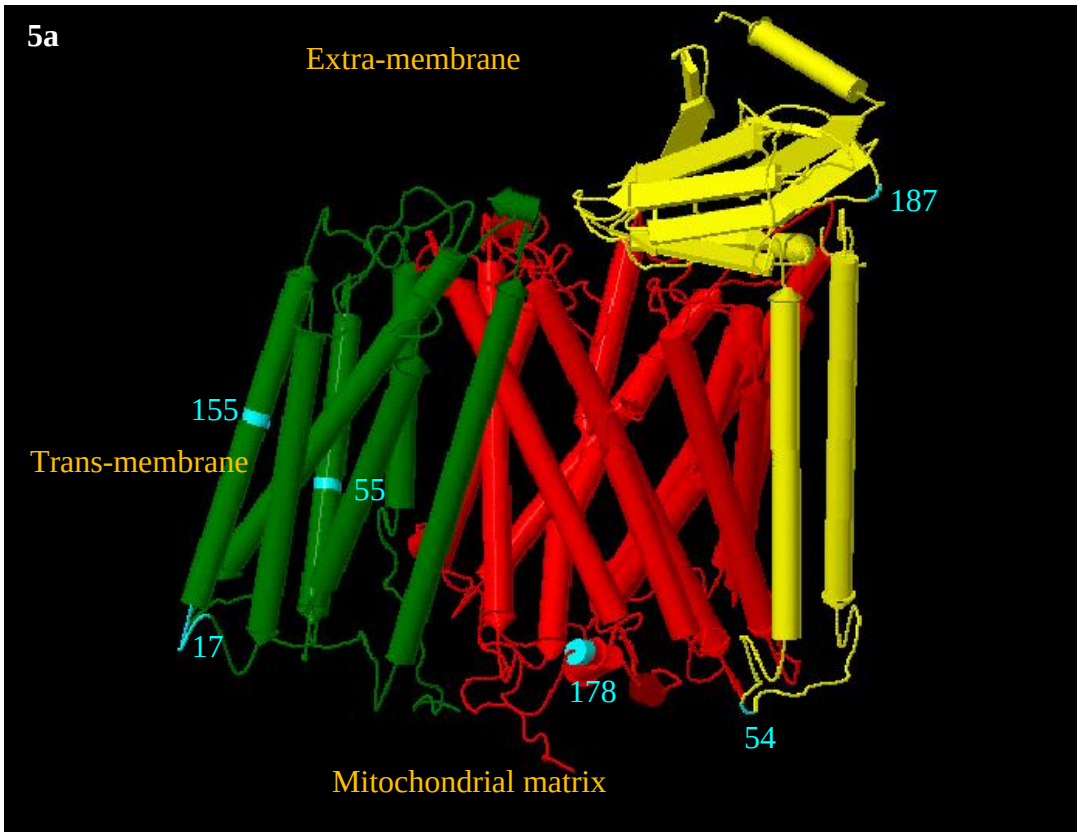

3d
Nqo13
ND4
422
414
415
389
54

Figure 4 Crystal structure of the entire respiratory complex III cytochrome b at 3.16 Å (PDB: 1BCC) from chicken (*Gallus gallus*).

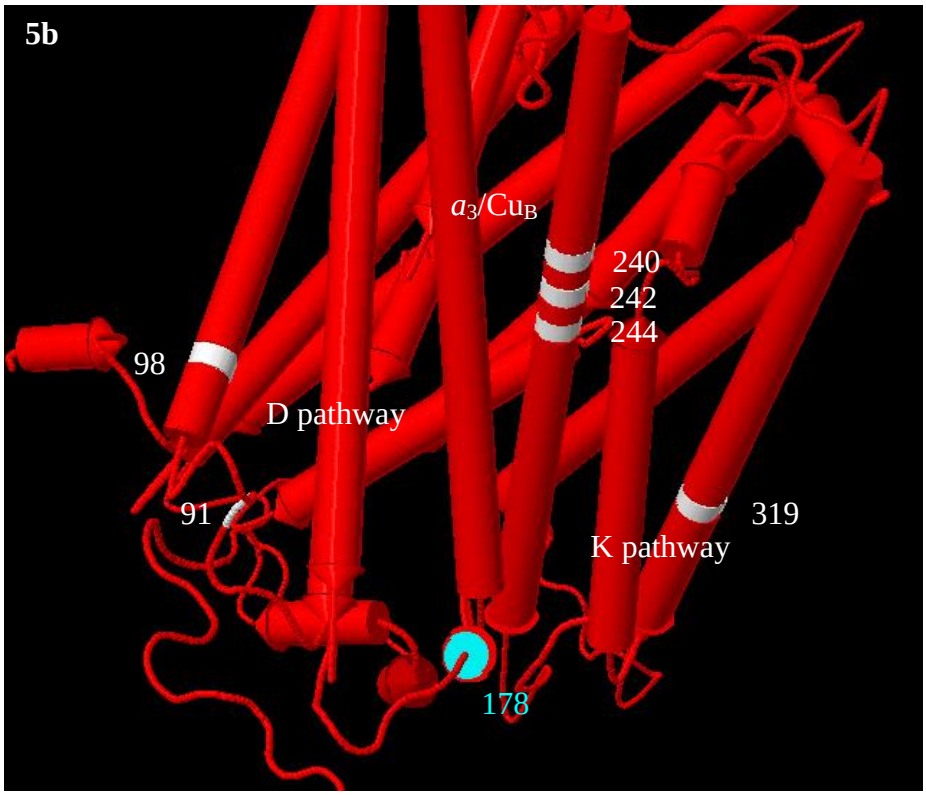


Note: White indicates antimycin-binding cavity formed by helices A (residues 33-54), D (172-204), E (221-245), and the amphipathic surface of helix A (65-72). Green indicates stigmatellin- and myxothiazol-binding pocket formed by residues 271, 275, 125-129, and 138-153. Cyan residues are those identified as positively selected amino acid sites in this study.

Figure 5 a) Structure of mitochondrial DNA encoded *MT-CO1* (red; also see in 5b), *MT-CO2* (yellow; also see in 5c), and *MT-CO3* (green) subunits from bovine heart cytochrome c oxidase at 2.8 Å (PDB: 1OCC).



Note: Amino acid sites in cyan are positively selected sites detected in this study.


5b
a_3/Cu$_B$
240
242
244
98
D pathway
91
319
K pathway
178

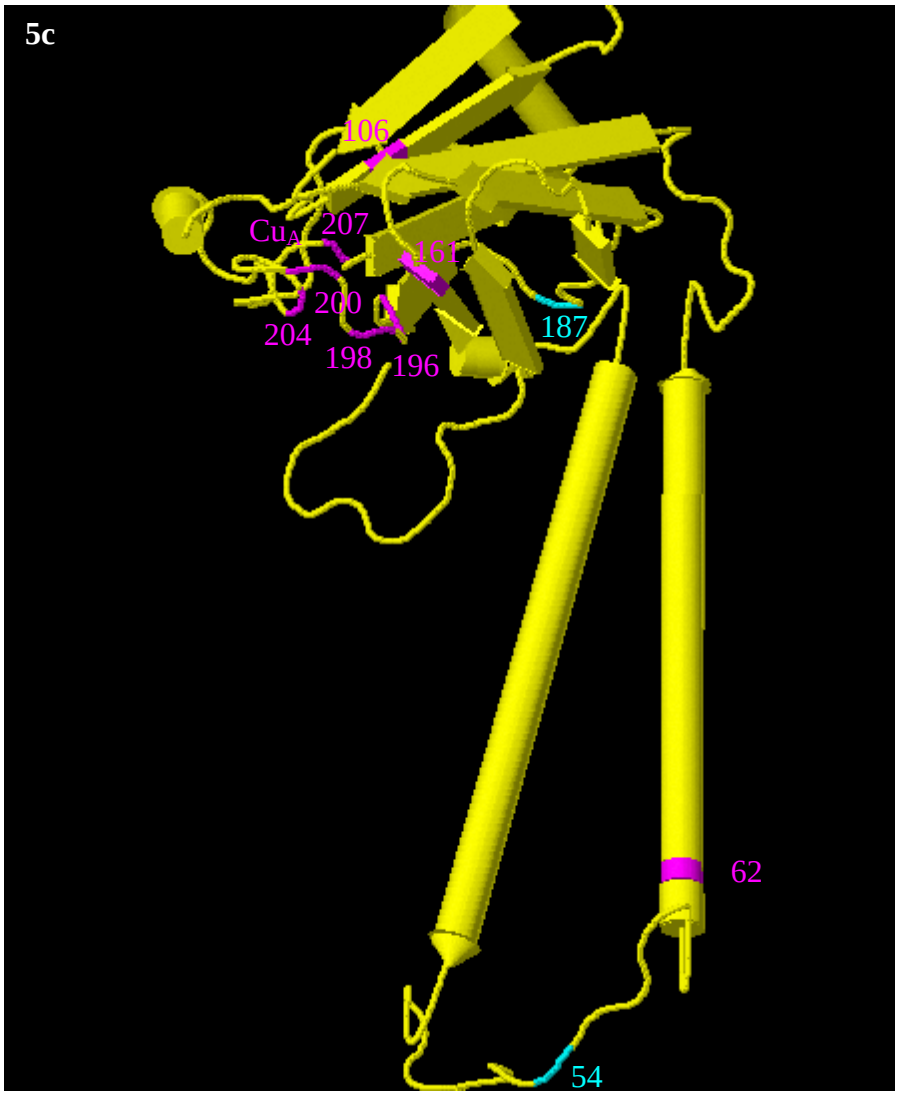
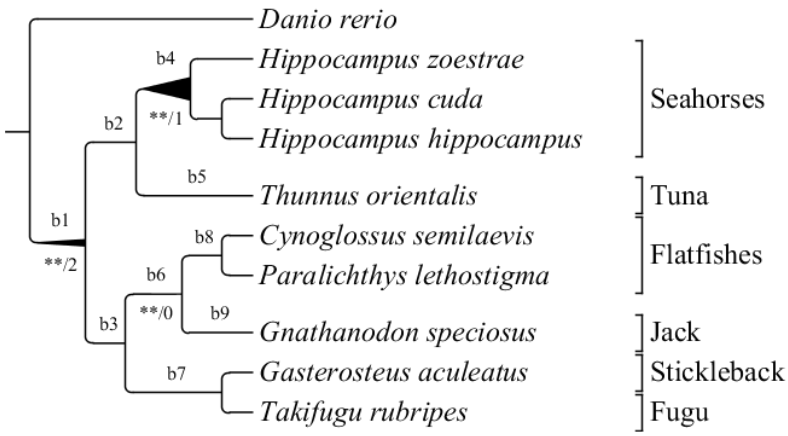

5c
106
Cu$_A$
207
161
200
204
198
196
187
62
54

Figure 6 Positive selection analysis on nuclear OXPHOS genes on each branch (arbitrary labels, b1-b8, appear above each branch).



Note: The number below each branch represents the number of positively selected sites with posterior probability higher than 0.8. The number in parentheses is the product of proportion of sites having $\omega > 1$ and ω_2 for that branch, as an indicator of the strength of positive selection. * represnts the presence of postive natural selection. The width of the branches is proportional to the strength of positive selection. Labled branches without values listed have no evidence of positive selection.

Table 1a Results of PAML random-sites models on 13 mitochondrial genes

Model	lnL	Parameters[a]			Null	LRTs	*p*
		ω_0/p	ω_1/q	$\omega_2/\ \omega_p$			
M0	-191327.6225	0.03421					
M1a	-189631.0183	0.02838(94.9%)	1(5.1%)				
M2a	-189631.0190	0.02838(94.9%)	1(5.1%)		M1a	0	
M3	-184769.7796	0.00609(72.6%)	0.11443(27.4%)	79.11600(0)	M0	13115.6857	**
M7	-184829.5771	0.22705	4.62358				
M8a	-184813.0237	0.23248	5.02542	1.00000			
M8	-184813.0237	0.23248	5.02544	1.00000	M7 M8a	33.1068 0	**

Table 1b Results of PAML random-sites models on 15 nuclear genes

Model	lnL	Parameters[a]			Null	LRTs	*p*
		ω_0/p	ω_1/q	$\omega_2/\ \omega_p$			
M0	-20381.7819	0.0753					
M1a	-20008.8853	0.0395(88.2%)	1.0000(11.78%)				
M2a	-20008.8853	0.0395(88.2%)	1.0000(11.78%)	8.8498(0)	M1a	0	
M3	-19855.8230	0.0000(17.34%)	0.01442(57.58%)	0.3297(25.08%)	M0	1051.9090	**
M7	-19840.1034	0.1753	1.5727				
M8a	-19840.0986	0.1763	1.5997				
M8	-19839.6389	0.1763	1.5930		M7 M8a	0.9290 0.9195	

Note: lnL, log likelihood; [a]ω values of each site class are shown for models M0-M3 ($\omega_0 - \omega_2$) with the proportion of each site class in parentheses. For M7-M8, the shape parameters, *p* and *q*, which describe the beta distribution are listed. LRTs represents the likelihood ratio tests; 2*(lnL(Model) – lnL(Null)).

Table 2a Results of PAML branch-site models on 13 mitochondrial genes.

Branch	Model A	Null model	LRTs	site class	0	1	2a	2b
b1	-189631.0183	-189631.0183	0	proportion	0.94985	0.05015	0.00000	0.00000
				foreground ω	0.02838	1.00000	1.00000	1.00000
b2	-189619.8114	-189623.914	8.2052**	proportion	0.94260	0.04932	0.00769	0.00040
				foreground ω	0.02825	1.00000	18.30494	18.30494
b3	-189710.2567	-189714.0692	7.6250**	proportion	0.92062	0.04856	0.02927	0.00154
				foreground ω	0.02811	1.00000	5.60205	5.60205
b4	-189591.2640	-189598.5732	14.6183**	proportion	0.92700	0.04823	0.02355	0.00123
				foreground ω	0.02799	1.00000	9.01795	9.01795
b5	-189728.0764	-189729.5338	2.9148	proportion	0.91976	0.04869	0.02997	0.00159
				foreground ω	0.02824	1.00000	292.74200	292.74200
b6	-189719.7435	-189722.2222	4.9573*	proportion	0.93956	0.04997	0.00994	0.00053
				foreground ω	0.02824	1.00000	5.43888	5.43888
b7	-189738.4458	-189738.4458	0	proportion	0.94477	0.05030	0.00468	0.00025
				foreground ω	0.02845	1.00000	1.00000	1.00000
b8	-189738.5873	-189738.5873	0	proportion	0.94946	0.05054	0.00000	0.00000
				foreground ω	0.02848	1.00000	1.00000	1.00000
b9	-189738.5873	-189738.5873	0	proportion	0.94946	0.05054	0.00000	0.00000
				foreground ω	0.02848	1.00000	1.00000	1.00000
b10	-189738.5873	-189738.5873	0	proportion	0.94946	0.05054	0.00000	0.00000
				foreground ω	0.02848	1.00000	1.00000	1.00000
b11	-189613.4925	-189622.0967	17.2084**	proportion	0.94688	0.04932	0.00362	0.00019
				foreground ω	0.02831	1.00000	79.27458	79.27458
b12	-189615.9415	-189619.0732	6.7338**	proportion	0.92922	0.04889	0.02080	0.00109
				foreground ω	0.02820	1.00000	999.00000	999.00000
b13	-189610.5139	-189616.7321	12.4364**	proportion	0.94411	0.04933	0.00624	0.00033
				foreground ω	0.02821	1.00000	8.65776	8.65776
b14	-189734.9950	-189736.0673	2.1446	proportion	0.94724	0.05043	0.00221	0.00012
				foreground ω	0.02842	1.00000	4.83612	4.83612
b16	-189624.9748	-189628.011	6.0725**	proportion	0.94529	0.04967	0.00479	0.00025
				foreground ω	0.02827	1.00000	7.66238	7.66238
b17	-189621.7633	-189627.8443	12.1622**	proportion	0.94863	0.04955	0.00173	0.00009
				foreground ω	0.02835	1.00000	336.12589	336.12589

b18	-189738.5873	-189738.5873	0	proportion	0.94945	0.05055	0.00000	0.00000
				foreground ω	0.02848	1.00000	1.00000	1.00000
b23	-189618.7575	-189625.3146	13.1142*	proportion	0.93161	0.04861	0.01879	0.00098
				foreground ω	0.02817	1.00000	999.00000	999.00000
b24	-189578.5777	-189597.2838	37.4121**	proportion	0.90294	0.04769	0.04689	0.00248
				foreground ω	0.02828	1.00000	999.00000	999.00000
b25	-189591.1234	-189603.6533	25.0598**	proportion	0.93264	0.04910	0.01735	0.00091
				foreground ω	0.02815	1.00000	9.15193	9.15193

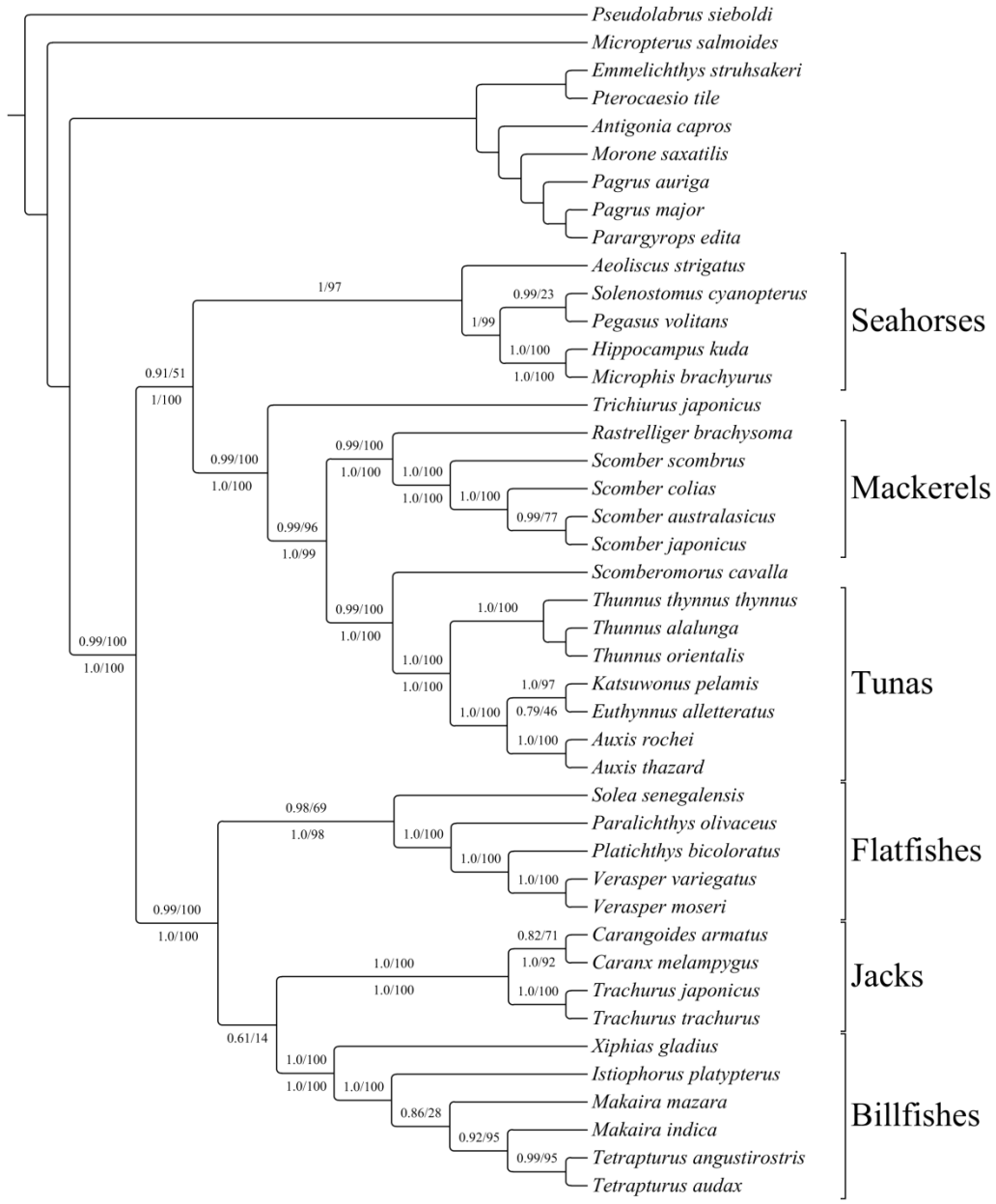
Table 2b Results of PAML branch-site models on 15 nuclear genes.

