UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INTEGRATING GENOMIC, MORPHOLOGICAL, AND PALEONTOLOGICAL DATA TO UNRAVEL MACROEVOLUTIONARY DRIVERS OF MORPHOLOGICAL DIVERSITY IN FISHES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

EMILY MARIE TROYER

Norman, Oklahoma

2024

INTEGRATING GENOMIC, MORPHOLOGICAL, AND PALEONTOLOGICAL DATA TO UNRAVEL MACROEVOLUTIONARY DRIVERS OF MORPHOLOGICAL DIVERSITY IN FISHES

A DISSERTATION APPROVED FOR THE

SCHOOL OF BIOLOGICAL SCIENCES

BY THE COMMITTEE CONSISTING OF

Dr. Hayley Lanier, Chair

Dr. Dahiana Arcila

Dr. Abigail Moore

Dr. Laura Stein

Dr. Richard Lupia

To the helpful folks at OSCER, especially Thang Ha and Horst Severini, for keeping the supercomputer up and running and troubleshooting the things that went kaput.

To my co-authors, for providing data, advice, new ideas, and feedback.

To my dissertation committee, for their helpful comments and feedback over the years.

To my lab-mates, past and present, both in Oklahoma and California, especially what remains of Lunch Bunch, for your support, commiseration, laughter, and random musings over the last five years.

To Hayley Lanier, for being the best stand-in advisor.

To Dahiana Arcila, for pushing me to become something beyond what I could have ever imagined the day I started my PhD.

To my parents and family, for not knowing exactly what it is I do, but supporting me anyways.

To my grandma, for all your love.

And most of all, to Isaac, for everything. Meep meep meep.

Table of Contents

List of Tables

Chapter 1

Appendix 1

Appendix 2

Appendix 3

List of Figures

Chapter 1

Chapter 2

Chapter 3

Appendix 1

Appendix 2

Appendix 3

Abstract

"*…from so simple a beginning, endless forms most beautiful and most wonderful have been, and are being, evolved*." — Charles Darwin, The Origin of Species.

Perhaps one of the most apparent, yet captivating, natural phenomena is the diversity of shape and form which has evolved on Earth. Morphological evolution is shaped by numerous factors across multiple biological scales, yet there is still much to be discovered using novel techniques and integrative approaches. By combining insights from phylogenetic comparative methods, paleoclimatic models, geometric morphometrics, and comparative transcriptomics, my dissertation provides a multi-faceted approach to understanding the factors (ecological, environmental, and genetic) contributing to the evolution of body size and shape in two diverse groups of fishes.

In Chapter 1, I use paleoclimate data in conjunction with a newly inferred phylogeny based on both extant and fossil species to examine how past ocean temperature is correlated with body size in tetraodontiform fishes (pufferfishes, boxfishes, ocean sunfishes, and allies). Numerous rules exist, which attempt to summarize the evolution of body size. These include Cope's rule, which states lineages tend to increase in size over evolutionary time scales, and Bergmann's rule which posits species tend to be larger in colder environments and smaller in warmer environments. These rules are generally well-supported in endotherms, but remain poorly understood in ectotherms. Using tetraodontiform fishes as a model clade, owing to their robust fossil record and disparity in body sizes, I find strong support for increasing body size over time in relation to decreasing oceanic temperatures. These results highlight the impact of paleoclimatic changes on aquatic ectotherms, which depend on their environment for temperature regulation and are potentially more susceptible to climatic changes compared with terrestrial vertebrates.

In Chapter 2, I continue investigating morphological evolution of tetraodontiform fishes, a clade that is extremely well-suited for these types of questions due to their extraordinarily unique morphological diversity, including spines and spikes in porcupinefishes, box-like armor in boxfishes, and inflatable bodies in pufferfishes. Here, I utilize data from three-dimensional CT scans of both fossil and extant species to investigate widescale drivers of morphological evolution in relation to habitat and key innovations. Habitat transitions and evolutionary innovations have been previously linked to increases in morphological diversification, but it is unclear whether these are universal drivers. Using tetraodontiform fishes as a model system, I show that these general rules may be more nuanced than previously thought. Coral reefs have long been suggested to increase morphological diversification in fishes, however I find that species living in other habitats display higher rates of skull shape evolution, suggesting reef association alone is not sufficient to spur high evolutionary rates. Additionally, I investigate a morphological novelty—the tetraodontiform beak—which is a fusion of the teeth into a beaked mouth in several families. I find that beaked families exhibit higher rates of morphological evolution compared with non-beaked families, suggesting that the beak may be an evolutionary innovation facilitating their diversification.