Branch	Model A	Null model	LRTs	site class	0	1	2a	2b
b1	-19976.5595	-19985.03038	16.9416**	proportion	0.86199	0.11098	0.02395	0.00308
				foreground ω	0.03654	1.00000	8.00791	8.00791
b2	-20008.8853	-20008.88562	0.0006	proportion	0.88208	0.11784	0.00007	0.0001
				foreground ω	0.03954	1.00000	1.00000	1.00000
b3	-20008.8853	-20008.8853	0	proportion	0.88215	0.11785	0.00000	0.00000
				foreground ω	0.03954	1.00000	1.00000	1.00000
b4	-20002.5866	-20008.09543	11.0176**	proportion	0.87846	0.11792	0.00320	0.00043
				foreground ω	0.03887	1.00000	179.95960	179.95960
b5	-20005.7808	-20006.94121	2.3208	proportion	0.88049	0.11738	0.00188	0.00025
				foreground ω	0.03929	1.00000	15.98495	15.98495
b6	-20008.0869	-20015.6358	15.0977**	proportion	0.87280	0.11539	0.01043	0.00138
				foreground ω	0.03923	1.00000	3.58708	3.58708
b7	-20005.0092	-20005.00916	0	proportion	0.74771	0.10039	0.13392	0.01798
				foreground ω	0.03825	1.00000	1.00000	1.00000
b8	-20005.5201	-20005.52013	0	proportion	0.56864	0.07567	0.31392	0.04177
				foreground ω	0.03851	1.00000	1.00000	1.00000
b9	-20007.3117	-20007.96269	1.3020	proportion	0.88060	0.11479	0.00408	0.00053
				foreground ω	0.03943	1.00000	4.44637	4.44637

Note: Estimated likelihood values under model A (allowing positive selection) and the null model (no positive selection), likelihood ratio tests (LRTs), and estimated parameters of model A. LRT critical values 3.84 at $p = 0.05$ (*) and 5.99 at $p = 0.01$ (**). Branches are as identified in Figure 2.

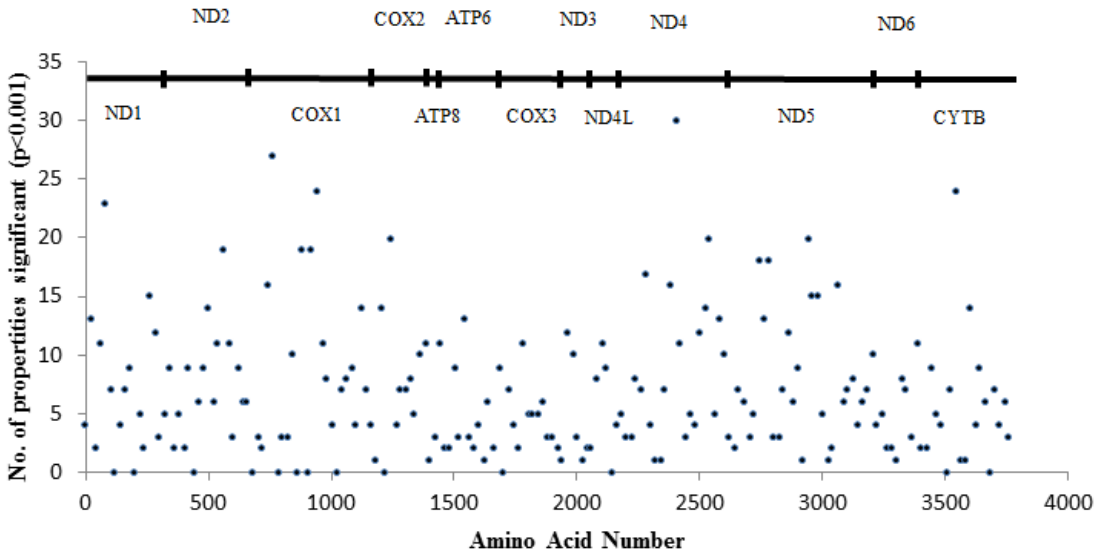
Supplementary Materials

Figure S1 Phylogeny of six fish groups based on Bayesian and maximum likelihood methods.



Note: Maximum liklihood analyses were partitioned by codon position; Bayesian analyses were not partitioned. Values above each branch are from analyses of 13 mitochondrial protein genes. Values below each branch are from analyses of a subset of taxa for which 7 nuclear genes plus 13 mitochondrial genes were available. Bayesian posterior probabilities are listed on the left of the / with bootstrap percentages on the right.

Figure S2 Summary of selection on all mitochondrial genes.



Note: Filled circles represent positively selected sites ($p < 0.001$) by TreeSAAP analysis. The dataset was analyzed for significance for the 31 physico-chemical properties using a sliding window with size of 20 amino acid sites. The y-axis represents the number of properties for which that site was determined to be subject to positive natural selection.

Table S1 Species list with mitochondrial genome accession number from NCBI.

Species	NCBI Accession
Aeoliscus_strigatus	NC_010270
Antigonia_capros	NC_003191
Auxis_rochei	NC_005313
Auxis_thazard	NC_005318
Carangoides_armatus	NC_004405
Caranx_melampygus	NC_004406
Emmelichthys_struhsakeri	NC_004407
Euthynnus_alletteratus	NC_004530
Hippocampus_kuda	NC_010272
Istiophorus_platypterus	NC_012676
Katsuwonus_pelamis	NC_005316
Makaira_indica	NC_012675
Makaira_mazara	NC_012680
Microphis_brachyurus	NC_010273
Micropterus_salmoides	NC_014686
Morone_saxatilis	NC_014353
Pagrus_auriga	NC_005146
Pagrus_major	NC_003196
Paralichthys_olivaceus	NC_002386
Parargyrops_edita	NC_008616
Pegasus_volitans	NC_010271
Platichthys_bicoloratus	NC_003176.1
Pseudolabrus_sieboldi	NC_009067
Pterocaesio_tile	NC_004408

Rastrelliger_brachysoma	NC_013485
Scomber_australasicus	NC_013725
Scomber_colias	NC_013724
Scomber_japonicus	NC_013723
Scomber_scombrus	NC_006398
Scomberomorus_cavalla	NC_008109
Solea_senegalensis	NC_008327
Solenostomus_cyanopterus	NC_010267
Tetrapturus_angustirostris	NC_012679
Tetrapturus_audax	NC_012678
Thunnus_alalunga	NC_005317
Thunnus_orientalis	NC_008455
Thunnus_thynnus_thynnus	NC_004901
Trachurus_japonicus	NC_002813
Trachurus_trachurus	NC_006818
Trichiurus_japonicus	NC_011719
Verasper_moseri	NC_008461
Verasper_variegatus	NC_007939
Xiphias_gladius	NC_012677

Table S2 PCR primers used for nuclear markers used in this study.

Locus	Primer names	Primers (5'~3')
rag1	rag1f1	GAGCTTCTCCCHGGHTTTCA
rag1	rag1r1	AAGTGRAAGCGGAAGGAGCG
rag1	rag1f2	ATGAAAGAGAGCAGGCTYATC
rag1	rag1r2	AGGGCTGCCCTCCAGCTGCG
rag1	rag1f3	TCTSAAAACATGGTGCTDCA
rag1	rag1r3	CCATCTYTCTCKATGATTTC
rag2	rag2f1	TTCCAGAGAGYTAYCTCATC
rag2	rag2r1	AGCAARGGRCTGCCCTGCAG
rag2	rag2f2	GGACAGTCYTTCCATBTGGC
rag2	rag2r2	GGCARCATTTGATCCARTAGCC
tmo-4c4	tmo4c4f1	AAGAARAGAGTGTTTGAAAATG
tmo-4c4	tmo4c4r1	ACAGCWCCCTCCTCRTAAAT
tmo-4c4	tmo4c4r2	ATRATCATRCTCTTRTTGTC
zic1	zic1f1	ATGCTCTTGGACGCAGGACCGCA
zic1	zic1r1	CCACAGCGGGGAACGGACA
zic1	zic1f2	CATCACTCAACAGGCGAAG
zic1	zic1r2	TTGGCTTTGAACGGYTTYCCTTC
myh6	myh6f1	ATGCYTAYCARTACATGCTGAC
myh6	myh6r1	GGTTRATYCTVACCACCATCCA
myh6	myh6f2	GAGAACCARTCBGTSCTCATCAC
myh6	myh6r2	CCAGTTGAACATYTTYTCRTA
btbd7	btbd7f145	CCAGTCGCTCAGCTGATCATGC
btbd7	btbd7r1093	ATGTGGTANAGCTCCATNGCCTC

Table S3 GenBank accession numbers of nuclear genes (rag1, rag2, tmo4c4, zic1, myh6, and btbd7) sequenced in this study.

	rag1	*rag2*	*tmo4c4*	*zic1*	*myh6*	*btbd7*
Abudefduf_vaigiensis			KP866724	KP866740		
Acanthocybium_solandri	KP866741					
Aphanopus_carbo	KP866742	KP866761	KP866710			
Auxis rochei	KP866743	KP866762	KP866703		KP137552	
Brama brama		KP866763	KP866725	KP866734	KP137555	
Carangoides_ferdau	KP866744	KP866764				
Caranx_ruber		KP866765	KP866704			
Centrolophus_sp.		KP866766	KP866719	KP866739	KP137558	KP121466
Centropomus undecimalis				KP866730	KP137554	
Centropomus_ensiferus			KP866709	KP866731		
Elassoma_evergladei		KP866767				
Etheostoma_vitreum		KP866768	KP866722			
Euthynnus_affinis	KP866745					KP121462
Gempylus_serpens	KP866746	KP866769	KP866712			
Gomphosus_varius			KP866721			
Gymnosarda_unicolor		KP866770	KP866717			
Icichthys_lockingtoni	KP866747	KP866771	KP866718	KP866738	KP137557	KP121465
Katsuwonus_pelamis	KP866748					
Lepidopus_altifrons	KP866749	KP866772	KP866713			
Lycodes terraenovae		KP866773	KP866708		KP137553	
Micropterus_salmoides		KP866774		KP866737		KP121464
Neoepinnula_americana	KP866750		KP866711			
Paralichthys_californicus	KP866751			KP866729		
Peprilus simillimus	KP866752	KP866775	KP866716		KP137556	
Perca_flavescens		KP866776	KP866723			
Platax_orbicularis			KP866720		KP137559	
Pterycombus_brama		KP866777	KP866715	KP866735		
Ruvettus_pretiosus	KP866753			KP866732		

Schedophilus_medusophagus	KP866754	KP866778				
Scomber combrus				KP866726		KP121459
Scomber japonicus	KP866755	KP866779	KP866702	KP866727	KP137551	KP121460
Scomberomorus maculatus				KP866728		KP121461
Sphyraena_argentea			KP866707			
Sphyraena_barracuda	KP866756					KP121463
Sphyraena_putnamae	KP866757		KP866706			
Taractichthys_longipinnis		KP866780	KP866714	KP866733		
Tetrapturus_albidus	KP866758	KP866781	KP866705			
Thunnus_albacares	KP866759					
Trachurus_lathami		KP866782				
Trichiurus_lepturus				KP866736		
Xiphias_gladius	KP866760					

Table S4 Number of amino acid sites showing positive selection signals in each mitochondrial gene on relevant branches.

Branch	ND1	ND2	CO1	CO2	ATP8	ATP6	CO3	ND3	ND4L	ND4	ND5	ND6	CYTB
b2	0	1	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	1
b3	0	1	0	0	0	2	1	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0
b4	0	0	1	2	0	0	1	0	1	0	2	2	0
	0	0	0	0	0	0	0	0	0	0	0	1	0
b6	1	0	0	0	0	1	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0
b11	0	0	0	0	0	1	0	0	0	1	0	1	0
	0	0	0	0	0	0	0	1	1	0	2	0	0
b12	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	0	0
b13	0	0	0	0	0	0	0	0	0	0	1	0	0
	0	2	0	0	0	1	1	1	0	1	1	0	3
b16	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	0	0	1	1	0	0
b17	0	0	0	0	0	0	0	0	0	1	1	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0
b23	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0
b24	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0
b25	1	2	0	0	0	0	0	0	0	2	1	0	0
	0	1	0	0	1	0	0	0	1	2	1	2	0
Total in PAML (>0.95)	2	4	1	2	0	5	2	0	1	4	6	3	1
Total in PAML (<0.95 >0.8)	1	4	0	0	1	2	1	2	2	4	5	4	4
Total in TreeSAAP	183	77	25	32	26	44	24	20	11	82	152	46	42

Note: The top number for each branch represents the number of amino acids having Bayes empirical Bayes (BEB) posterior probability higher than 0.95 identified in PAML; the bottom number represents the number of amino acids having BEB posterior probability higher than 0.80 and lower than 0.95 identified in PAML. The last row lists the total number of amino acid sites identified as positively selected by TreeSAAP. See Supplementary Table S4 for additional details.

Table S5 Positively selected amino acid sites for each mitochondrial gene on various branches.

	ND1	ND2	COI	COII	ATP8	ATP6	COIII	ND3	ND4L	ND4	ND5	ND6	CYTB
b2	10 I	419 L [9] 475 Q 602 L* [2]	1074 T	1343 V		1596 N* [6]		2007 S		2306 I 2363 S 2404 V [3] 2423 I 2472 A 2525 I	3172 I [3]		3627 A*
b3	3 S 29 V [2] 159 I [3] 177 T 274 L	371 N 595 T 652 I 654 S*	847 P 1005 I 1185 Q		1441 T [2] 1466 S [6]	1520 G [7] 1524 M* [7] 1527 T 1529 Q 1531 L 1534 L 1536 L 1596 N* [6] 1652 F [10]	1854 E* 1857 Q		2111 G 2122 L	2224 C [2] 2556 I [2]	2643 S [2] 3078 P 3176 L	3330 W 3346 A 3347 L [1] 3364 E [1] 3365 V 3415 L	3475 S 3477 I [2] 3478 A 3479 T 3482 S 3500 M 3576 N [9] 3580 Q 3613 L [2] 3632 D 3652 G* [1] 3655 T 3680 L 3724 I 3739 T [1] 3761 V 3763 P 3784 T [5]
b4	181 I	435 L [2] 517 V [3] 559 T [1] 563 A [1] 602 L [2] 641 L [4] 658 I	849 I 851Q** [4] 854 T 1187 Q	1205 V 1236 S 1237 T 1241 N 1243L** [3] 1333 V 1361 M 1372 T 1375 T 1376L** [8] 1416 L [4]		1550 I	1706 A 1750 L [3] 1760 I 1791 F 1870 F* [9] 1924 Y		2100 T 2205 F	2278 A [7] 2530 A 2611 I	2656 I 2697 S [5] 3214 I [3] 3220 A [4] 3239 V** [14]	3248 Y 3251 I 3253 L 3256 S** 3390 V	3460 I [6] 3482 S 3627 A [2]
b6	7 T* [7] 259 E [9]			1238 K 1244 L		1483S*	1746 I [6] 1922 I				2635 H* [8]	3249 V [1]	
b11						1483S*		2050 F** [2]	2080 V* [9]	2192 F* [3]	2710 L* [3] 2722* I [11]	3336 M* [3]	
b12			891 T 1133 I [1]	1239 L	1433 L [1]	1503 F [2]	1740 I [1]		2100 T	2612 A [6]			3421 S [8]
b13	9 I	332* I [1] 601* M				1483* S 1543 L [3]	1740 I [1] 1746 I [6] 1750 L [3] 1754 Y [2]	2054 S [3]		2228 K*[11]	2656 I 2661 S [2] 3070 S** [2] 3152 F* [9] 3240 L [1]		3543 A [9] 3611 T** [3] 3714 S* [10] 3794 M [4]
b16		411 D* [2]			1451 K					2568 L*	2690 F [8] 2846 M*		3627 A [2] 3714 S [10]
b17	177 T*	411 D [2] 470 P [2]	935 A [4]	1289 L				1980 F		2564 F*	2915 L* [7]		
b23		475 Q 633 I [6]			1467 K [12]	1544 M				2473 A [2]	2657 I [1] 2910 P [5] 3208 T [0]		
b24		398 S				1500 W		2042 L	2124 T 2167 Q	2318 N	2702 T [7] 2717 S [6] 3214 I [3]		3464 L 3609 A
b25	158 A* [7]	417 T [2] 585 G* [6] 592 S* [10] 600 N* [2] 656 M [11] 670 T [1]			1440 A* [7] 1459 V [2]	1663 L			2168 A* [6]	2217 W** 2305 L 2364 T* 2572 L* [2] 2633 F* [10]	2654 T* [6] 2823 S [16] 3148 S* [2] 3153 T [1] 3161 L [13]	3291 F* [2] 3330 W 3347 L [1] 3370 L [9] 3411 E* [19]	3543 A [9]

Note: Each site lists the ordinal number of this site in the aligned sequence matrix, and the amino acid (based on the top sequence in the alignment). Symbols indicate bayes empirical bayes (BEB) probabilities as follows: ** >0.99, * 0.95 - 0.98, # 0.80 - 0.94, no symbol 0.5 - 0.79. If the site is also identified as having at least one property significant in TreeSAAP, the number of such properties is shown in square brackets. Sites of each gene in the same color indicate they occur on multiple branches.

Chapter 3: Effects of environmental factors and life history on evolutionary rates of mitochondrial protein-coding genes in bony fishes

Abstract

Understanding the factors that affect the rates of nucleotide substitution is central to evolutionary biology, population genetics, and mutation research and has been a long-standing biological question. Mitochondrial genes have been popular molecular markers in population genetic studies. However, how the evolution of mitochondrial genes is affected by environmental factors (*e.g.*, temperature and salinity) and life history traits (*e.g.*, migration) has not been examined in vertebrate species at a broad phylogenetic scale. Fishes inhabit environments that include a great variation in temperature (*e.g.*, tropical, subtropical, temperate, and deep cold water), salinity (*e.g.*, fresh water, brackish, and salt water), and also exhibit variation in migration behavior (anadromous, catadromous, amphidromous, oceanodromous, potamodromous, and non-migratory). The newly published fish tree of life provides a broad evolutionary background for such analysis. We investigated how the rates of evolution of mitochondrial protein-coding genes are affected by temperature, salinity, and migratory behavior in 972 bony fish species. Our results showed that tropical fishes have the highest synonymous substitution rate (dS) and nonsynonymous substitution rate (dN), and deep cold water fishes also have dS and dN higher than subtropical and temperate fishes. One possible explanation for these findings is that substitution rates are not only affected by the absolute temperature (low or high), but might also be affected by the stability of the temperature (relatively constant or variable). Similar patterns were found in fishes from environments that differed in salinity levels: fishes that can live in both fresh and salt water (variable salinity environment) had lower dN and dS than fishes

that can only live in either fresh or salt water (stable salinity environment). Migratory fishes had lower substitution rates than non-migratory fishes. This latter pattern is probably because migratory fishes with high energy demands usually have high oxidative phosphorylation (OXPHOS) efficiency, thus their OXPHOS systems have lower tolerance for substitutions.

Introduction

Understanding the factors that affect nucleotide substitution rates is central to evolutionary biology, population genetics, and mutation research (Martin and Palumbi 1993), and has been a long-standing biological question. Teasing apart the effect of the major factors that directly or indirectly cause the variation of substitution rates among species becomes more complicated when the factors are not completely independent (*e.g.*, body size is correlated with factors such as metabolic rate and generation time). Variation in substitution rates has been explained by one or a combination of several factors such as body size (Gillooly et al. 2005; Martin and Palumbi 1993; Popadin et al. 2007), metabolic rate (Martin and Palumbi 1993; Palumbi 1992), generation time (Laird et al. 1969), environmental factors (e.g. temperature, salinity, and ultraviolet exposure) (Allen et al. 2006; Gillooly et al. 2005; Wright et al. 2003), population size (Nikolaev et al. 2007; Popadin et al. 2007), as well as natural selection directly on substitution rates (Barrett et al. 2010). Among these factors, environmental factors can act as "selective" sieves, which influence levels of genotypic diversity generated by random mutations (Gillespie 1991); they can also modulate the rate of introduction of variation or even the type of new mutants (Cairns et al. 1988). Therefore, investigation the effects of environmental factors on nucleotide substitutions can enhance our understanding of the

genetic basis of organismal adaptation to specific environmental habitats and yield insights into the mechanisms that drive evolutionary changes.

Temperature is an environmental factor that affects nearly all biological processes, including individual metabolic rates, growth and reproduction rates, and it also limits organisms' distributions. Rohde (1992) proposed that ectothermic organisms that inhabit warmer environments have faster evolutionary rates compared to their relatives that inhabit colder environments as a result of shorter generation time, higher mutation rates, and stronger selection. Much evidence supports this prediction (Allen et al. 2006; Gillman et al. 2009). Although fishes are living in water with a relatively more stable temperature environment compared to terrestrial organisms, there is great variation in water temperature: ranging from around -1.9°C in polar areas to about 34°C in the tropics and deserts. In addition, some temperate environments fluctuate between such extremes seasonally. A large number of fish species (more than 1,000) have mitochondrial genome DNA sequences available, which provides an opportunity to investigate how water temperature influences organismal nucleotide substitution rates.

There has to date been no direct evidence that various salinities can affect substitution rates. However, many adaptive responses to salinity change such as turnover of gill cells, gill cell differentiation, expression of ATPases and active ion transporters (such as osmotic stress transcription factor 1 [OSTF1] and homologs of transcription factor II B [TFIIB] have been observed (Evans 2002; Fiol and Kültz 2005). As such, some genetic changes to alterations of saline environments may be adaptive. At the biochemical level, it is known that salt affects protein and DNA structure, as well as their interactions (Pegram et al. 2010). Therefore, salt could

potentially affect substitution rates of DNA sequences. In addition, fish osmoregulation is correlated with an individual's energy consumption or metabolic rate, which is also one of the most important factors that can affect nucleotide substitution rates. Fishes are among a few vertebrate groups that live in water with a great variety of salinity levels (*e.g.*, fresh water, brackish, and marine), which provides a great opportunity to investigate how different water salinity levels affect nucleotide substitution rates.

Another factor that is related to the differences in metabolic rate in fishes is locomotion. Fishes exhibit great variation in their locomotion ability. Studies have shown that migration is energetically costly in fishes (Gross 1987; Koch and Wieser 1983). Therefore, it is predicted that migratory fishes like tunas, salmons, billfishes, and mackerels have much higher energy demands (and higher metabolic rates) than very sessile fishes like seahorses and flounders. Higher metabolic rates are associated with increased production of reactive oxygen species, the by-products of aerobic respiration, which could damage DNA and lead to higher nucleotide substitution rates (Demple and Harrison 1994; Friedberg et al. 1995; Martin and Palumbi 1993; Yakes and Van Houten 1997). In addition, high turnover and frequent repair may increase replication errors and contribute to higher rates of evolution (Martin and Palumbi 1993). However, Sun et al. (2011) showed that migratory fishes have lower dN (and dN/dS ratio) than non-migratory fishes. We therefore, investigated whether there is correlation between fish migratory ability and nucleotide substitution rates and how they might be related.

The oxidative phosphorylation (OXPHOS) system plays a critical role in organisms' aerobic respiration because it is the major metabolic system of energy production (ATP) under aerobic condition in almost all eukaryotes. It is also unique in

that it is composed of subunits encoded from both mitochondrial genes (13) and nuclear genes (70+). Mitochondrial OXPHOS genes differ from nuclear genes in that they are localized in mitochondria, where reactive oxygen species are most abundant and their mutagenic effect is the strongest. Therefore, mitochondrial OXPHOS genes are more sensitive to factors that are associated with organismal metabolic rates, such as temperature, salinity, and locomotion ability. The evolution of OXPHOS genes not only records the accumulation of random-neutral changes, but also reflects the consequence of natural selection during the evolutionary process of specific physiological and/or behavioral traits.

This study was conducted on a well-resolved phylogeny of 972 fish species to investigate the effects of two important environmental factors (temperature and salinity) and one life-history trait (locomotion) on the substitution rates of all 13 mitochondrial protein-coding genes, all of which encode proteins in OXPHOS pathway. We investigated potential associations between these environmental and life-history traits and nucleotide substitution rates, and found that several factors exhibit positive correlations.

Methods

We collected temperature, salinity, and migratory information of 972 fish species from Fishbase (www.fishbase.org, Froese and Pauly 2000). Fishes were classified into four categories based on the average annual water temperature: “tropical” (359 species), “subtropical” (212 species), “temperate” (221 species), and “bathydemersal” (or cold-deep water) (180 species). Tropical and subtropical areas have relatively stable and high temperatures all year around, with water temperature

above 20°C and 10-20°C, respectively. Temperate areas have a wider temperature range (native fishes can tolerate minimum water temperatures below 10°C) with warm summers and cold winters. Bathydemersal areas (or cold-deep water) are 200-2,000 meters depth with a stable and low temperature (0-3°C).

We used two schemes to categorize fishes based on salinity. The simple scheme classifies fishes into three categories: freshwater (fishes that can only live in fresh water) (288 species), marine (fishes that can only live in salt and/or brackish water) (489 species), and fresh-marine (fishes that can live both in fresh water and salt water– usually as a result of migrating between waters of differing salinity) (195 species). The more complicated scheme classifies fishes into five categories based on utilization of more than one salinity environment: fully fresh water (288 species), fresh-brackish (74 species), fresh-marine (121 species), brackish-marine (93 species), and fully marine (489 species) (In a technical sense, "fresh water" represents water with less than 0.5 g/L of total dissolved mineral salt; "brackish" represents water that contains > 0.5 and < 30 g/L of total dissolved salts; and "marine" represents water pertaining to the sea (i.e., ~30 grams per liter or greater), from the open oceans to the high tide mark and into estuaries).

Based on their migratory ability, fishes were first categorized into "non-migratory" (681 species) and "migratory" (291 species) groups. The migratory group was further categorized into "diadromous migratory" (118 species) and "non-diadromous migratory" (187 species). Diadromous fishes are those that migrate predictably between fresh and salt water at relatively fixed times in their lives. They are a little less than 1% of all fish species, and many of them are very important

commercially (Helfman et al. 2009). Diadromous migratory fishes contain three different forms: “anadromous” (54 species), “catadromous” (33 species), and “amphidromous” (31 species); and “non-diadromous migratory” fishes contain two forms: “oceanodromous” (107 species) and “potamodromous” (66 species) groups. Anadromous fishes are diadromous fishes that spend most of their lives in the sea and migrate to fresh water to breed. Catadromous fishes are diadromous fishes that spend most of their lives in fresh water and migrate to the sea to breed. Amphidromous fishes are diadromous fishes whose migration from fresh water to salt water, or vice versa, is not for the purpose of breeding, but occurs regularly in the juvenile stage of the life cycle. Oceanodromous fishes are truly migratory fishes that live and migrate only in the sea. Potamodromous fishes are truly migratory fishes whose migrations occur entirely within fresh water.

We collected DNA sequences of 13 mitochondrial protein-coding genes of all species from National Center for Biotechnology Information (NCBI). Each gene was aligned using Geneious 7.1 and then sequences of all 13 genes were concatenated together. The concatenated alignments were used to estimate dN and dS of each species. We used the branch model in the codeml program implemented in PAML (Yang 2007) to estimate these rates. The branch model estimates substitution rates of each branch (including the external branches and all internal branches) on the phylogenetic tree. The phylogeny from Betancur-R et al. (2013) was used as the “backbone” phylogeny at the family level and concatenated sequences of 13 mitochondrial protein-coding genes were used to determine the phylogenetic relationships to the species level using RaxML (Stamatakis 2006). We used Model 1 in codeml, which allows the overall substitution

rates and the ratio of dN/dS to vary among all branches. Because of the large number of taxa, the phylogeny with 972 species was separated into four smaller trees with roughly equal taxon numbers, with about 10 species in common between each pair of adjacent subtrees. Only dN, dS, and dN/dS values associated with the external branches were used in the subsequent analyses.

When examining the correlation between environmental temperature, salinity or migratory ability and substitution rates across a range of species, one problem that prevents the use of general statistical methods is that related species may share the same traits due to shared ancestry. Because many related species may share a trait that evolved once on the phylogeny, the species are not independent samples of the trait, rather they are replicates of a single event. This phylogenetic non-independence violates basic assumptions of most standard statistical procedures. We used two common methods to correct for phylogenetic non-independency: phylogenetically independent contrasts (PIC) (Felsenstein 1985, 1988), and phylogenetic generalized least squares methods (PGLS, Martins and Hansen 1997). Both methods were performed in the R computing environment using the 'ape' (Paradis 2011), 'geiger' (Harmon, et al. 2009), and 'nlme' (Pinheiro, et al. 2007) packages. The PIC method transforms the tip data into values of internal nodes that are statistically independent and then internal node values are compared. The GLS method can be seen as an extension of the method of ordinary least squares, which estimates the unknown parameters in a linear regression model, with the goal of minimizing the sum of the vertical distances between each data point and the corresponding point on the regression line (Paradis 2011). The assumptions that observations have the same variance and covariance is equal to zero

are relaxed with GLS. Because the taxon number is relatively large, as long as one method shows that substitution rates are correlated with other traits, it is considered, as that there is correlation after correction for phylogeny.

If phylogenetically independent contrasts showed significance, post-hoc statistical methods were applied. Kruskal-Wallis one-way analysis of variance was conducted on ranks followed by all pairwise multiple comparisons (Dunn's method), which were performed when the test for normality (Shapiro-Wilk) failed.

Results

Temperature

PIC, but not PGLS, shows that dN is correlated with water temperature after correction for phylogeny. The results of post-hoc tests show that tropical and cold-water fishes have significantly higher dN than subtropical and temperate fishes (Figure 1a). Both PIC and PGLS show that dS is correlated with temperature after correction for phylogeny: tropical fishes have significantly higher dS than subtropical, temperate and cold water fishes. The cold water fishes do not have the lowest dS, instead, they have significantly higher dS than temperate fishes (Figure 1b).

PGLS, but not PIC,reveals that dN/dS is correlated with water temperature after correction for phylogeny: cold water fishes have significantly higher dN/dS than all other groups (Figure 1c).

Salinity

Salinity - 3 categories

We first compared fishes living in freshwater, marine, and both freshwater and marine (including brackish water).

Both PIC and PGLS show that dN and dS are correlated with salinity, after correction for phylogeny. Fishes that can live in both fresh water and salt water have significantly lower dN than fishes living in either fresh water or salt water (Figure 2a), and significantly lower dS than fishes living salt water (Figure 2b).

PIC shows dN/dS is not correlated with salinity after correction for phylogeny. Even though PGLS shows dN/dS is correlated with salinity after correction for phylogeny, there are no significant differences among any groups (Figure 2c).

Salinity - 5 categories

Our second comparison is among fishes divided into five groups: fresh, fresh-brackish, fresh_ marine, marine-brackish, and marine. Both PIC and PGLS show that dN and dS are correlated with salinity after correction for phylogeny. Fresh-marine fishes have significantly lower dN and dS than fresh, marine-brackish, or marine fishes (Figure 3a, 3b). PIC shows dN/dS is not correlated with salinity after correction for phylogeny. Even though PGLS shows dN/dS is correlated with salinity after correction for phylogeny, there are no significant differences among any groups (Figure 3c).

Migration

Migration – 2 categories

We first compared non-migratory and migratory fishes. Both PIC and PGLS show that dN is correlated with migratory status after correction for phylogeny. Non-migratory fishes have significantly higher dN than migratory fishes (Figure 4a). PGLS, but not PIC, shows that dS is correlated with migratory status after correction for phylogeny. Non-migratory fishes have significantly higher dS than migratory fishes (Figure 4b). PGLS, but not PIC, shows that dN/dS is correlated with migratory status

after correction for phylogeny. Non-migratory fishes have significantly higher dN/dS than migratory fishes (Figure 4c).

Migration - 2 categories within migratory

Our second comparison is conducted within migratory fishes: between non-diadromous migratory fishes that migrate within water of single consistent salinity-level (either fresh or salt water) and diadromous migratory fishes that migrate among waters with different salinity levels (including fresh, brackish, and marine). There are no significant differences in dN, dS, or dN/dS between these two categories (Figure 5).

Migration – 6 categories

Our third comparison is among non-migratory and differentially-migratory fishes (anadromous, catadromous, amphidromous, ocenodromous, and potamodromous).

PGLS, but not PIC, shows that dN and dS are correlated with migratory status after correction for phylogeny. One consistent pattern for both dN and dS is that amphidromous fishes have significantly higher dN and dS than other migratory and non-migratory fishes (Figure 6a, 6b). There is considerable variation among certain groups (such as non-migratory group has significantly higher dS than that of the anadromous group), but it seems there is no simple unifying pattern. There are no significant differences in dN/dS among any groups.

Discussion

The environment affects the evolution of organisms in many different ways and on many different, often interacting characteristics. Fitness is clearly complex and a more complete view of evolution requires understanding organismal modifications of

physiology, metabolism, genotypic variation, and many other factors (Thaler 1994). Thus, understanding environmental factors that affect nucleotide substitution rates will help us decipher the mechanisms that generate and maintain genetic variation and will shed light on adaptive responses to different habitats.

Mitochondrial OXPHOS genes differ from the majority of nuclear genes in that they are localized in mitochondria (thus, outside of the nucleus), and are directly exposed to mutagenetic agents (such as reactive oxygen species, generated through oxidative phosphorylation), which will in turn affect the substitution rates. The efficiency of OXPHOS can be inferred from the metabolic rates, which is affected by the animals' body temperature and locomotion abilities. Fishes are among the few groups of vertebrates that live in water with various salinity levels. Salinity has been shown to affect gill molecular composition and expression levels of particular enzymes (Evans 2002; Fiol and Kültz 2005). Whether it also affects rates of nucleotide substitution in mitochondrial protein-coding genes has not been previously investigated.

We found that fishes from tropical and cold waters have higher dN's than species from subtropical and temperate regions. Nonsynonymous substitutions change amino acids and usually result in change in organismal biological processes. Thus, nonsynonymous substitutions are more likely subject to natural selection. We predicted that tropical fishes would have higher nonsynonymous substitution rates due to higher metabolic rates, which indicates potentially higher mutation rates (*e.g.*, greater production of reactive oxygen species as well as high turnover and replication errors). Our finding that tropical fishes have the highest dS is consistent with this prediction. However, if the degree of temperature is the major factor that affects substitution rates,

one would expect to see cold-water fishes having the lowest substitution rates. However, our results did not support this prediction. There are two possible explanations why deep cold-water fishes do not have the lowest dN: relaxed natural selection or adaptive evolution.

Relaxed natural selection hypothesis: one similarity between tropical and deep cold-water environments is that they are both relatively stable temperature environments compared to subtropical and temperate regions. Thus, fishes from relatively stable environments may need less energy to deal with the inconsistency of temperature, compared to fishes from unstable temperature environments. Fishes from stable (temperature) environments might tolerate certain levels of substitutional changes, thus, relaxed selective constraint on genes in OXPHOS may be observed.

Adaptive evolution hypothesis: "deep" water is considered to be that below 200 meters or more (as deep as 8,000 meters) and there are many environmental characteristics that greatly differentiate these bottom waters from surface waters. Helfman et al (2009) summarized five of these physical factors: tremendous pressures, constant low temperature (2-5 °C), immense space, no sunlight and scarce food resources. These factors have been strong selective forces on fishes. In this environment, fishes show convergent adaptations in physiology, morphology, and behavior. For example, most bathypelagic (regions of water around 1,000-4,000 m) fishes have lost their gas bladders, have greatly reduced eyes, and their body musculature and skeletons are reduced, likely as energy saving mechanisms; lanternfishes have light organs to attract mates; some males in benthopelagic taxa (a group of benthic or bottom-associated species that swim just above the bottom, usually

along the upper continental slope at depths of less than 1,000 m) produce sounds to attract females; mesopelagic (regions of water 200-1,000 m) fishes have very large eyes, often measuring 50% of head length, large pupils and lenses and lengthened eyes (Helfman et al. 2009). It has been reported in eels that the mitochondrial respiration measurements were significantly increased by pressure (Simon et al. 1992, Theron et al. 2000); mitochondrial respiration of pressure-acclimated eels appears to be able to respond to a given energy demand, while consuming less oxygen than non-pressure-acclimated eels. Elevated nonsynonymous substitution rates could facilitate adaptations to this extreme deep-water environment. This idea is consistent with the observation that some deep cold-water fishes have the highest dN/dS values.

Synonymous substitution rates have been positively correlated with temperature and one would expect to see tropical fishes having the highest dS and cold water fishes having the lowest dS. Again, our results do not entirely support these predictions: dS of tropical fish was the highest among different temperature groups, but dS of cold-water fishes was not the lowest among them. Thus, dS appears to not be affected by temperature in a linear way. There could be two explanations for this result: 1) some synonymous substitutions are not neutral (*e.g.*, translational efficiency could have a greater impact in very cold conditions) and therefore the previous explanations for the elevated dN of cold-water fishes also apply; 2) there could be no simple monotonic correlation between temperature and dS. Previous studies also found no correlation between substitution rates and temperature. Hebert et al. (2002) examined substitution rates of *Daphnia* species in different temperature environments, and found no evidence

that evolutionary rates are slower in cold-water compared to warm-water congeneric species (Hebert et al. 2002).

Besides ATP production, heat generation (proton leak or mitochondrial uncoupling) is another important role of mitochondria. This occurs when protons, which are pumped to the inter-membrane space by electron transport, re-enter the mitochondrial matrix without contributing to ATP synthesis, due to facilitated diffusion (Voet, et al. 1999). This process is mediated by thermogenin or uncoupling proteins (UCP1, 2, 3) in mammals and birds (Mozo et al. 2005).

Sun et al. (2011) stated that fishes living in cold environments had greater heat production to maintain or elevate their body temperatures, thus mitochondrial protein-coding genes of fishes living in colder zones accumulated fewer mutations and thus have experienced stronger selective constraints. However, the fact that "fishes living in cold environments have greater heat production to maintain or elevate their body temperatures" is true for less than 0.1% of described fishes, *e.g.*, regionally endothermic tunas (family Scombridae) and lamnid sharks (family Lamnidae), which warm their aerobic swimming musculature and/or other regions (Block et al. 1993; Dickson and Graham 2004), billfishes (family Istiophoridae and Xiphiidae, which warm the eye and brain region only) (Block 1986), and whole-body endothermy present in opah (family Lampridae) (Wegner et al. 2015), but it is not common in the majority fishes. Instead, the majority of fishes are ectotherms and internal physiological sources of heat are of relatively small or quite negligible importance in body temperature regulation. We interpret the high dS and dN in tropical fishes as the result of higher metabolic rates rather than selective constraints on fishes living at lower temperatures. Yet, we found

deep-water fishes had the highest dN/dS ratio (Figure 1c), which indicates either relaxed selection or adaptive responses to deep-water.

Our observation that there is no simple unidirectional trend between substitution rates and salinity concentration suggests that the variation in substitution rates may not be only affected by the salinity concentration, but that it also could be affected by the variability of salinity levels, as in the case of environmental temperature. Two sets of comparisons (Figures 2 and 3) among fishes from different salinity environments showed the same pattern: fishes from inconsistent salinity (they can tolerant both fresh water and marine) had lower dN and dS than fishes from consistent salinity environments (either fresh water or marine). These findings suggest that fishes from inconsistent salinity environments need more energy to deal with changes in salinity level (*e.g.*, osmoregulation), and thus experience strong functional constraint on OXPHOS genes, while fishes from consistent salinity environments experience relaxed functional constraint, leading to the elevated values for dN and dS.

Many studies have looked at the physiological consequences of salinity changes on specific taxa, but not in a broad range of organisms. For example, in Senegalese sole (*Solea senegalensis*) changes in salinity causes a rise in oxygen consumption and greater metabolite mobilization (Herrera et al. 2012). Using cortisol treatment on gilthead seabream (*Sparus auratus*) to mimic salinity change, also caused significant changes in the energy metabolism of osmoregulatory as well as non-osmoregulatory organs (Laiz-Carrión et al. 2002). However, few studies have aimed specifically to investigate the effects of various salinity concentrations on nucleotide substitution rates. Hebert et al. (2002) found accelerated molecular evolution in halophilic crustaceans and

suggested that this result did not appear to be linked to selection or other agents that are known to influence the rate of mutations, such as UV exposure, generation time, or metabolic rate. Rather, it could be because ionic strength under different salinities has potent effects on DNA-protein interactions, as well as on the structural properties of both DNA and proteins, and these factors might account for the lowered fidelity of DNA replication.

Generally, migratory fishes have higher aerobic energy demands compared to non-migratory fishes due to locomotion costs. Thus, we predicted that migratory fishes should have higher substitution rates as a result of higher aerobic energy consumption. However, considering the functional constraint of natural selection on the OXPHOS complexes in fishes with different energy demands, fishes with high energy consumption need high OXPHOS efficiency in ATP production, and thus would undergo stronger purifying selection to suppress non-synonymous substitutions. This is consistent with the observation that migratory fishes have lower dN and dS. The lower dN/dS of migratory fishes also supports the contention that migratory fishes undergo stronger selective pressure than non-migratory fishes.

We did not find significant differences in dN, dS, or dN/dS among diadromous migratory fishes (i.e., those that migrate between fresh and salt water) and non-diadromous migratory fishes (i.e., those that migrate within fresh or salt water), probably because the energy used for dealing with the change between different salinities has no significant effects on substitution rates, or the effects could be obscured by other factors (*e.g.*, swimming performance).

It is interesting that amphidromous fishes have significantly higher dN and dS than non-migratory fishes and other migratory fishes. Amphidromy may serve as an intermediate or stepping stone condition in the evolution of anadromy or catadromy (Gross 1987; Helfman, et al. 2009). Anadromy is largely a Northern hemisphere, high-latitude phenomenon, while catadromy is more common at low latitudes and in the Southern hemisphere, and amphidromy has a bimodal distribution at middle latitudes in both hemispheres, with greater representation in the Southern hemisphere (McDowall 1987). Compared with anadromous and catadromous fishes, amphidromous fishes do migrate for the purpose of breeding and migration regularly takes place in the juvenile stage of the life cycle, while anadromous and catadromous fishes migrate mainly for breeding. Compared with oceanodromous and potamodromous fishes, amphidromous fishes migrate across waters with differing salinity while oceanodromous and potamodromous fishes are restricted to less variable salinity environments. It is difficult to compare the amount of ATP consumption among the five fish groups due to complicated factors such as different swimming speed and duration, osmoregulation, reproduction and allocation of energy to different activities. Thus, it is not clear what causes the significant difference that we observed among groups.

Temperature, salinity and migration are not totally independent of each other. For example, vertically-migrating mesopelagic species swim through and function across a temperature range of as much as 20°C and that could cause changes in substitution rates. Laternfish species that migrate vertically have greater amounts of DNA per cell (i.e., larger genome sizes) than do species that are non-migratory (Ebeling et al. 1971). Increased DNA could potentially allow for multiple enzyme systems that

function at the different temperatures encountered by the fishes (Ebeling et al. 1971). In addition, migratory species often move between different salinity environments, which subject them to multiple potential selective factors. The most general conclusion that can be drawn from our results is that organisms in stable environments appear to tolerate more variation in the OXPHOS system, while in more variable environments selection is more stringent in preserving existing OXPHOS function.