Lastly, in Chapter 3, I investigate morphological evolution through a genetic lens by employing comparative transcriptomics and differential expression analyses to identify candidate genes involved in miniaturization in gobiid fishes. While large body size is traditionally seen as advantageous, numerous transitions to miniaturization, the extreme reduction of adult body size, have evolved across the Tree of Life. Despite how common miniaturization appears, its genetic

mechanisms are poorly understood. Miniaturization is especially common among fishes, with species in the family Gobiidae (gobies) being an exceptional case. Gobiid fishes are part of an ecological grouping called “cryptobenthic reef fishes” which are the poster child for small-bodied fishes. Even within the already small-bodied gobiid phylogeny, there are multiple, independent transitions to extreme small size, allowing for tests of genetic convergence within a comparative macroevolutionary framework. Here, I assemble the first *de novo* transcriptomes for six species of gobiid fishes, which represent three clades each containing a closely related large-bodied and small-bodied species. I identify sets of statistically significant orthologs which are differentially expressed between large-bodied and small-bodied species in each clade. From these, I identify several candidate genes potentially involved in miniaturization, including *ybx1* and *bzw2*, both known to affect cell growth and development. These candidate genes offer insight into the genetic convergence on miniature body size and provide a framework for future studies.

Overall, my dissertation provides new insights into the large-scale processes and dynamics which have shaped the evolution of morphological diversity in fishes. By combining data from both fossil and extant taxa, as well as analyzing the evolution of size and shape through both a morphological and genomic lens, we can appreciate the complexities and nuances of morphological evolution and gain a more complete picture of the evolutionary processes which shape life on our planet.

The impact of paleoclimatic changes on body size evolution in marine fishes

Published in: *Proceedings of the National Academy of Sciences*, 11 July 2022

https://doi.org/10.1073/pnas.2122486119

Emily M. Troyer[1,2], Ricardo Betancur-R.[1], Lily C. Hughes[3], Mark Westneat[3], Giorgio Carnevale[4], William T. White[5], John J. Pogonoski[5], James C. Tyler[6], Carole C. Baldwin[7], Guillermo Ortí[8], Andrew Brinkworth[9], Julien Clavel[10] & Dahiana Arcila[1,2*]

[1] *Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Richards Hall Norman, OK 73019, USA.*
[2] *Department of Ichthyology, Sam Noble Oklahoma Museum of Natural History, 2401 Chautauqua Avenue, Norman, OK 73072, USA.*
[3] *University of Chicago, Department of Organismal Biology and Anatomy, and Committee on Evolutionary Biology, 1027 E. 57th St, Chicago IL, 60637, USA.*
[4] *Università degli Studi di Torino, Dipartimento di Scienze della Terra, Torino, Italy.*
[5] *CSIRO Australian National Fish Collection, National Research Collections Australia, Castray Esplanade, Hobart, Tasmania, 7001, Australia.*
[6] *Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, 10th & Constitution Ave. NW, Washington, DC 20560, USA.*
[7] *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th & Constitution Ave. NW, Washington, DC 20560, USA.*
[8] *The George Washington University, Department of Biological Sciences, Washington, D.C. 20052, USA.*
[9] *Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AZ, UK.*
[10] *Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France.*