Figures

Figure 1 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS among fishes living in tropical, subtropical, temperate, and cold waters.

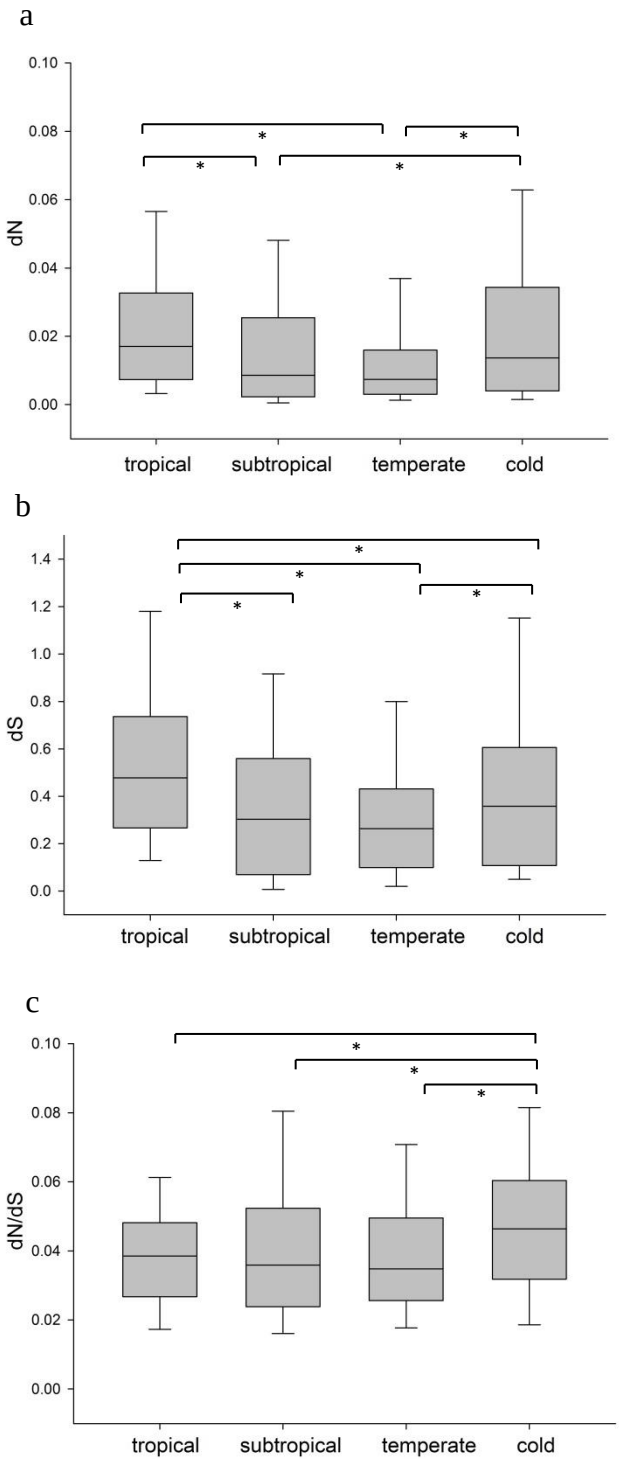
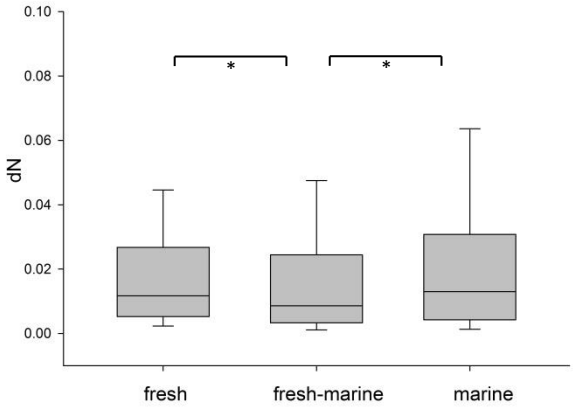
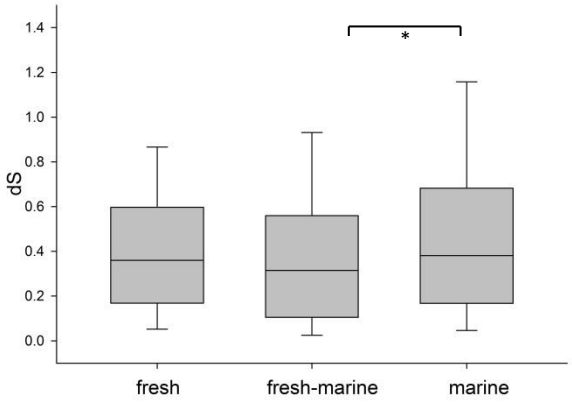


Figure 2 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS among fishes living in fresh water, marine, and both fresh water and marine environments.

a



b



c

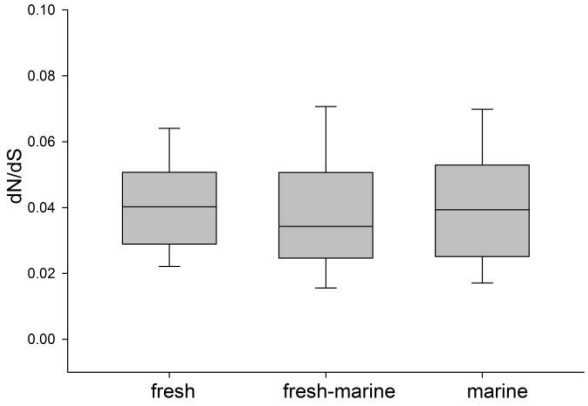
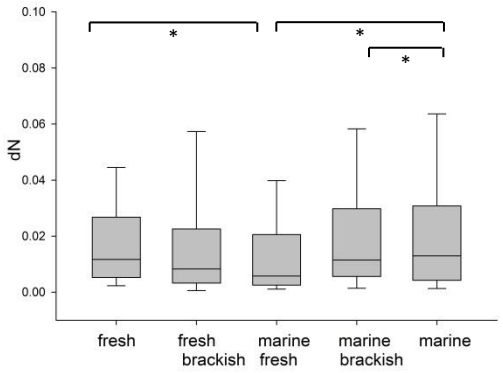
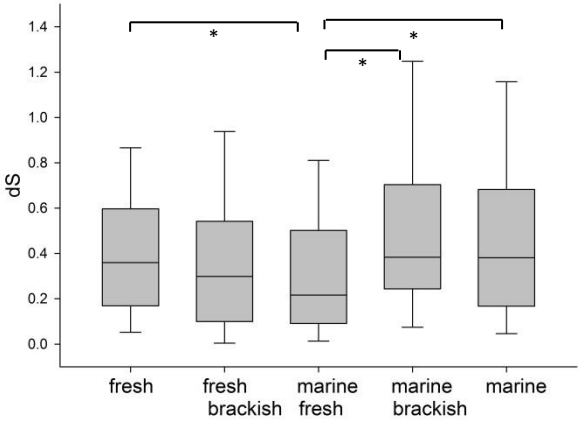


Figure 3 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS among fishes living in fresh water, fresh-brackish, marine-fresh, marine-brackish, and marine environments.

a



b



c

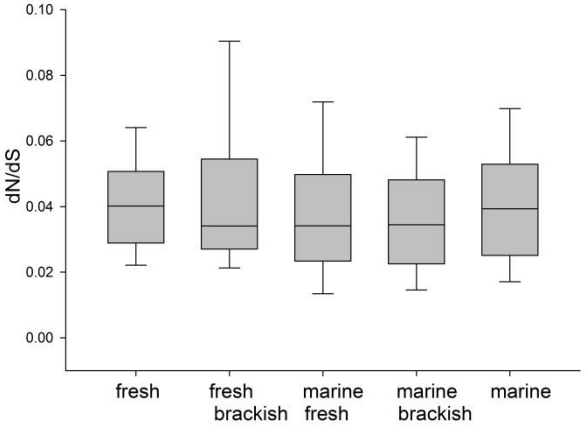
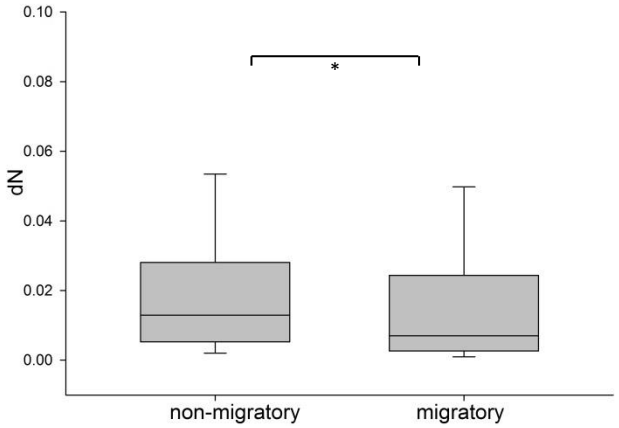
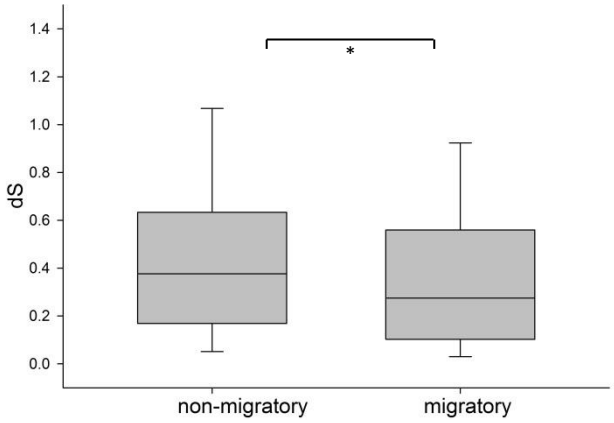


Figure 4 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS between non-migratory and migratory fishes.

a



b



c

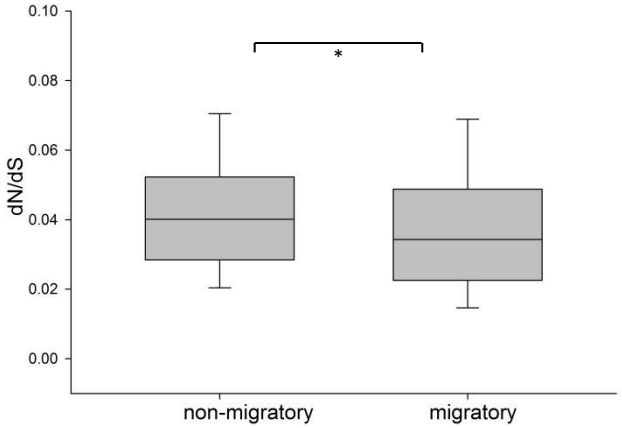
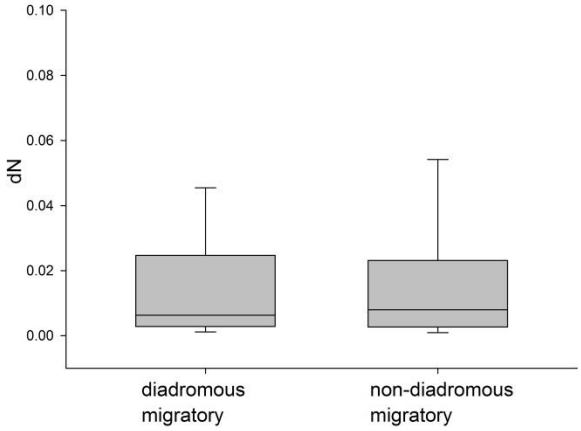
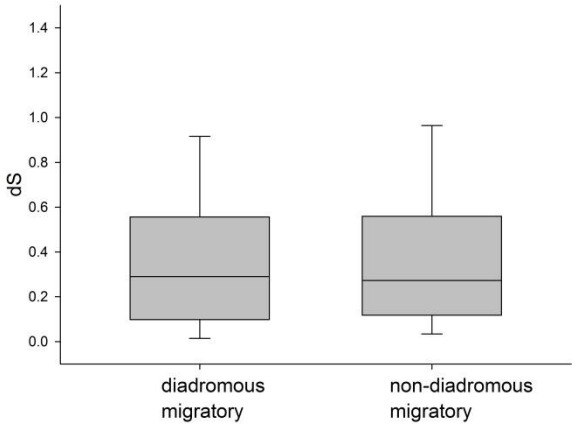


Figure 5 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS (c) between diadromous and non-diadromous migratory fishes.

a



b



c

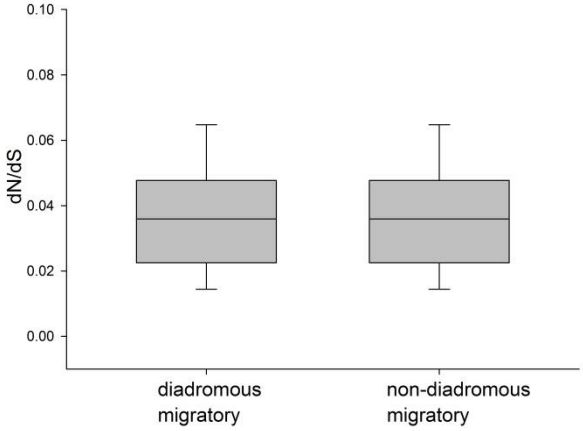
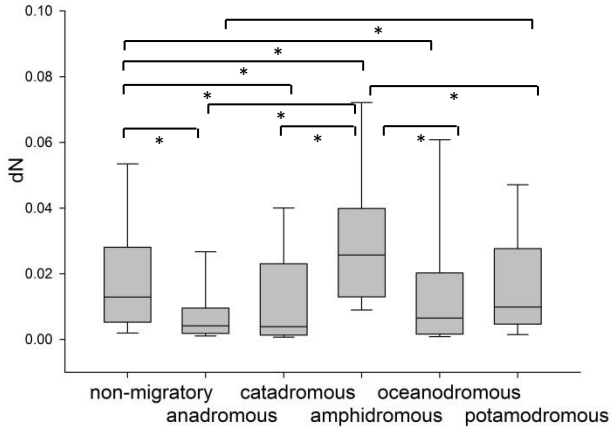
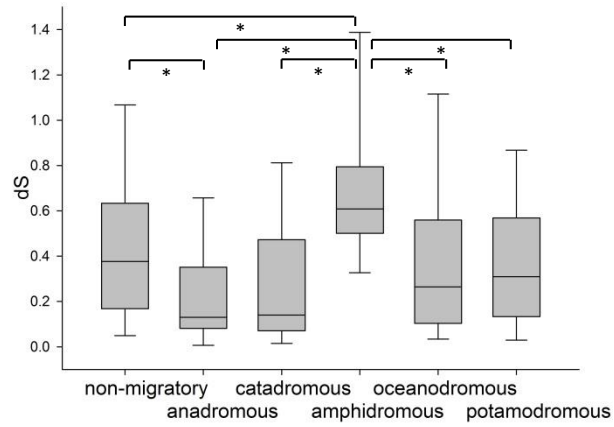


Figure 6 Comparison of (a) non-synonymous substitution rate (dN), (b) synonymous substitution rate (dS), and (c) dN/dS among non-migratory and different types of migratory fishes (anadromous, catadromous, amphidromous, ocenodromous, and potamodromous).

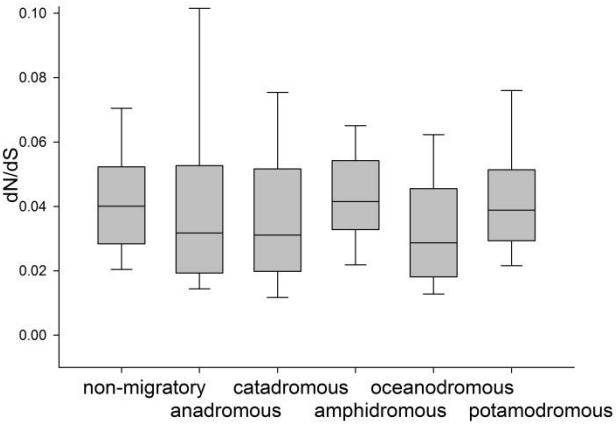
a



b



c



Chapter 4: Various patterns of positive natural selection on bony fishes with different energy demands

Abstract

The recently developed codon-based branch-site model using maximum likelihood methods is a powerful tool for detecting positive natural selection on protein-coding sequences, thus providing an effective means of narrowing down plausible candidate genes or residues for further testing. The oxidative phosphorylation (OXPHOS) pathway is the primary source of ATP in eukaryotic cells and there is much evidence for the association of positive selection on mitochondrial OXPHOS genes on branches leading to organisms with high energy demands or organisms well-adapted to cold environments. However, Zhang and Broughton (2015) found a different pattern: positive selection was found on branches leading to fishes with low energy demands, while negative selection dominated on the branches leading to high-performance fishes. To test if this is an exceptional pattern, we identified 11 groups of bony fishes with some interesting characteristics that make them high energy demanding (and two groups are adapted to cold environments), and examined the evidence for positive selection on branches leading to the target group, its sister group, and the most recent common ancestor of the target group and its sister group. We found that in most cases positive selection was associated with the target group, but we also found various patterns that indicated positive selection on mitochondrial OXPHOS genes, including the pattern found in Zhang and Broughton (2015). It appears that the pattern of positive selection is case-specific and determined by each group's particular evolutionary scenario.

Introduction

Mutations in DNA sequences are subject to several evolutionary forces. Generally, most mutations are either neutral, where fixation or deletion is determined by genetic drift, or selected against, excluded from the gene pool by negative (purifying) selection (Muller 1950; Ohta and Kimura 1973). Only few mutations that increase the fitness of an organism could be fixed by positive natural selection (Gillespie 1991; Nielsen and Yang 1998). Negative selection is useful for detecting regions of high functional importance (Keightley 2012; Ward and Kellis 2012), while positive selection is usually associated with adaption and the evolution of new form or function, and thus, has become a major interest in the study of molecular evolution (Gillespie 1991; Yang 2006). It is remarkably difficult to provide unequivocal evidence for positive selection because it often operates episodically on a few amino acid sites, and the signal may be masked by negative selection and neutral evolution in the same gene (Zhang, et al. 2005). The most convincing cases of adaptive molecular evolution have been identified through comparison of synonymous (silent; dS) and nonsynonymous (amino acid-changing; dN) substitution rates in protein-coding DNA sequences, with $dN/dS > 1$ providing evidence of positive selection in action on the protein molecule (Nei and Kumar 2000). Powerful statistic methods have been developed in the past decades for detecting adaptive molecular evolution and the most powerful method is to use maximum likelihood to detect the presence of a few amino acid sites on pre-specified lineages using codon-based models (Yang and Bielawski 2000; Yang and Nielsen 2002; Zhang, et al. 2005). This provides an effective means of narrowing down plausible

candidate genes or residues for further testing, thus helping to identify substitution hotspots due to adaptive processes.

The oxidative phosphorylation (OXPHOS) pathway is the primary source of ATP in eukaryotic cells. Consequently, adaptation in genes involved in OXPHOS may facilitate evolution of organismal traits that have differing ATP demands. The OXPHOS system is localized in the mitochondrial inner membrane and consists of five complexes (I-V). Each complex (except for Complex II) is composed of protein subunits encoded by both mitochondrial and nuclear genes. All 13 protein-coding genes in mitochondrial genomes encode proteins involved in the OXPHOS pathway. Modifications in the structure of OXPHOS complexes I-V caused by substitutions in constituent proteins could affect "slip reactions", resulting in either more or fewer protons pumped by the electron-transport chain for each electron pair transferred (the $H^+/2e$ ratio), thus affecting the efficiency of ATP production (reviewed in Rand 2005). There has been ample evidence of positive selection acting on mitochondrial OXPHOS genes associated with the evolution of a variety of energetically demanding characteristics (reviewed in Garvin, et al. 2015), including the origin of large brains in anthropoid primates (Doan, et al. 2004), powered flight in bats (Shen, et al. 2010), and adaptation to cold environments in polar bears (Welch et al. 2014).

Nearly all previous studies mentioned above have restricted tests of positive selection on OXPHOS genes to specific branches of interest, usually those leading to taxa that evolved high ATP demands. The results of almost all of these studies support the prediction that the evidence of positive selection was always associated with organisms with high energy demands, which makes intuitive sense: positive selection

favors the substitutions that could increase OXPHOS efficiency in these organisms. However, we previously (Zhang and Broughton 2015) conducted positive selection analyses on high-performance fishes (tunas and billfishes), but also included many other branches on the phylogeny. We did not find evidence of positive selection on the branches directly leading to tunas or billfishes. Instead, we found evidence of positive selection on the branches leading to their sister groups including mackerels and jacks, and sedentary fishes such as seahorses and flatfishes. In this case, the selection strength was inversely proportional to the locomotive ability. This raised the question: is this really an exceptional pattern among all the studies that have detected positive selection on branches leading to organisms with high energy demands? Or, is our finding due to the fact that most previous investigations have simply not examined selection on low-performance taxa (Foote et al. 2010; Garvin et al. 2011)?

Ray-finned fishes (Actinopterygii) include the vast majority of fishes and form the largest class of vertebrates (Nelson 2006). It is an extremely diverse group and one aspect of their diversity is the various levels of energetic consumption. We identified 11 fish groups with characteristics indicating either high energy demands or those live in extreme environments, compared to their sister groups, and examined the patterns of positive selection on mitochondrial OXPHOS genes. The objective was to determine whether positive selection is generally associated with groups with high energy demands. We performed tests for detecting positive selection on the branches leading to the target group, its sister group, and the branch leading to the most recent common ancestor of each clade.

Two groups of fishes appear to have high energy demands because of their ability to generate electric fields. The first group consists of the elephantfishes (mormyrids), belong to the order Osteoglossomorpha (bonytongues), a group that has well-developed teeth on the tongue and are among the most primitive living teleosts, while the second group consists of the black ghost knifefishes (Gymnotiformes). These fishes live in fresh water on all major continents, except Europe (Helfman et al. 2009). There are at least two characteristics that could make this group be very energy demanding. First, mormyrids, along with the related aba (*Gymnarchus niloticus*), have evolved an electrical sense system that can both generate and detect weak electric fields. The electrical sense system enables this group of fishes to be active nocturnally and occur in turbid waters, using their electric fields to sense their environment and communicate with conspecifics (Helfman et al. 2009). Evidence suggests that electric signal generation is very metabolically costly in gymnotiform fishes (Markham et al. 2013), so it is likely to be so in mormyrids as well. Secondly, mormyrids have the largest cerebellum of any fish, and the brain is responsible for approximately 60 % of body O_2 consumption, a figure three times higher than that for any other vertebrate studied (Nilsson 1996). From an evolutionary and ecophysiological perspective, the energetic cost of this brain may be more important than its absolute size (Nilsson 1996). Incidentally, mormyrids also have a strong learning capacity (Burghardt 2005).

Gymnotiformes also have the ability to produce and receive weak electric impulses, but this appears to have evolved independently from the mormyrids (Helfman, et al. 2009). The black ghost knifefishes, which are restricted to Central and South America, live in fresh water and are nocturnally active. The electrogenic tissue is

combined with modified lateral line organs. Electrical output in gymnotiforms is continual and at high frequencies, as compared to the pulsed, low-frequency output of mormyrids (Helfman et al. 2009). Although the electric output is very weak (on the order of fractions of a volt), the energy used to generate it is extremely high (Salazar, et al. 2013). For example, the glass knifefish (*Eigenmannia virescens*) generates electric organ discharges (EODs) at frequencies that can exceed 500 Hz, and simulations predict a metabolic cost of ~1-3 x 10^{10} ATP molecules per electrocyte action potential (Markham et al. 2013).

Another group of fishes expected to have high energy demands are the flyingfishes classified in the order Beloniformes, which contains silvery, marine fishes that are active at, and sometimes above, the surface of the water. The suborder Belonoidei, which contains needlefishes (Belonidae), halfbeak (Hemiramphidae), and flyingfishes (Exocoetidae), has gradually evolved to specialized aerial locomotion, a strategy of either predation or predator-escape, such as jumping and gliding that is rare in fishes. Flyingfishes can take off into the air and travel for 30 seconds and as far as 400 meters in a series of up to 12 flights (Davenport 1994). There is no direct evidence of how much energy is inversted for this aerial locomotion and it is suggested that leaving the water and gliding in the air is energetically efficient, as is performed by penguins and fast mammalian swimmers (Reidman 1990). However, fishes need to achieve a high speed in order to take-off the water. This process could cost lots of energy, probably more energy than its sister group (ricefishes). Flyingfishes are also an important marine food in Japan, consumed both fresh and as ingredient for fish cake. In 2001, the total catch of flying fishes was 8,286 tons (Nagase, et al. 2005).

The Belonoidei's sister groups, the medakas or ricefishes (genus *Oryzias*), are less active in locomotion, and have been used extensively in genetic, embryological, and physiological studies as a fish model organism (Wittbrodt et al. 2002). Compared to another model organism, zebrafish (*Danio rerio*), medaka also has the same advantages of short generation time (2-3 months) and transparent eggs, but medaka serves as a complementary model organism of zebrafish because: 1) it has clearly-defined sex chromosomes, and sex determination has been intensively studied (little is known about sex determination in zebrafish and no genetic markers have been found to be sex-linked (Woods, et al. 2000)); 2) it is easy to work with because it is very hardy and less susceptible to disease; 3) the number of mutations recovered through large-scale mutagenesis screens in zebrafish that correspond to those that have been identified in mouse and human is very limited, probably due to the gene duplications that have occurred in teleosts (Wittbrodt et al. 2002).

Additional groups of high-energy demand fishes are cods and icefishes, two groups of fishes that are well-adapted to the polar regions. Cods, together with some other important commercial fishes worldwide (haddocks, hakes, pollocks, and whitings), belong to the Gadiformes. Most gadid species live at high latitudes of the Northern hemisphere and have successfully adapted to a cold environments (4-18°C). Cold environments present challenges for ectothermic fishes because low temperatures depresses cellular processes, metabolic rates and performance level. Therefore, many compensatory adaptations have occurred in cods to adjust to the cold environment, including: (1) the production of antifreeze compounds (Farrell and Steffensen 2005); (2) they are frequently observed to rest in contact with ice and taking refuge inside holes in

ice (Helfman et al. 2009); (3) and they have reduced kidney glomeruli to help retain antifreeze compounds (Eastman 2013). The adjustment of aerobic scope is identified as a crucial step in thermal adaptation, and can be achieved in increments via adjusting mitochondrial aerobic capacities. For example, adjustments in mitochondrial densities and efficiencies have been observed in cod during adaptation to cold along a latitudinal cline (Fischer 2003). In addition, in cod white muscle, increased cytochrome c oxidase activity is observed, but mitochondrial-encoded COX2 mRNA levels were not limiting for functional activities, which indicates post-transcriptional control or limitations by other transcripts of the COX complex (Lucassen, et al. 2006).

The only group of teleost fish that succeeded in adapting to the challenging environmental conditions of Antarctic Ocean, namely low temperature, the presence of sea ice, habitat reduction, and extreme seasonality of primary production, are the icefish (Notothenioidei) (Bargelloni, et al. 1998). The Antarctic region has been at its present location with its present climate for about 20-25 million years, having separated from Australia during the early Cenozoic (Eastman 2013). Some remarkable adaptations of icefish include the production of antifreeze glycoproteins (Hudson et al. 1979), absence of heat shock response (Hofmann et al. 2000), lack of oxygen-binding proteins (hemoglobin and myoglobin) (Ruud 1954), high mitochondrial densities in heart and skeletal muscles (Archer and Johnston 1991), and increased mitochondrial volume (Urschel and O'Brien 2008). ATP production through mitochondrial aerobic respiration in icefish is challenging in this cold environment because the rate of biochemical reactions is acutely sensitive to temperature (Somero 2002). Some adaptations to cold environment include higher enzyme concentrations (such as cytochrome c oxidase)

and/or increased enzyme efficiency. Transcriptome analysis shows that duplications of genes were significantly enriched for proteins with mitochondrial localization, mitochondrial function and biogenesis (Coppe, et al. 2013).

We also chose to focus on fishes that exhibit various migratory patterns: potamodromous paddlefish (Polypteriformes, Polypteridae), anadromous sturgeon (Acipenseriformes, Acipenseridae), amphidromous ladyfish (Elopiformes), catadromous eels (Anguilliformes, Anguillidae), anadromous salmons (Salmoniformes, Salmonidae), anadromous smelts (Osmeriformes, Osmeridae), amphidromous gobies (Gobiiformes, Gobiidae), and oceanodromous mackerels (Scombriformes, Scombridae). Anadromous fishes are diadromous fishes that spend most of their lives in the sea and migrate to fresh water to breed. Catadromous fishes are diadromous fishes that spend most of their lives in fresh water and migrate to the sea to breed. Amphidromous fishes are diadromous fishes whose migration from fresh water to the seas, or vice versa, is not for the purpose of breeding, but occurs regularly in the juvenile stage of the life cycle. Oceanodromous fishes are truly migratory fishes that live and migrate wholly in the sea. Potamodromous fishes are truly migratory fishes whose migrations occur wholly within freshwater. Fish swimming is energetically costly and thus migratory fishes usually have higher energy consumption from swimming than their non-migratory sister group (Gross 1987). For example, Koch and Wieser (1983) suggested that oxygen consumption is positively correlated with swimming activity at any temperature and that a reduction of swimming activity would lead to a saving of 371 KJ/kg every month. Moreover, diadromous migratory (anadromous, catadromous, and amphidromous) fishes spend extra energy dealing with osmoregulation compare to non-diadromous

migratory fishes if their swimming speed and duration levels are similar. Thus, it is interesting to compare positive selection in these different migratory fish groups.

Anadromous sturgeons are restricted to the northern hemisphere. All species spawn in fresh water. Some species such as American lake sturgeon (*Acipenser fulvescens*) and three river sturgeons (*Scaphirhynchus* spp.) are restricted to freshwater, and some are anadromous species, spending part of their lives at sea and returning to fresh water to spawn (Binkowski and Doroshov 1985; Van Winkle 2002). All the sturgeons in our analysis are Asian anadromous species. Some large sturgeons can feed on fishes, but they are generally slow-swimming feeders on benthic invertebrates (Carroll and Wainwright 2003). Sturgeons can grow to a length of several meters and a weight of hundreds to thousands of kilograms (*e.g.,* white sturgeon (*Acipenser transmontanus*) can attain a length of 3.6 m and a weight of 800 kg, and a beluga (*Huso huso*) can attain a length of 8.6 m and a weight of 1,300 kg; Helfman et al. 2009)). Sturgeons are long-lived vertebrates. Belugas have been aged at 118 years, and white sturgeon at 70-80 years (Casteel 1977; Scott and Crossman 1973). Sexual maturity is attained slowly and after maturation, females may only spawn every 3-5 years (Smith 1985). However, fecundity is relatively high, with ovaries accounting for 25% of body mass, making a large female exceedingly valuable (*e.g.,* a beluga female captured in 1924 from the Tikhaya Sosna River of Russia weighted 1,227 kg and yielded 245 kg of caviar. High-grade caviar can sell for more than $150/oz or $5,000/kg, making the fish potentially worth in excess of $1 million; Helfman et al. 2009)!

Paddlefishes are not benthic swimmers, but instead move through the open waters of large, free-flowing rivers, feeding on zooplankton or other fishes. Adult North

American paddlefish (*Polyodon spathula*) are potamodromous, and are passive filter feeders; typically, they swim through the water both day and night with their mouths open, straining zooplankton and aquatic insect larvae indiscriminately (Rosen and Hales 1981). North American paddlefish may live for 30 years and attain 2.2 m length and 83 kg mass. The Chinese paddlefish (*Psephurus gladius*) are probably anadromous, with adults moving upriver to spawn and juveniles moving down to the East China Sea to mature (Wei et al. 1997). There are numerous ampullary receptors on the surface of the rostral paddle, which accounts for one-third of the body length in adults. These receptors are sensitive enough to detect weak electric fields created by individual zooplankton such as water fleas (*Daphnia*) from distances of up to 9 cm (Wilkens, et al. 2002).

We also included the eels (the "true" eels) (close to 20 species) in Anguillidae in our analysis. These fish are catadromous. Take the American eel (*Anguilla rostrata*), which spawn in the Sargasso Sea. The larvae do not metamorphose until they are 2-3 years old. Larvae float past the North American continent and migrate upriver, moving by selective tidal stream transport and transforming into transparent, miniature (50 mm long) eels known as glass eels. As they move upriver, they become pigmented and are called elvers. Elvers grow into juvenile yellow eels that then live in fresh water for 3-40 years. Mature eels turn into silvery-bronze color "silver eels", which then travel as much as 5,000 km back to the Sargasso Sea to spawn and then die (Avise, et al. 1990; Tesch et al. 1977).

Pacific salmon is a classic cases of anadromous migration. Adults lay eggs in rivers or streams. Eggs hatch into alevins, transform into silvery smolts that then move

out into the ocean. They grow into juveniles and adults and may cover distances of several thousand kilometers in the ocean. When they mature, they will migrate back to the river, even the exact place where they were incubated, cease feeding, change color, spawn and die (Quinn et al. 2006).

Methods

We collected migratory information of 972 fish species from fishbase.org (Froese and Pauly 2000). We also collected DNA sequences of 13 mitochondrial protein-coding genes of the same species from the National Center for Biotechnology Information (NCBI). Each gene was aligned and then concatenated together. Phylogeny from Betancur-R et al. (2013) was used as the "skeletal" phylogeny at the family level and concatenated sequences of 13 mitochondrial protein-coding genes were used to determine the relationship at the species level using RaxML (Stamatakis 2006). For the designated fish groups that we tested for positive natural selection, we kept only the target species and their close relatives. This resulted in a much smaller phylogenetic tree of about 40 species (and the corresponding DNA sequences) for each group. The branch-site model in the CODEML program as implemented in PAML (Yang 2007) was used to detect the signature of positive Darwinian selection. The branch-site model aims to detect episodic Darwinian selection along particular branches on a tree that affects only a few codons in a protein-coding gene, with selection measured by the nonsynonymous/synonymous rate ratio (ω = dN/dS) and positive Darwinian selection indicated by $\omega > 1$ (Yang and Nielsen 2002). Two models (model A and null model) were used on each target foreground branch. The model A allows partial sites having ω (dN/dS) larger than 1, while the null model allows all sites having ω equal to or smaller

than 1. Likelihood Ratio Test (twice the difference between model A and the null model) larger than 3.84 suggests that model A is significantly better than the null model, which provides evidence of positive Darwinian selection on the foreground branch.

Results

Among the 11 analyzed groups, a total of six different patterns of positive selection were detected (summarized in Figure 10). For the majority of them, the signature of positive selection was detected on the branch leading to the target group (Figure 10 a-d). There were also variations when positive selection occurred on the branch leading to the target group. For example, signature of positive selection appeared only on the branch leading to the target group (Figure 10a), as in gobies (Figure 9); evidence of positive selection appeared on branches leading to the target group and adjacent ancestor (Figure 10b), as in the ghost knifefish (Figure 2) and icefishes (Figure 5); evidence of positive selection appeared on branches leading to the target group and its sister group (Figure 10c), as in smelts (Figure 8); evidence of positive selection appeared on branches leading to the target group, sister group, and its adjacent ancestor (Figure 10d), as in flyingfishes (Figure 3) and sturgeons (Figure 6).

Three groups shared the pattern where no signature of positive selection was detected on the branch leading to the target group, while evidence of positive selection was detected on branches leading to its sister group and the adjacent ancestor (Figure 10e), as in elephantfish (Figure 1), salmons (Figure 8), and mackerels (Figure 9).

We did not find evidence of positive selection on the branches leading to the target group, sister group, or adjacent ancestor (Figure 10g) in two groups: cods (Figure 4) and eels (Figure 7).

Discussion

Positive Darwinian selection at a molecular level acts in a diversifying or a directional manner favoring specific changes during the adaptation to current or changing environments. Computational statistical methods are useful tools to detect evidence of positive selection on functionally-important genes and to narrow down the few amino acids that may be the targets of selection. However, current analyses are limited to single/few pre-specified branches of interest. Previous studies have focused almost exclusively on branches where high-energy demand presumably evolved. Here, we examined the evidence of positive selection in 11 groups of bony fishes with varied energy demands. We found various patterns of the distribution of positive selection and there seemed to be no simple unifying pattern: the occurrence of positive selection was case-specific and appeared to be determined by the evolutionary histories of particular lineages and tradeoffs between functional constraints that were specific to these lineages.

Zhang and Broughton (2015) found a different pattern of positive selection compared to previous studies. We found that the strength of positive selection was inversely proportional to apparent energy demand, being strongest on the least mobile taxa, while negative (purifying) selection was most intense on highly-mobile taxa. This pattern suggested an evolutionary scenario that a fairly efficient OXPHOS system had evolved probably under positive selection in the ancestors. However, over subsequent

time, purifying selection prevented further alteration. While low performance fishes do not need as much energy as tunas or billfishes, they could presumably tolerate more substitutions in the OXPHOS pathway. It is also possible that positive selection favored some substitutions that decrease OXPHOS efficiency. Here we tested the generality of such patterns. We found three groups, elephantfishes, salmons, and mackerels, which displayed a similar pattern of positive selection as found in Zhang and Broughton (2015). We found that, in 6 out of these 11 groups, the evidence for positive selection was detected on the branch directly leading to the target group. However, there were also variations: some positive selection was also detected on the branches leading to the sister groups, and/or the most common ancestral node. Thus, it is difficult to predict a specific relationship between positive selection and energetic demands or, indirectly, OXPHOS efficiency. When the signature of positive selection was found only on the branch leading to the target group (Figure 10a), that finding supports the hypothesis that positive selection favored substitutions that increased the OXPHOS efficiency during the evolutionary history from the most common ancestor to the target species. When the evidence of positive selection was found on the branches leading to the target group, in addition to the most recent common ancestor (Figure 10b), that finding suggests that OXPHOS adaptation began earlier in the lineage. When the signature of positive selection was found on branches leading to the target group and its sister group (Figure 10c), it could be that positive selection favored substitutions for divergent traits. And in the case of positive selection on branches leading to the target group, sister group and most recent common ancestor (Figure 10d), there could be a combination of factors involved.

Besides fish with high energy demands, we also analyzed two fish groups that are well-adapted to cold environments. Cold environments present challenges for ectothermic fishes because low temperatures slow down cellular processes, metabolic rates and performance. Therefore, many compensatory adaptations occur at the physiological and/or behavioral levels. The adjustment of aerobic scope is identified as a crucial step in thermal adaptation, and incremental change in mitochondrial aerobic capacity through variation of enzyme density and activity has been observed (Fischer 2003). There are many different ways that enzyme activity can be increased, including variability in DNA sequence, rates of transcription, translation, or protein modification. For example, significant incremental changes in red and white muscle COX activities were observed upon cold acclimation in rainbow trout (*Oncorhynchus mykiss*) without an increase of COX1 mRNA (Battersby and Moyes 1998). Similarly, significant incremental changes in white muscle COX activities, but not in the liver, were observed upon cold acclimation in cod (*Gadus morhua* L.) without significant increase in COX2 mRNA (Lucassen et al. 2006). In contrast, both mitochondrial-encoded COX1 and COX2 and nuclear-encoded COX4 mRNA were overrepresented, relative to enzyme capacities in white muscle and liver of the eurythermal common eelpout (*Pachycara brachycephalum*) during cold acclimation (Hardewig et al. 1999). Recently, Welch et al. (2014) also observed increased evolutionary rates in the mitochondrial cytochrome c oxidase I gene in polar bears (*Ursus maritimus*), but not its sister species, brown bears (*U. arctos*), which live in warmer climates.

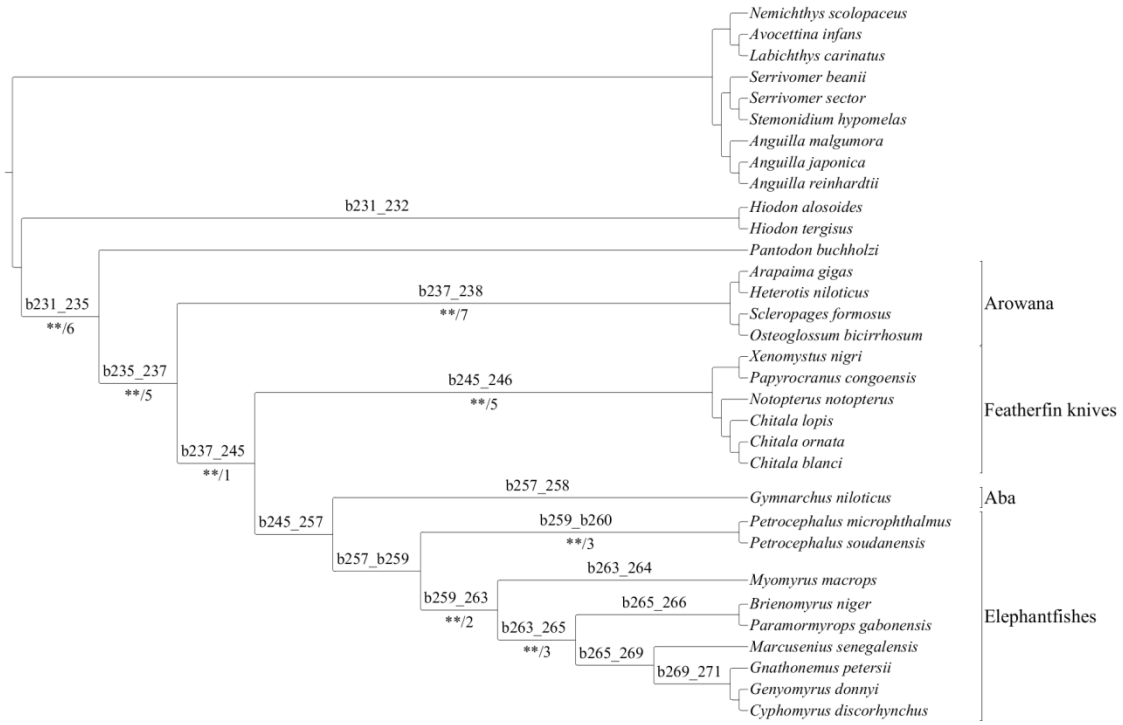
The two groups of fishes living in cold environments upon which we focused were cods living in the Arctic and icefishes living in the Antarctic oceans. We did not

find positive selection in cods, but did find positive selection in icefishes. There are two possible explanations for this result: 1) cods may have adapted through mechanisms not involving OXPHOS or not involving structural variation of OXPHOS proteins; or 2) the Arctic has undergone repeated warming and cooling until about 3 million years ago, when the present cold conditions stabilized, leaving less time for organisms to adapt to current conditions (Briggs 1995). Some species in sub-Arctic and Arctic waters may still be in transition to life in the permanent cold, while those in the Antarctic have developed features of permanent cold adaptation over a much longer period of time (20-25 million years) (Eastman 2013).