Abstract

Body size is an important species trait, correlating with lifespan, fecundity, and other ecological factors. Over Earth's geological history, climate shifts have occurred, potentially shaping body size evolution in many clades. General rules attempting to summarize body size evolution include Bergmann's rule, which states that species grow to larger sizes in cooler environments and smaller sizes in warmer environments; and Cope's rule, which poses that lineages tend to increase in size over evolutionary time. Tetraodontiform fishes (including pufferfishes, boxfishes, and ocean sunfishes) provide an extraordinary clade to test these rules in ectotherms owing to their exemplary fossil record and the great disparity in body size observed among extant and fossil species. We examined Bergmann's and Cope's rules in this group by combining phylogenomic data (1,103 exon loci from 185 extant species) with 210 anatomical characters coded from both fossil and extant species. We aggregated data layers on paleoclimate and body size from the species examined, then inferred a set of time-calibrated phylogenies using tip-dating approaches for use in downstream comparative analyses of body size evolution using models that incorporate paleoclimatic information. We find strong support for a temperature-driven model in which increasing body size over time is correlated with decreasing oceanic temperatures. On average, extant tetraodontiforms are 2–3 times larger than their fossil counterparts, which otherwise evolved during periods of warmer ocean temperatures. These results provide strong support for both Bergmann's and Cope's rules, trends that are less studied in marine fishes compared to terrestrial vertebrates and marine invertebrates.

Significance Statement

General rules are useful tools for understanding how organisms evolve. Cope's rule (tendency to increase in size over evolutionary time) and Bergmann's rule (tendency to grow to larger sizes in cooler climates) both relate to body size, an important factor that affects the biology, ecology, and physiology of organisms. These rules are well studied in endotherms but remain poorly understood among ectotherms. Here, we show that paleoclimatic changes strongly shaped the trajectory of body size evolution in tetraodontiform fishes. Their body size evolution is explained by both Cope's and Bergmann's rules, highlighting the impact of paleoclimatic changes on aquatic organisms, which rely on their environment for temperature regulation and are likely more susceptible than terrestrial vertebrates to climatic changes.

Introduction

Paleoclimatic changes are recognized as strong factors affecting the macro-evolutionary dynamics of clades, including their distribution, ecology, and diversification (1). Throughout the course of Earth's geological history, several large, dynamic climatic shifts have occurred, such as the end-Permian extinction event (*ca.* 252 Ma), the Cretaceous–Paleogene (K-Pg) extinction event (*ca.* 66 Ma), and the Paleocene-Eocene Thermal Maximum (PETM; *ca.* 55.6 Ma) (2, 3). These periods are often marked by large changes in temperature, ocean acidification and anoxia, as well as increases in volcanic activity (3). These environmental shifts have led to mass extinction events in fishes (4) and changes in rates and magnitude of body size evolution in amphibians, birds, and mammals, (5, 6) among others. Morphological responses to paleoclimate change can be directly observed from

the fossil record. As body size correlates with many aspects of a species' biology, physiology, and ecology, its evolution should be associated with shifts in climate (5, 7, 8).

Bergmann's rule attempts to summarize body size responses to climatic changes, stating that species within a clade (or populations within a species) tend to grow to larger sizes in cooler environments and smaller sizes in warmer environments. While Bergmann's rule can apply at multiple evolutionary scales (9), from an interspecific viewpoint it can be defined as an ecogeographical trend where species' body size varies as a negative function of temperature. Originally studied in mammals, this trend has now been identified in a range of animals such as crustaceans, amphibians, and ray-finned fishes (8, 10–12). Various explanations have been proposed for Bergmann's rule, from heat conservation in endotherms to oxygen availability in ectotherms (13, 14). Another broad hypothesis summarizing body size patterns is Cope's rule, stating that species tend to increase in size over evolutionary time. Explanations for Cope's rule are thought to be linked to fitness advantages at larger body sizes or an increase in size variance as lineages diversify from a smaller ancestor following a passive trend (15, 16). Cope's rule could simply be an evolutionary or temporal manifestation of Bergmann's rule if lineages evolve larger body sizes during periods of climatic cooling (8). This idea, termed the Cope-Bergmann hypothesis by Hunt and Roy (8), has received considerably less attention than studies that examine body size trends relating to Cope's and Bergmann's rules separately (but see 8, 17).