Because there are no direct measurements of oxygen utilization and efficiency of the OXPHOS pathway for each species or for the reconstructed common ancestors, the actual effects of different nucleotide substitutions on OXPHOS function remain unknown. The results presented here are strongly suggestive that such a relationship exists. However, investigations exploring the functional consequences of specific substitutions across a range of organisms will enhance our understanding of OXPHOS function and evolution.

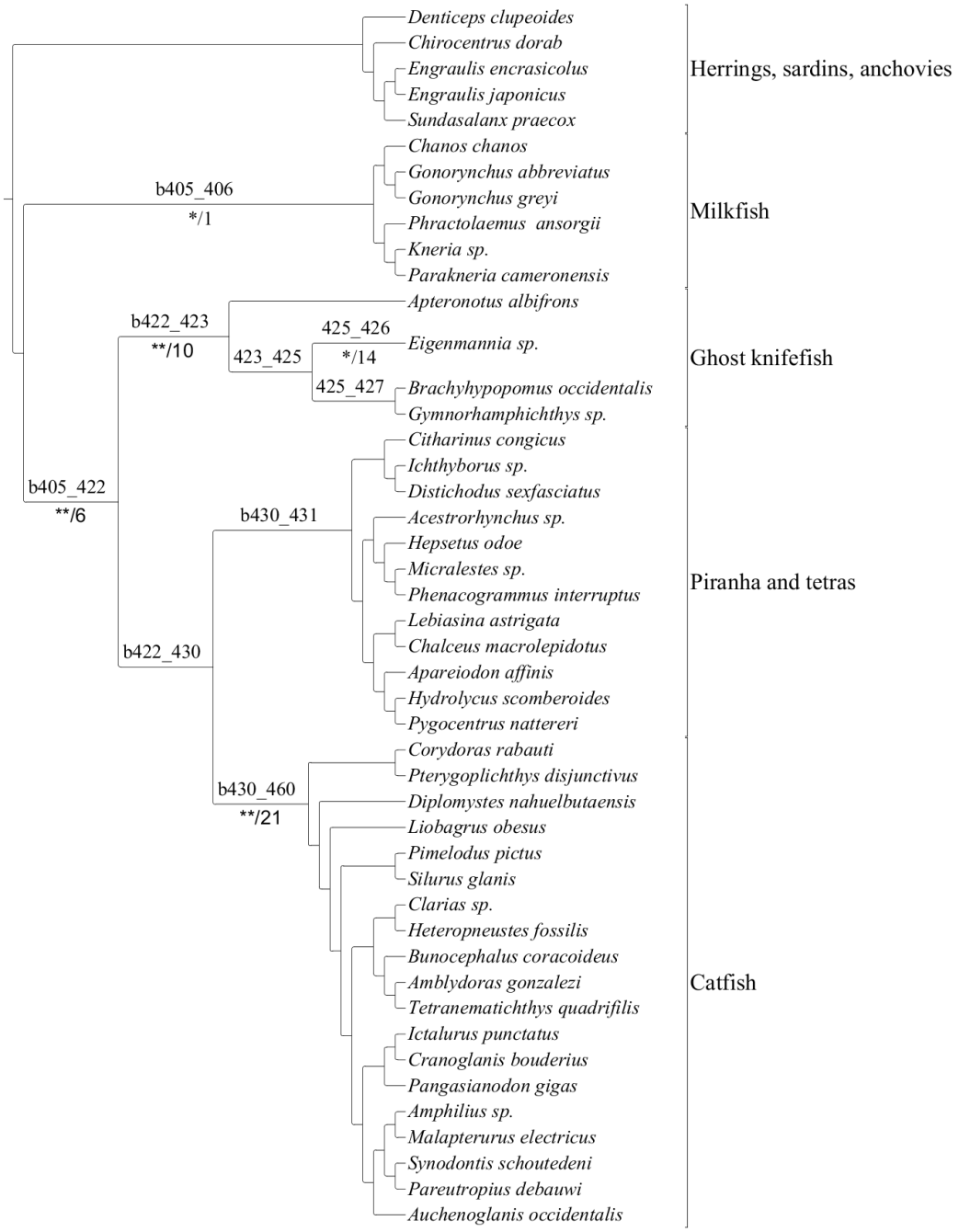
Figures

Figure 1 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b231_235, appear above each branch) of the elephantfish group.



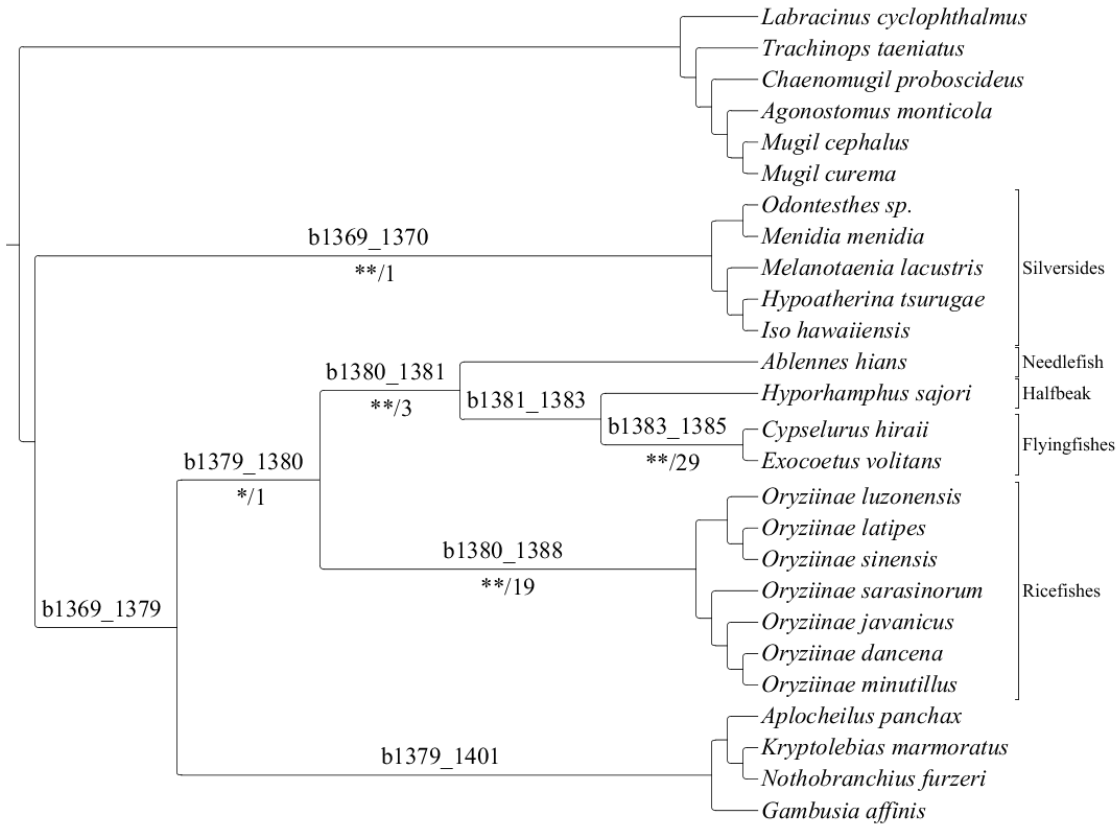
Note: "*" below branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below branch have no evidence of positive selection.

Figure 2 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b405_422, appear above each branch) of the knifefish group.



Note: "*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below the branch have no evidence of positive selection.

Figure 3 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b1369_1379, appear above each branch) of the flyingfish group.



Note: "*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labeled branches without information below the branch have no evidence of positive selection.

Figure 4 Positive selection analysis on 13 mitochondrial genes on branches (arbitrary labels, such as b1033_1035, appear above each branch) of the cod group.

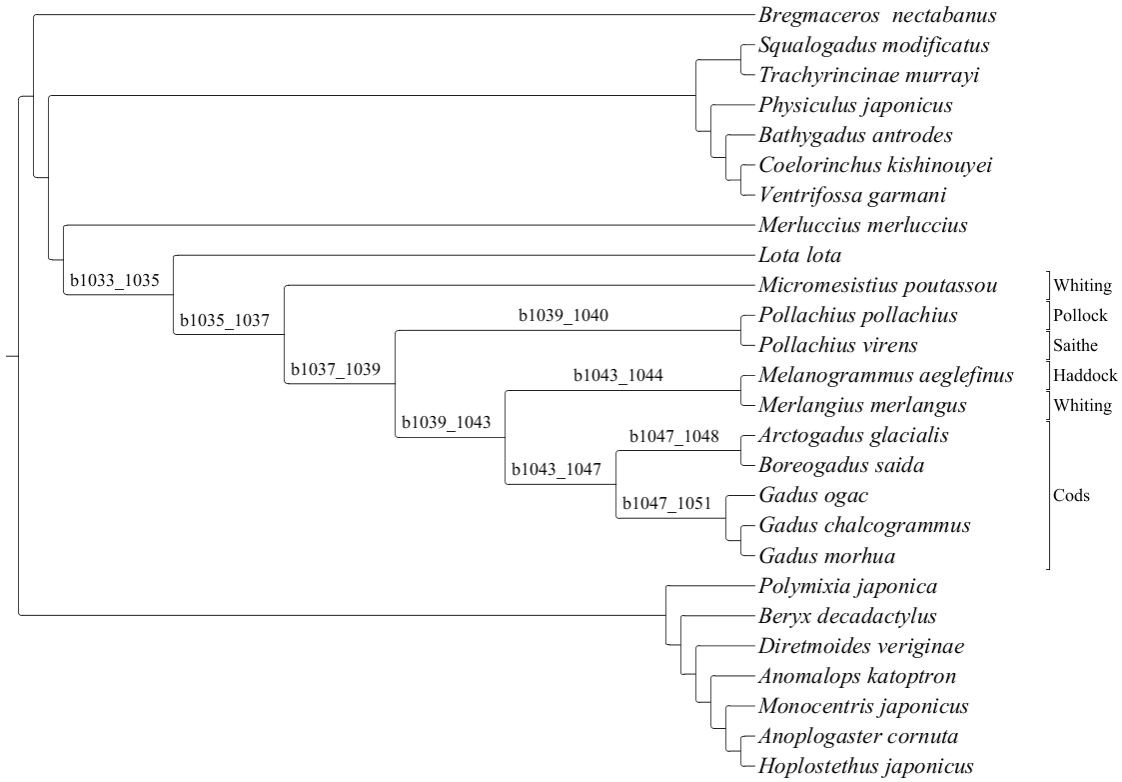
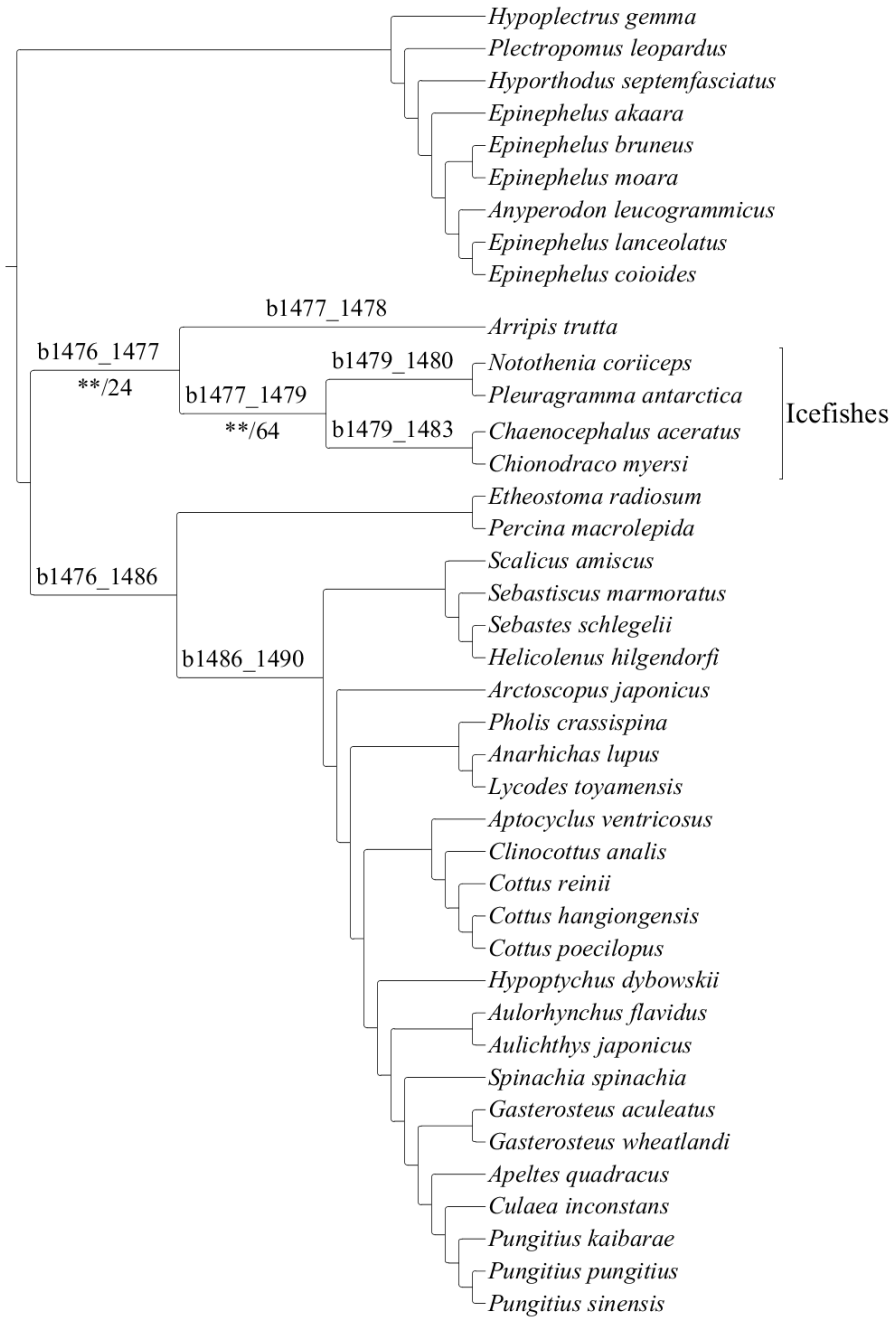
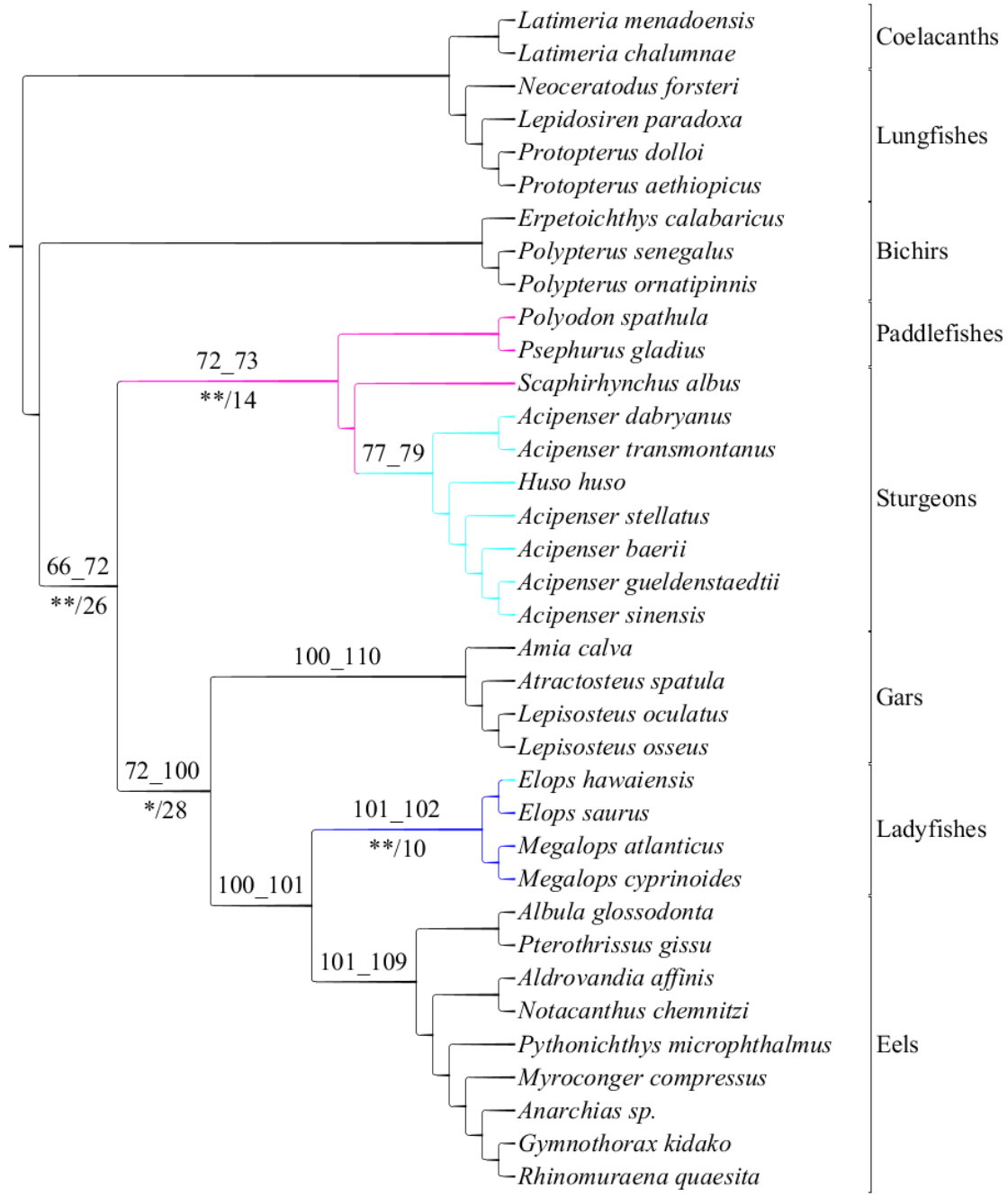


Figure 5 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b1476_1477, appear above each branch) of the icefish group.



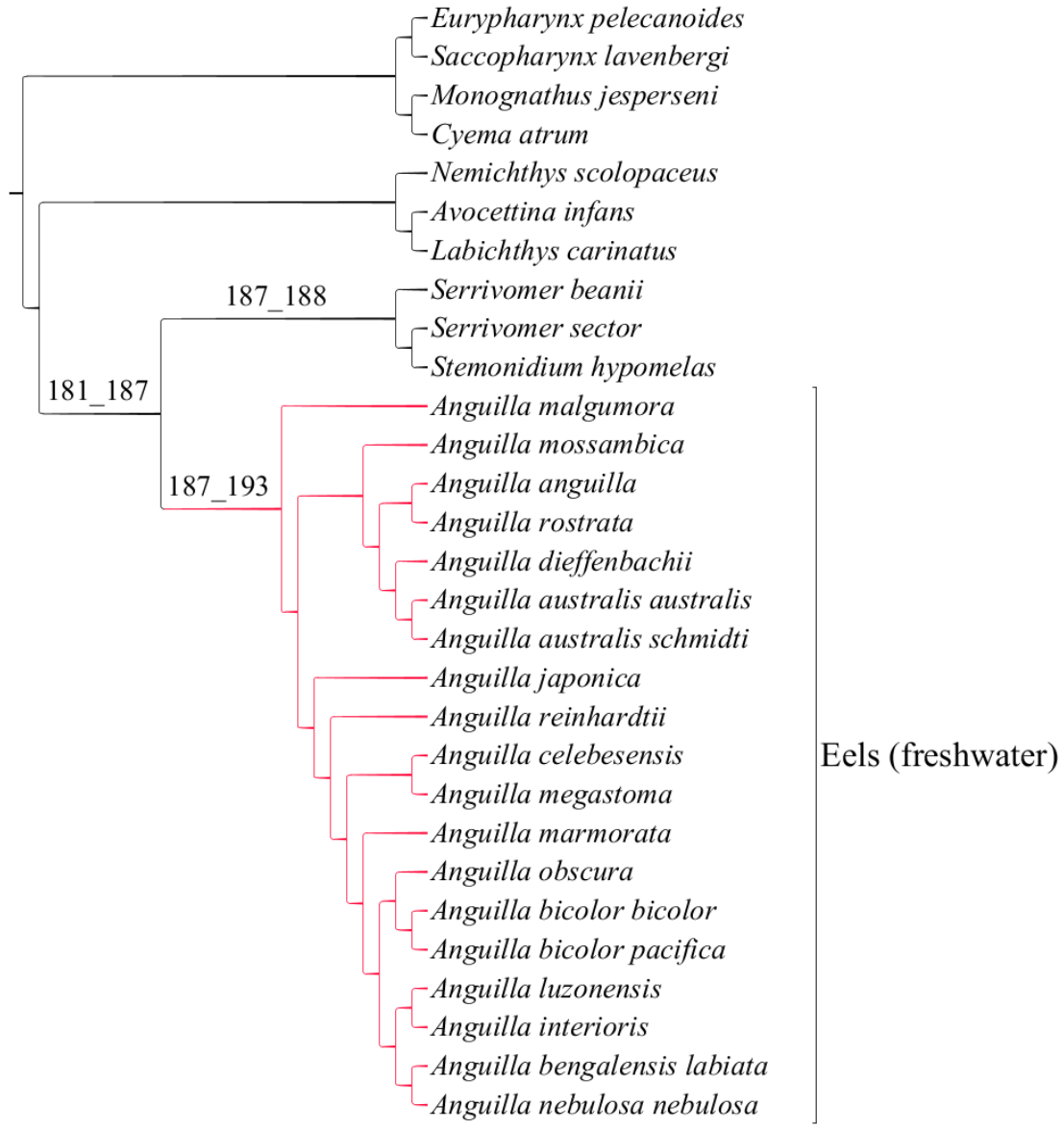
Note: "*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below the branch have no evidence of positive selection.

Figure 6 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b66_72, appear above each branch) of the sturgeon group.



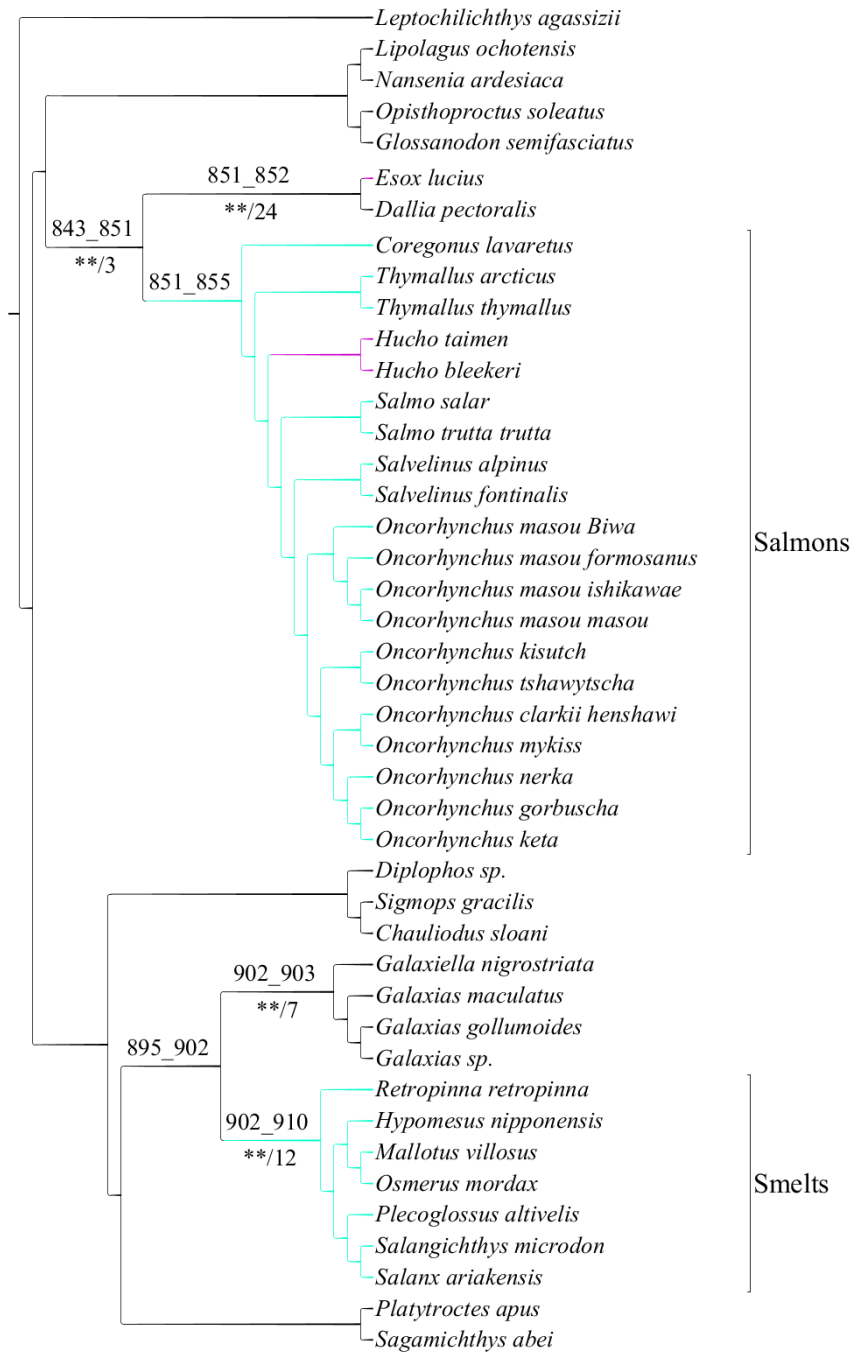
Note: "*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below the branch have no evidence of positive selection. Magenta indicates potamodromous; cyan indicates anadromous; blue indicates amphidromous.

Figure 7 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b181_187, appear above each branch) of the eel group.



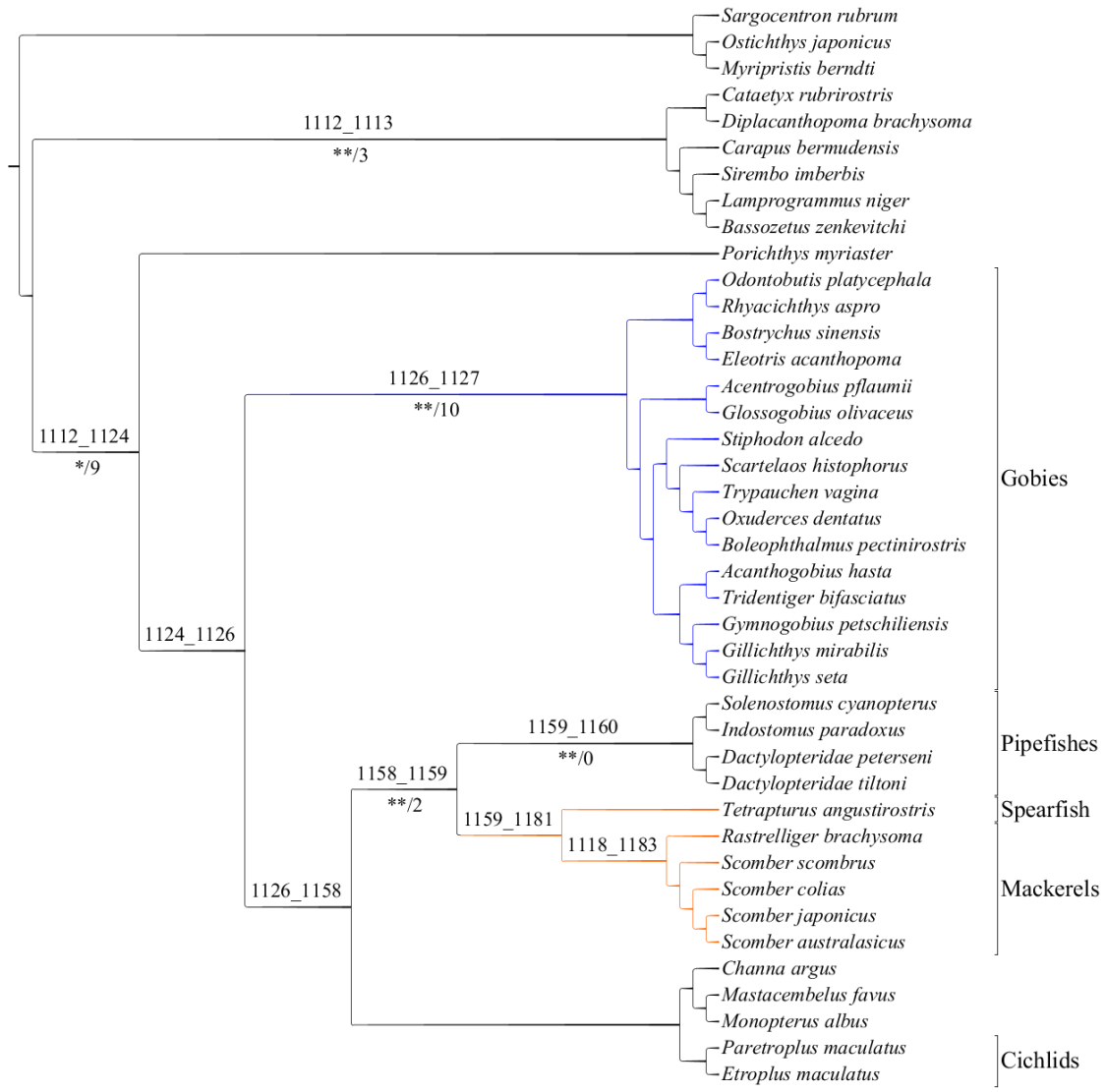
Note: red indicates catadromous

Figure 8 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b843_851, appear above each branch) of the salmon group.



Note: "*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below the branch have no evidence of positive selection. Magenta indicates potamodromous; cyan indicates anadromous.

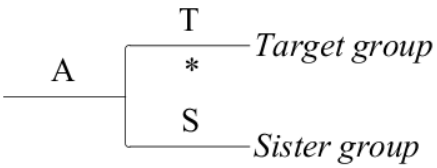
Figure 9 Positive selection analyses on 13 mitochondrial genes on branches (arbitrary labels, such as b1112_1124, appear above each branch) of the goby group.



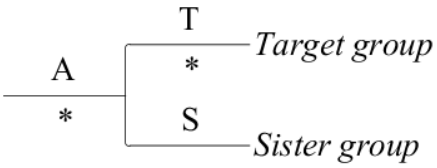
"*" below a branch indicates that the likelihood ratio tests are significant at 0.05, and "**" below branch indicates that the likelihood ratio tests are significant at 0.01; both indicate evidence of positive selection. The number below each branch after * or ** represents the number of positively selected sites with posterior probability higher than 0.8. Labled branches without information below branch have no evidence of positive selection. Blue indicates amphidromous; orange indicates oceanodromous.

Figure 10 Summary of patterns of positive selection.

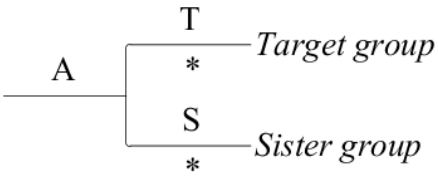
a. Signature of positive Darwinian selection was on branch leading to target group (as in gobies Figure 9)



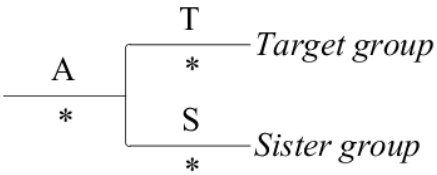
b. Signature of positive Darwinian selection was on branches leading to target group and ancestral node (as in ghost knifefish (Figure 2) and icefishes (Figure 5))



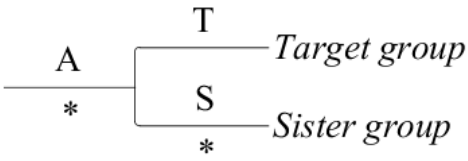
c. Signature of positive Darwinian selection was on branches leading to target group and sister group (as in smelts (Figure 8))



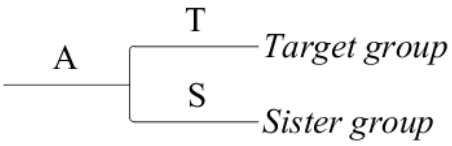
d. Signature of positive Darwinian selection was on branches leading to target group, sister group and most recent common ancestor (as in flyingfishes (Figure 3) and sturgeons (Figure 6))



e. Signature of positive Darwinian selection was on branches leading to sister group and ancestral node (as in elephantfish (Figure 1), salmons (Figure 8), and mackerels (Figure 9))



f. Signature of positive Darwinian selection was not detected on branches leading to target group, sister group or ancestral node (as in cods (Figure 4) and eels (Figure 7))



References

Adkins RM, Honeycutt RL 1994. Evolution of the primate cytochrome c oxidase subunit II gene. Journal of Molecular Evolution 38: 215-231.

Allen AP, Gillooly JF, Savage VM, Brown JH 2006. Kinetic effects of temperature on rates of genetic divergence and speciation. Proceedings of the National Academy of Sciences USA103: 9130-9135.

Altringham JD, Johnston IA 1990. Modelling muscle power output in a swimming fish. Journal of Experimental Biology 148: 395-402.

Amsterdam A, et al. 2004. Identification of 315 genes essential for early zebrafish development. Proceedings of the National Academy of Sciences USA 101: 12792-12797. doi: 10.1073/pnas.0403929101

Andrews TD, Easteal S 2000. Evolutionary rate acceleration of cytochrome c oxidase subunit I in simian primates. Journal of Molecular Evolution 50: 562-568.

Archer SD, Johnston IA 1991. Density of cristae and distribution of mitochondria in the slow muscle fibers of Antarctic fish. Physiological Zoology: 242-258.

Avise JC, Nelson WS, Amold J, Koehn RK, Williams GC, Thorsteinsson V. 1990. The evolutionary genetic status of Icelandic eels. Evolution 44: 1254-1262.

Ballard J, Melvin R 2010. Linking the mitochondrial genotype to the organismal phenotype. Molecular Ecology 19: 1523-1539.

Ballard JWO, Melvin RG, Katewa SD, Maas K 2007. Mitochondrial DNA variation is associated with measurable differences in life history traits and mitochondrial metabolism in *Drosophila simulans*. Evolution 61: 1735-1747.

Ballard JWO, Rand DM 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annual Review of Ecology, Evolution, and Systematics 36: 621-642.

Ballard JWO, Whitlock MC 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744.

Baradaran R, Berrisford JM, Minhas GS, Sazanov LA 2013. Crystal structure of the entire respiratory complex I. Nature 494: 443-448.

Bargelloni L, Marcato S, Patarnello T 1998. Antarctic fish hemoglobins: evidence for adaptive evolution at subzero temperature. Proceedings of the National Academy of Sciences USA 95: 8670-8675.

Barrett RD, et al. 2010. Rapid evolution of cold tolerance in stickleback. Proceedings of the Royal Society B: Biological Sciences: rspb20100923.

Battersby BJ, Moyes CD 1998. Influence of acclimation temperature on mitochondrial DNA, RNA, and enzymes in skeletal muscle. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 275: R905-R912.

Bazin E, Glémin S, Galtier N 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312: 570-572.

Beal M 1996. Mitochondria, free radicals, and neurodegeneration. Current Opinion in Neurobiology 6: 661-666. doi: 10.1016/s0959-4388(96)80100-0

Beamish F. 1978. Swimming Capacity: Academic Press.

Benjamini Y 2010. Discovering the false discovery rate. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72: 405-416.

Betancur-R R, et al. 2013. The Tree of Life and a new classification of bony fishes. PLoS Currents Tree of Life.013 Apr 18. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Binkowski FP, Doroshov SI 1985. North American Sturgeons: Biology and Aquaculture Potential. Dr W. Junk Publishers, Dordrecht(Netherlands). 1985.

Birky C 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annual Review of Genetics 35: 125-148. doi: 10.1146/annurev.genet.35.102401.090231

Blake R 1983. Energetics of leaping in dolphins and other aquatic animals. Journal of the Marine Biological Association of the United Kingdom 63: 61-70.

Blank JM, Farwell CJ, Morrissette JM, Schallert RJ, Block BA 2007. Influence of swimming speed on metabolic rates of juvenile Pacific bluefin tuna and yellowfin tuna. Physiological and Biochemical Zoology 80: 167-177.

Block BA 1986. Structure of the brain and eye heater tissue in marlins, sailfish, and spearfishes. Journal of Morphology 190: 169-189.

Block BA, Booth D, Carey FG 1992. Direct measurement of swimming speeds and depth of blue marlin. Journal of Experimental Biology 166: 267-284.

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, Finnerty JR, Stewart AF, Kidd J 1993. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. Science 260: 210-210.

Boekema E, Braun H-P 2007. Supramolecular structure of the mitochondrial oxidative phosphorylation system. The Journal of Biological Chemistry 282: 1-4. doi: 10.1074/jbc.R600031200

Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J 2012. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biology 13: R122.

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics doi: 10.1093/bioinformatics/btu170.

Brand M 2005. The efficiency and plasticity of mitochondrial energy transduction. Biochemical Society Transactions 33(5): 897-904.

Briggs JC. 1995. Global Biogeography: Elsevier.

Brill RW, Bushnell PG 1991. Metabolic and cardiac scope of high energy demand teleosts, the tunas. Canadian Journal of Zoology 69: 2002-2009.

Brown W, George M, Wilson A 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences USA76: 1967-1971.

Burghardt GM. 2005. The Genesis of Animal Play: Testing the Limits: MIT Press.

Caccone A, et al. 2004. Extreme difference in rate of mitochondrial and nuclear DNA evolution in a large ectotherm, Galápagos tortoises. Molecular Phylogenetics and Evolution 31: 794-798. doi: 10.1016/j.ympev.2004.02.004

Cairns J, Overbaugh J, Miller S 1988. The origin of mutants. Nature 335: 142-145.

Campos Y, et al. 1997. Bilateral striatal necrosis and MELAS associated with a new T3308C mutation in the mitochondrial ND1 gene. Biochemical and Biophysical Research Communications 238: 323-325.

Carroll AM, Wainwright PC 2003. Functional morphology of prey capture in the sturgeon, *Scaphirhynchus albus*. Journal of Morphology 256: 270-284.

Casteel RW 1976. Fish Remains in Archaeology and Paleo-environmental Studies. Academic Press.

Cecchini G 2003. Function and structure of complex II of the respiratory chain*. Annual Review of Biochemistry 72: 77-109.

Comeron J, Kreitman M 1998. The correlation between synonymous and nonsynonymous substitutions in *Drosophila*: mutation, selection or relaxed constraints? Genetics 150: 767-775.

Coppe A, et al. 2013. Genome evolution in the cold: Antarctic icefish muscle transcriptome reveals selective duplications increasing mitochondrial function. Genome Biology and Evolution 5: 45-60.

Croteau D, Bohr V 1997. Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. Journal of Biological Chemistry 272: 25409-25412. doi: 10.1074/jbc.272.41.25409

da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A 2008. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 9: 119.

Dalziel AC, Moyes CD, Fredriksson E, Lougheed SC 2006. Molecular evolution of cytochrome c oxidase in high-performance fish (Teleostei: Scombroidei). Journal of Molecular Evolution 62: 319-331.

Davenport J 1994. How and why do flying fish fly? Reviews in Fish Biology and Fisheries 4: 184-214.

de Bruijn NG 1946. A combinatorial problem. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series A 49: 758.

De Grassi A, et al. 2005. Evolution of nuclearly encoded mitochondrial genes in Metazoa. Gene 354: 181-188.

Demple B, Harrison L 1994. Repair of oxidative damage to DNA: enzymology and biology. Annual Review of Biochemistry 63: 915-948.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science 289: 2342-2344.

Dickson KA 1995. Unique adaptations of the metabolic biochemistry of tunas and billfishes for life in the pelagic environment. Environmental Biology of Fishes 42: 65-97.

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

Doan JW et al. 2004. Coadaptive evolution in cytochrome c oxidase: 9 of 13 subunits show accelerated rates of nonsynonymous substitution in anthropoid primates. Molecular Phylogenetics and Evolution 33: 944.

Dunn K, Bielawski J, Yang Z 2001. Substitution rates in *Drosophila* nuclear genes: implications for translational selection. Genetics 157: 295-305.

Eastman JT. 2013. Antarctic fish biology: evolution in a unique environment: Academic Press.

Ebeling, Alfred W., Neils B. Atkin, and Paulette Y. Setzer. Genome sizes of teleostean fishes: increases in some deep-sea species. American Naturalist(1971): 549-561.

Efremov RG, Sazanov LA 2011. Structure of the membrane domain of respiratory complex I. Nature 476: 414-420.

Evans DH 2002. Cell signaling and ion transport across the fish gill epithelium. Journal of Experimental Zoology 293: 336-347.

Evans M, Bernatchez L 2012. Oxidative phosphorylation gene transcription in whitefish species pairs reveals patterns of parallel and nonparallel physiological divergence. Journal of Evolutionary Biology 25: 1823-1834.

Farrell AP 1991. Invited perspectives in physiological zoology. From hagfish to tuna: a perspective on cardiac function in fish. Physiological Zoology 64: 1137-1164.

Farrell AP, Steffensen JF. 2005. Fish Physiology: Physiology of Polar Fishes: Gulf Professional Publishing.

Felsenstein J 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution: 783-791.

Felsenstein J 1988. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521-565.

Fiol DF, Kültz D 2005. Rapid hyperosmotic coinduction of two tilapia (*Oreochromis mossambicus*) transcription factors in gill cells. Proceedings of the National Academy of Sciences USA 102: 927-932.

Fischer T 2003. The effects of climate induced temperature changes on cod (*Gadus morhua* L.): linking ecological and physiological investigations= Die Effekte klimabedingter Temperaturveränderungen auf den Kabeljau (*Gadus morhua* L.): eine Verknüpfung von ökologischen und physiologischen Untersuchungen. Berichte zur Polar-und Meeresforschung (Reports on Polar and Marine Research) 454.