Species' responses to climate will vary, but ectotherms that rely on their environment for temperature regulation are likely more susceptible than endotherms to climatic changes (18). Temperature controls a variety of aspects of ectotherm biology and is strongly linked with an organism's fitness, affecting growth rates and overall body size (14, 18). Understanding ectotherm morphological responses to global paleoclimate change may benefit greatly from examining clades with a rich fossil record. Ectothermic invertebrates have been studied in great detail, particularly brachiopods and marine arthropods (8, 19–21), due to their exceptional fossil record (22). Among ectothermic vertebrates, Cope's and Bergmann's rules have been tested in amphibians and reptiles (5, 23), but comparatively less in ray-finned fishes (but see 7, 24).

Fishes in the order Tetraodontiformes provide a model clade to test patterns of body size evolution in relation to paleoclimate events, owing to their exemplary and well-studied fossil record and extraordinary morphological diversity (4, 25). They constitute a circumglobally distributed taxonomic order of mostly marine, subtropical/tropical dwelling fishes, represented by *ca*. 450 living species, including the charismatic pufferfishes, triggerfishes, and ocean sunfishes. Tetraodontiforms exhibit a diverse array of body shapes, from nearly square (boxfishes) to globose (pufferfishes) and laterally compressed (filefishes). Species in this order also feature remarkable variation in adult body size, ranging from just 25 mm total length (TL) (e.g., *Rudarius excelsus, Carinotetraodon salivator*) to 3.4 m TL (e.g., *Mola mola, Masturus lanceolatus*). The tetraodontiform fossil record extends to the late Cretaceous with representatives from 12 exclusively fossil families; all 10 extant families are also present in the paleontological record, and on average body size is smaller among fossil taxa (25). Their morphological diversity, coupled with a robust fossil record, provides a unique system to test the Cope-Bergmann hypothesis in ectotherms.

This study aims to investigate patterns of body size evolution in relation to paleoclimate events in ectotherms using tetraodontiform fishes as a model group. We test the following questions: 1) are paleoclimatic changes correlated with changes in tetraodontiform body size? and 2) if a correlation between paleoclimate and body size is observed, do tetraodontiforms follow the Cope-Bergmann hypothesis? That is, does their body size evolution correlate with a

paleotemperature curve where tetraodontiforms are evolving towards a larger body size during periods of climate cooling? To address these questions, we estimate a new time-calibrated phylogeny for tetraodontiforms using total-evidence dating approaches that combine genome-wide data from extant species with a morphological matrix coded from both fossil and extant species. We also incorporate body length data and paleotemperature records spanning the past 100 Ma into a series of evolutionary model fitting analyses. We hypothesize that tetraodontiform evolution has been driven by past temperature changes and body size is strongly linked to past climate.

Materials and Methods

Taxonomic sampling and genomic data

Extended materials and methods are reported in the *SI Appendix, Supplementary Materials and Methods.* We generated new genomic data for 141 individuals representing 131 species of the order Tetraodontiformes and four species of its sister group, the Lophiiformes (33) (Dataset S1). All tissue samples are associated to voucher specimens deposited in museum collections (*SI Appendix:* Table S13). We shipped DNA extractions to Arbor Biosciences for library preparation and target enrichment. Sequencing of libraries was conducted at the University of Chicago Genomics Facility (Illumina HiSeq 4000). Target capture used the Eupercaria probe set of Hughes et al. (26, 69) to enrich 1,105 single-copy nuclear exon markers. We assessed sequence quality and removed two exons due to high levels of missing data, leaving 1,103 exons in total. We excluded one newly sequenced species (*Rhinecanthus verrucosus*) due to low capture efficiency. After quality control, we aligned all exons by considering their reading frames. We further increased our taxonomic sampling by adding sequences from 55 additional tetraodontiform species and one outgroup species retrieved from NCBI (Dataset S2). Our combined genomic dataset contains 185 tetraodontiforms and five outgroup taxa.

Phylogenomic inference

We inferred phylogenetic trees and associated support values in a maximum likelihood (ML) framework in IQTREE v.1.6.12 (70) (Dataset S4). In addition, we conducted a multi-species coalescent analysis in ASTRAL-III (71), based on IQTREE ML gene trees (*SI Appendix*, Fig. S1, Dataset S5). To account for the effect of missing data in our dataset, we conducted two concatenation-based ML analyses using all 1,103 exon markers: one including all newly sequenced taxa and 4 previously published transcriptomes (134 tetraodontiform species, 47% missing data overall; *SI Appendix*, Fig. S2, Dataset S6), and a second excluding taxa with more than 65% missing data (102 tetraodontiform species, < 33% missing data overall; *SI Appendix*, Fig. S3, Dataset S7). Because the topology and branch lengths were largely in agreement between the two analyses, all downstream phylogenetic analyses use the complete dataset.

Integration of fossil and extant species

To combine the fossils and extant tetraodontiform species, we used the morphological matrix of Arcila and Tyler (4), which consists of 210 characters coded for 17 extant and 52 fossil tetraodontiform species, plus two additional outgroup taxa. We combined the morphological matrix with our genomic dataset for a total of 237 tetraodontiform species and seven outgroups. Our analyses use the GTRGAMMA and Mk models with four partitions; three for the molecular sequences (one for each codon position); and one for the morphological dataset.

Phylogenetic uncertainty and total evidence dating using the fossilized birth death process

In addition to the phylogenomic analyses described above, we conducted divergence time estimations under a total evidence, or tip-dating, framework using the Fossilized Birth Death (FBD) model in MrBayes v 3.2.7a (72). To account for topological uncertainty, we assembled 15 largely independent genomic subsets containing *ca.* 50 randomly selected loci subsampled from the complete genomic dataset (1,103 loci total) (Dataset S8). All subsets overlap in only five "anchor" genes to maintain the same set of species for each subset (29, 30). In addition to genomic data, each subset contains the morphological dataset with fossil and extant taxa. We provide a list with the fossil ages in the Supplementary Information (Dataset S3) as well as a list of prior distributions used for node dating from previous studies that included Tetraodontiformes in their analyses (SI Appendix: Table S14). We ran all 15 subsets in MrBayes. After six months of total runtime, only five (of the 15) subsets reached convergence based on estimated sample size (ESS) values close to or above 200, and we retained those for downstream comparative analyses. Because there is no consensus yet on whether the superfamily Plectocretacicoidea should be considered as stem tetraodontiforms and the exclusion of this superfamily has the potential to drastically affect the age estimations (4, 31), we used two different schemes (including and excluding plectocretacicoids). We sampled 100 trees from the posterior distribution of each retained subset (500 trees total). In addition, we constructed a MCC tree from 10,000 trees evenly sampled from the posterior of all five subsets using TreeAnnotator (73).

Trait data

We compiled standard length (SL) data for most fossil and extant tetraodontiform species in our dataset using museum collection databases, public datasets, and published articles (Dataset S3). Given the highly fragmented nature of some fossil specimens, we excluded three extinct species from the body size analysis. Additionally, because of the bias for smaller specimens in museum collections, we omitted any measurements from individuals that were more than 20% smaller than the maximum recorded size and averaged the largest specimens to obtain a mean maximum SL per species. We performed all analyses using log transformed values. Mean maximum length was chosen as an indicator for how large a species could potentially reach. Although some tetraodontiform clades have a more three-dimensional body shape compared to other fish groups, phylogenetic generalized least squares (PGLS) analyses between SL and volume (SL vs. volume: $p = 0.0005$), and between SL and surface area (SA) (SL vs. SA: $p = 0.0248$), revealed a strong correlation (*SI Appendix,* Figs. S21-S22). These analyses were restricted to a subset of 41 tetraodontiform species (across all 10 extant families) for which CT scans were generated or obtained from Morphosource (Dataset S9). Measurements other than SL were not included due to homology concerns arising from fragmented fossil specimens.

Paleotemperature data

We obtained temperature curves that span the nearly 100 Ma evolutionary history of tetraodontiforms from Scotese et al (2). These authors used oxygen isotope data to reconstruct past global average ocean temperatures and sea surface temperatures between tropical latitudes (15°N–15°S). These two contrasting temperature curves most accurately capture the spectrum of past environmental affinities in this group (see Results).