Foote AD, et al. 2011. Positive selection on the killer whale mitogenome. Biology Letters 7: 116-118.

Friedberg E, Walker G, Siede W, PVD 1995. DNA repair and mutagenesis. Trends in Biochemical Sciences 20: 440.

Froese R, Pauly D (eds). 2000. Fishbase: a global information system on fishes: FishBase. World Wide Web electronic publication. www.fishbase.org.

Gabriel W, Lynch M, Burger R 1993. Muller's ratchet and mutational meltdowns. Evolution: 1744-1757.

Garvin MR, Bielawski JP, Gharrett AJ 2011. Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. PloS One 6: e24127.

Garvin MR, Bielawski JP, Sazanov LA, Gharrett AJ 2015. Review and meta analysis of natural selection in mitochondrial complex I in metazoans. Journal of Zoological Systematics and Evolutionary Research 53: 1-17.

Gershoni M, Templeton A, Mishmar D 2009. Mitochondrial bioenergetics as a major motive force of speciation. BioEssays 31: 642-650. doi: 10.1002/bies.200800139

Gillespie JH. 1991. The Causes of Molecular Evolution: Oxford University Press.

Gillman LN, Keeling DJ, Ross HA, Wright SD 2009. Latitude, elevation and the tempo of molecular evolution in mammals. Proceedings of the Royal Society B: Biological Sciences 276: 3353-3359.

Gillooly JF, Allen AP, West GB, Brown JH 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proceedings of the National Academy of USA 102: 140-145.

Goldberg A, et al. 2003. Adaptive evolution of cytochrome c oxidase subunit VIII in anthropoid primates. Proceedings of the National Academy of Sciences USA 100: 5873-5878. doi: 10.1073/pnas.0931463100

Grabherr MG, et al. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature Biotechnology 29: 644.

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.

Gross MR editor. American Fisheries Society Symposium. 1987.

Grossman LI, Schmidt TR, Wildman DE, Goodman M 2001. Molecular evolution of aerobic energy metabolism in primates. Molecular Phylogenetics and Evolution 18: 26-36.

Haag-Liautard C, et al. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biology 6: e204.

Harmon L, et al. 2009. geiger: Analysis of evolutionary diversification. R package version 1.

Hassanin A, Ropiquet A, Couloux A, Cruaud C 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the

tribe Caprini (Bovidae, Antilopinae). Journal of Molecular Evolution 68: 293-310.

Hebert PD, Remigio EA, Colbourne JK, Taylor DJ, Wilson CC 2002. Accelerated molecular evolution in halophilic crustaceans. Evolution 56: 909-926.

Helfman G, Collette BB, Facey DE, Bowen BW. 2009. The Diversity of Fishes: Biology, Evolution, and Ecology: John Wiley & Sons.

Herrera M, et al. 2012. Physiological short-term response to sudden salinity change in the Senegalese sole (*Solea senegalensis*). Fish Physiology and Biochemistry 38: 1741-1751.

Hill GE 2014. Cellular respiration: the nexus of stress, condition, and ornamentation. Integrative and Comparative Biology 54: 645-657.

Hofmann GE, Buckley BA, Airaksinen S, Keen JE, Somero GN 2000. Heat-shock protein expression is absent in the Antarctic fish *Trematomus bernacchii* (family Nototheniidae). Journal of Experimental Biology 203: 2331-2339.

Hudson AP, DeVries AL, Haschemeyer AE 1979. Antifreeze glycoprotein biosynthesis in Antarctic fishes. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry 62: 179-183.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294: 2310-2314.

Janssen RJ, Nijtmans LG, van den Heuvel LP, Smeitink JA 2006. Mitochondrial complex I: structure, function and pathology. Journal of Inherited Metabolic Disease 29: 499-515.

Johnson K, et al. 2003. Dramatically elevated rate of mitochondrial substitution in lice (Insecta: Phthiraptera). Molecular Phylogenetics and Evolution 26: 231-242. doi: 10.1016/s1055-7903(02)00342-1

Kaltenpoth M, et al. 2012. Accelerated evolution of mitochondrial but not nuclear genomes of Hymenoptera: new evidence from crabronid wasps. PloS One 7. doi: 10.1371/journal.pone.0032826

Keightley PD 2012. Rates and fitness consequences of new mutations in humans. Genetics 190: 295-304.

Kenyon L, Moraes C 1997. Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. Proceedings of the National Academy of Sciences of the USA 94: 9131-9135. doi: 10.2307/42569

Koch F, Wieser W 1983. Partitioning of energy in fish: Can reduction of swimming activity compensate for the cost of production? Journal of Experimental Biology 107: 141-146.

Kocher TD, et al. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences USA 86: 6196-6200.

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

Laiz-Carrión R, et al. 2002. Energy metabolism in fish tissues related to osmoregulation and cortisol action. Fish Physiology and Biochemistry 27: 179-188.

Lauder GV, Madden PG, Mittal R, Dong H, Bozkurttas M 2006. Locomotion with flexible propulsors: I. Experimental analysis of pectoral fin swimming in sunfish. Bioinspiration & Biomimetics 1: S25.

Lenn T, Leake M, Mullineaux C 2008. Are *Escherichia coli* OXPHOS complexes concentrated in specialized zones within the plasma membrane? Biochemical Society Transactions 36: 1032-1036.

Li C, Ortí G, Zhang G, Lu G 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evolutionary Biology 7: 44.

Lightowlers R, Chinnery P, Turnbull D, Howell N 1997. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends in Genetics : 13: 450-455. doi: 10.1016/s0168-9525(97)01266-3

Little A, Lougheed S, Moyes C 2010. Evolutionary affinity of billfishes (Xiphiidae and Istiophoridae) and flatfishes (Plueronectiformes): Independent and trans-subordinal origins of endothermy in teleost fishes. Molecular Phylogenetics and Evolution 56: 897-904.

Lucassen M, Koschnick N, Eckerle L, Pörtner H-O 2006. Mitochondrial mechanisms of cold adaptation in cod (*Gadus morhua* L.) populations from different climatic zones. Journal of Experimental Biology 209: 2462-2471.

Lynch M 2007. The Origins of Genome Architecture. Sinauer Associates.

Maddison WP, Maddison D 2001. Mesquite: a modular system for evolutionary analysis.

Markham MR, Kaczmarek LK, Zakon HH 2013. A sodium-activated potassium channel supports high-frequency firing and reduces energetic costs during rapid modulations of action potential amplitude. Journal of Neurophysiology 109: 1713-1723.

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

Martins EP, Hansen TF 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist: 646-667.

McDowall R editor. 1987. American Fisheries Society Symposium.

McKenzie M, Chiotis M, Pinkert CA, Trounce IA 2003. Functional respiratory chain analyses in murid xenomitochondrial cybrids expose coevolutionary constraints of cytochrome b and nuclear subunits of complex III. Molecular Biology and Evolution 20: 1117-1124.

Meiklejohn CD, Montooth KL, Rand DM 2007. Positive and negative selection on the mitochondrial genome. Trends in Genetics 23: 259-263.

Meredith RW, et al. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334: 521-524.

Mishmar D, et al. 2003. Natural selection shaped regional mtDNA variation in humans. Proceedings of the National Academy of Sciences USA 100: 171-176.

Miya M, et al. 2013. Evolutionary origin of the scombridae (tunas and mackerels): Members of a paleogene adaptive radiation with 14 other pelagic fish families. PloS One 8: e73535.

Moon T, Altringham J, Johnston I 1991. Energetics and power output of isolated fish fast muscle fibres performing oscillatory work. Journal of Experimental Biology 158: 261-273.

Moritz C, Dowling T, Brown W 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annual Review of Ecology and Systematics 18: 269-292.

Moyer GR, Winemeller KO, McPhee MV, Turner TF 2005. Historical demography, selection, and coalescence of mitochondrial and nuclear genes in Prochilodus species of northern South America. Evolution 59: 599-610.

Moyes CD, Mathieu-Costello OA, Brill RW, Hochachka PW 1992. Mitochondrial metabolism of cardiac and skeletal muscles from a fast (*Katsuwonus pelamis*) and a slow (*Cyprinus carpio*) fish. Canadian Journal of Zoology 70: 1246-1253.

Mozo J, Emre Y, Bouillaud F, Ricquier D, Criscuolo F 2005. Thermoregulation: what role for UCPs in mammals and birds? Bioscience Reports 25: 227-249.

Muller HJ 1950. Our load of mutations. American Journal of Human Genetics 2: 111.

Nagase M, Aimi T, Suginaka K, Kitamoto Y, Morinaga T 2005. Complete mitochondrial DNA sequence of the Japanese flying fish *Cypselurus hiraii*. Fisheries Science 71: 914-923.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics: Oxford University Press.

Niehuis O, Judson A, Gadau J 2008. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids *of Nasonia giraulti* and N. *vitripennis*. Genetics 178: 413-426. doi: 10.1534/genetics.107.080523

Nielsen R, Yang Z 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929-936.

Nikolaev SI, et al. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. Proceedings of the National Academy of Sciences USA 104: 20443-20448.

Nilsson G 1996. Brain and body oxygen requirements of *Gnathonemus petersii*, a fish with an exceptionally large brain. Journal of Experimental Biology 199: 603-607.

Ohta T, Kimura M 1973. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetical Research 22: 201-204.

Opdal SH, Egeland T, Musse MA, Rognum TO 2002. Possible role of mtDNA mutations in sudden infant death. PediatricNneurology 27: 23-29.

Osada N, et al. 2002. Search for genes positively selected during primate evolution by 5′-end-sequence screening of cynomolgus monkey cDNAs. Genomics 79: 657-662.

Palumbi SR 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357: 153.

Paradis E. 2011. Analysis of Phylogenetics and Evolution with R: Springer Science & Business Media.

Pazos F, Valencia A 2008. Protein co-evolution, co-adaptation and interactions. EMBO Journal 27: 2648-2655. doi: 10.1038/emboj.2008.189

Pegram LM, et al. 2010. Why Hofmeister effects of many salts favor protein folding but not DNA helix formation. Proceedings of the National Academy of Sciences USA 107: 7716-7721.

Pichaud N, Ballard JWO, Tanguay RM, Blier PU 2012. Naturally occurring mitochondrial DNA haplotypes exhibit metabolic differences: insight into functional properties of mitochondria. Evolution 66: 3189-3197.

Pinheiro J, Bates D, DebRoy S, Sarkar D 2007. Linear and nonlinear mixed effects models. R package version 3: 57.

Pollock D, Taylor W, Goldman N 1999. Coevolving protein residues: maximum likelihood identification and relationship to structure. Journal of Molecular Biology 287: 187-198. doi: 10.1006/jmbi.1998.2601

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. Proceedings of the National Academy of Sciences USA 104: 13390-13395.

Quinn TP, Stewart IJ, Boatright CP 2006. Experimental evidence of homing to site of incubation by mature sockeye salmon, *Oncorhynchus nerka*. Animal Behaviour 72: 941-949.

Rand D, Haney R, Fry A 2004. Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology & Evolution 19: 645-653. doi: 10.1016/j.tree.2004.10.003

Rand DM 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends in Ecology & Evolution 9: 125-131.

Rdevelopment C 2008. TEAM. 2008. R: Language and environment for statistical computing. R Foundation for Statistical Computing. Vienne. disponible en ligne à: r-project. org.[Visité le 09-10-03].

Riedman M 1990. The pinnipeds: seals, sea lions, and walruses. No. 12. Univ of California Press.

Richter O-M, Ludwig B 2003a. Cytochrome c oxidase—structure, function, and physiology of a redox-driven molecular machine. Reviews of physiology, biochemistry and pharmacology: 47-74.

Richter O-M, Ludwig B. 2003b. Cytochrome c oxidase—structure, function, and physiology of a redox-driven molecular machine. In. Reviews of physiology, biochemistry and pharmacology: Springer. p. 47-74.

Rohde K 1992. Latitudinal gradients in species diversity: the search for the primary cause. Oikos: 514-527.

Ronquist F, Huelsenbeck JP 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Rosen RA, Hales DC 1981. Feeding of paddlefish, *Polyodon spathula*. Copeia: 441-455.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303: 223-226.

Ruud JT 1954. Vertebrates without erythrocytes and blood pigment. Nature 173: 848-850

Sackton T, Haney R, Rand D 2003. Cytonuclear coadaptation in Drosophila: disruption of cytochrome c oxidase activity in backcross genotypes. Evolution 57: 2315-2325. doi: 10.2307/3448782

Salazar VL, Krahe R, Lewis JE 2013. The energetics of electric organ discharge generation in gymnotiform weakly electric fish. Journal of Experimental Biology 216: 2459-2468.

Saraste M 1999. Oxidative phosphorylation at the fin de siècle. Science 283: 1488-1493. doi: 10.1126/science.283.5407.1488

Scheffler IE 1999. Mitochondrial electron transport and oxidative phosphorylation. Mitochondria: 141-245.

Schmidt TR, Goodman M, Grossman LI 1999. Molecular evolution of the COX7A gene family in primates. Molecular Biology and Evolution 16: 619-626.

Schmidt TR, Wu W, Goodman M, Grossman LI 2001. Evolution of nuclear-and mitochondrial-encoded subunit interaction in cytochrome c oxidase. Molecular Biology and Evolution 18: 563-569.

Schriefer JE, Hale ME 2004. Strikes and startles of northern pike (*Esox lucius*): a comparison of muscle activity and kinematics between S-start behaviors. Journal of Experimental Biology 207: 535-544.

Schulz MH, Zerbino DR, Vingron M, Birney E 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28: 1086-1092.

Scott WB, Crossman EJ 1973. Freshwater Fishes of Canada. Fisheries Research Board of Canada Bulletin 184.

Shadel G, Clayton D 1997. Mitochondrial DNA maintenance in vertebrates. Annual Review of Biochemistry 66: 409-435. doi: 10.1146/annurev.biochem.66.1.409

Sharp P, Li W 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Molecular Biology and Evolution 4: 222-230.

Shen Y-Y, et al. 2010. Adaptive evolution of energy metabolism genes and the origin of flight in bats. Proceedings of the National Academy of Sciences USA 107: 8666-8671.

Simon B, et al. 1992. Muscle energetics in yellow freshwater eels (*Anguilla anguilla L.*) exposed to high hydrostatic pressure (101 ATA) for 30 days. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry102.2: 205-208.

Simon DK, et al. 2003. A heteroplasmic mitochondrial complex I gene mutation in adult-onset dystonia. Neurogenetics 4: 199-205.

Singh M, Gupta A, Lakra W 2012. In silico 3-D structure prediction of cytochrome b protein of sisorid catfish *Glyptothorax ngapang*. Indian Journal of Biotechnology 11: 156-162.

Smith TI 1985. The fishery, biology, and management of Atlantic sturgeon, *Acipenser oxyrhynchus*, in North America. Environmental Biology of Fishes 14: 61-72.

Somero GN 2002. Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. Integrative and Comparative Biology 42: 780-789.

Stamatakis A 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690. doi: 10.1093/bioinformatics/btl446

Streelman J, Karl SA 1997. Reconstructing labroid evolution with single–copy nuclear DNA. Proceedings of the Royal Society of London. Series B: Biological Sciences 264: 1011-1020.

Sun Y-B, Shen Y-Y, Irwin DM, Zhang Y-P 2011. Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution. Molecular Biology and Evolution 28: 39-44.

Tang J, Wardle C 1992. Power output of two sizes of Atlantic salmon (*Salmo salar*) at their maximum sustained swimming speeds. Journal of Experimental Biology 166: 33-46.

Tesch F, Greenwood J, Greenwood P, Henderson I. 1977. The Eel: Biology and Management of Anguillid Eel: English Ed: Chapman and Hall.

Thaler DS 1994. The evolution of genetic intelligence. Science 264: 224.

Theron M, Guerrero F, and Sebert P. 2000. Improvement in the efficiency of oxidative phosphorylation in the freshwater eel acclimated to 10.1 MPa hydrostatic pressure. Journal of Experimental Biology 203.19: 3019-3023.

Tsukihara T, et al. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Science. 272: 1136-1144.

Ugalde C, Janssen RJ, van den Heuvel LP, Smeitink JA, Nijtmans LG 2004. Differences in assembly or stability of complex I and other mitochondrial OXPHOS complexes in inherited complex I deficiency. Human Molecular Genetics 13: 659-667.

Urschel MR, O'Brien KM 2008. High mitochondrial densities in the hearts of Antarctic icefishes are maintained by an increase in mitochondrial size rather than mitochondrial biogenesis. Journal of Experimental Biology 211: 2638-2646.

Van Winkle W (ed) 2000. Biology, Management, and Protection of Sturgeon American Fisheres Society Symposium 28.

Videler J, Nolet B 1990. Costs of swimming measured at optimum speed: scale effects, differences between swimming styles, taxonomic groups and submerged and surface swimming. Comparative Biochemistry and Physiology Part A: Physiology 97: 91-99.

Vilella AJ, et al. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Research 19: 327-335.

Voet D, Voet JG, Pratt CW 1999. Fundamentals of Biochemistry. Chap. V, Wiley & Sons, Chichester: 515.

Vogel S. 1994. Life in Moving Fluids: The Physical Biology of Flow: Princeton University Press: Princeton.

Wainwright PC, Bellwood DR, Westneat MW 2002. Ecomorphology of locomotion in labrid fishes. Environmental Biology of Fishes 65: 47-62.

Walker JA, Westneat MW 2002. Performance limits of labriform propulsion and correlates with fin shape and motion. Journal of Experimental Biology 205: 177-187.

Walker JE 1992. The NADH: ubiquinone oxidoreductase (complex I) of respiratory chains. Quarterly Reviews of Biophysics 25: 253-324.

Wallace DC 2010. Mitochondrial DNA mutations in disease and aging. Environmental and Molecular Mutagenesis 51: 440-450.

Ward LD, Kellis M 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science 337: 1675-1678.

Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiological Genomics 2: 143-147.

Webb PW 1994. The biology of fish swimming. Pp. 45-62. In: Mechanics and Physiology of Animal Swimming. Cambridge University Press, Cambridge.

Wei Q, et al. 1997. Biology, fisheries, and conservation of sturgeons and paddlefish in China. Environmental Biology of Fishes 48: 241-255.

Welch AJ, et al. 2014. Polar Bears Exhibit Genome-wide signatures of bioenergetic adaptation to life in the Arctic environment. Genome Biology and Evolution 6: 433-450.

Wegner NC, Nicholas C., et al. Whole-body endothermy in a mesopelagic fish, the opah, Lampris guttatus. Science 348.6236 (2015): 786-789.

Wildman DE, Wu W, Goodman M, Grossman LI 2002. Episodic positive selection in ape cytochrome c oxidase subunit IV. Molecular Biology and Evolution 19: 1812-1815.

Wilkens LA, Hofmann MH, Wojtenek W 2002. The electric sense of the paddlefish: a passive system for the detection and capture of zooplankton prey. Journal of Physiology-Paris 96: 363-377.

Willett C, Burton R 2001. Viability of cytochrome c genotypes depends on cytoplasmic backgrounds in Tigriopus californicus. Evolution 55: 1592-1599. doi: 10.2307/2680377

Willett CS, Burton RS 2003. Environmental influences on epistatic interactions: viabilities of cytochrome c genotypes in interpopulation crosses. Evolution 57: 2286-2292.

Wittbrodt J, Shima A, Schartl M 2002. Medaka - a model organism from the far East. Nature Reviews Genetics 3: 53-64.

Wolfe K, Sharp P 1993. Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. Journal of Molecular Evolution 37: 441-456. doi: 10.1007/bf00178874

Woods IG, et al. 2000. A comparative map of the zebrafish genome. Genome Research 10: 1903-1914.

Woolley S, Johnson J, Smith MJ, Crandall KA, McClellan DA 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671-672.

Wright SD, Gray RD, Gardner RC 2003. Energy and the rate of evolution: inferences from plant rDNA substitution rates in the western Pacific. Evolution 57: 2893-2898.

Wu W, Goodman M, Lomax MI, Grossman LI 1997. Molecular evolution of cytochrome c oxidase subunit IV: evidence for positive selection in simian primates. Journal of Molecular Evolution 44: 477-491.

Yakes FM, Van Houten B 1997. Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress. Proceedings of the National Academy of Sciences USA 94: 514-519.

Yang Z. 2006. Computational Molecular Evolution: Oxford University Press Oxford.

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591.

Yang Z, Bielawski JP 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution 15: 496-503.

Yang Z, Nielsen R 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution 19: 908-917.

Yang Z, Nielsen R 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution 46: 409-418.

Yoshida M, Muneyuki E, Hisabori T 2001. ATP synthase - a marvellous rotary engine of the cell. Nature Reviews Molecular Cell Biology 2: 669-677.

Zeyl C, Andreson B, Weninck E 2005. Nuclear-mitochondrial epistasis for fitness in *Saccharomyces cerevisiae*. Evolution 59: 910-914. doi: 10.2307/3449037

Zhang and Broughton 2015. Heterogeneous natural selectin on oxidative phosphorylation genes among fishes with extremes of aerobic performance. BMC Evolutionary Biology (in review).

Zhang J, Nielsen R, Yang Z 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution 22: 2472-2479.

Zhang Z, et al. 1998. Electron transfer by domain movement in cytochrome bc1. Nature 392: 677-684.