Evolutionary model fitting

We conducted model fitting analyses in R version 4.0.2 (74). We fitted models of continuous character evolution using the R package 'geiger' and the 'fitContinuous' function. To more explicitly assess an increase in body size in response to paleoclimatic changes over time (Cope-Bergmann rule), we fitted a climate-dependent OU model, where the parameter of the model which represents the optimum trait value is time-variable and follows a temperature curve (75). We tested this model using the two temperature curves mentioned above from Scotese et al (2). Additionally, we tested the climate OU model on a deep-sea curve from Cramer et al. (35). See Results and Supplementary Materials and Methods for a description of other models tested. To account for tree uncertainty, we tested all models with the 500 trees selected from the posterior distribution. We account for interspecific sources of variation by incorporating measurement error into our model fitting analyses (see Supplementary Materials and Methods for more details).

Ecomorphological correlations

To further determine if tetraodontiforms adhere to the Cope-Bergmann rule, we examined patterns of body size in relation to past ocean temperature changes for the two temperature curves mentioned above. We performed ancestral state reconstructions of body sizes for all nodes and mapped these onto the MCC tree using the 'contMap' function in the R package 'phytools' (76). Resulting estimated ancestral sizes for each node were incorporated into a PGLS analysis and modeled for each of the two temperature curves from Scotese et. al. (2). We tested various evolutionary models for the PGLS analysis (e.g. OU model, BM model, and a non-phylogenetically informed model (i.e. ordinary least squares (OLS)) to determine best fit. (See Supplementary Materials and Methods for more details).

Results

Phylogenomic Inference, Total-Evidence Dating, and Tree Uncertainty

Under a total-evidence framework which combines a phylogenomic dataset based on 1,103 nuclear markers and 210 morphological characters (4), we used Bayesian inference and divergence time analyses to generate the most complete tetraodontiform phylogeny to date (Fig. 1*A*). Our approach used extensive taxon sampling which included both newly generated sequences (Dataset S1) and previously published sequences (Dataset S2) (26) for a total of 185 extant taxa (out of *ca*. 450) and 52 fossil taxa. In addition, we tested other phylogenetic inference methods such as concatenation-based maximum likelihood (ML) and summary multi-species coalescent approaches (*SI Appendix*, Figs. S1-S5). Phylogenetic placements are remarkably stable and congruent with past molecular studies for the group (27, 28) conducted before the advent of genomic datasets, supporting the monophyly of all families and the seven main suborders. Results show some differences between the major approaches conducted, with placements for some lineages (e.g., Triacanthidae + Triacanthodidae) along the backbone varying due to the short internodes at the base of the trees (*SI Appendix,* Figs. S1-S5).

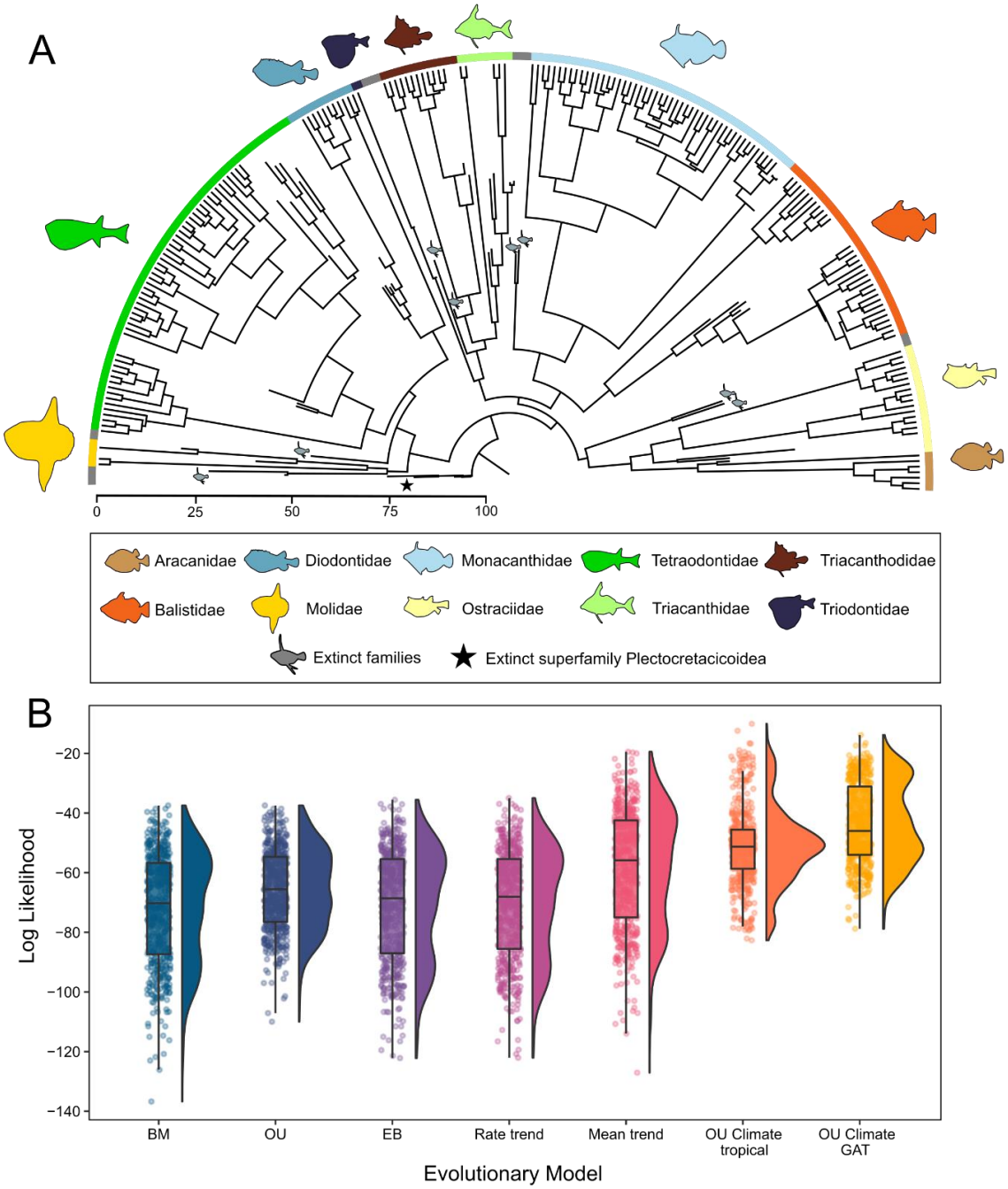


Figure 1. Tip-dating tree inferred for Tetraodontiformes and evolutionary model fitting results, including superfamily Plectocretacicoidea. *A*, Maximum clade credibility (MCC) tree derived from a total-evidence dating analysis using the fossilized birth-death model in MrBayes (including plectocretacicoids; see *SI Appendix*, Fig. S6*A* for a tree excluding plectocretacicoids). MCC tree is derived from 10,000 trees evenly sampled from the posterior distribution of five independent subsets. See *SI Appendix*, Fig. S4 for an expanded version of this tree. ***B*,** Raincloud plots (half-violin plots and boxplots) for each model of body size evolution tested, representing the distribution of likelihood scores from 500 trees evenly selected from the posterior distribution of five independent gene subsets in the Bayesian analysis (see also Table 1). Dots represent the raw likelihood score for each of the 500 trees analyzed, for each model. Evolutionary models include Early-burst (EB), Brownian motion (BM), rate trend, Ornstein-Uhlenbeck (OU), mean trend, a climate OU model using tropical ocean temperatures, and a climate OU model using global average ocean temperatures (GAT).

To account for phylogenetic and divergence time uncertainty on downstream comparative analyses, we analyzed five, independent and largely non-overlapping genomic subsets, each with a sufficient number of genes to reduce sampling error (29, 30). This approach contrasts with the traditional way of analyzing trees obtained from a Bayesian posterior distribution, which are typically estimated using a concatenated alignment with a scant number of genes. We also used two separate fossil schemes, by either including or excluding the superfamily Plectocretacicoidea, whose placement within Tetraodontiformes has been controversial (29, 31). Given the Mesozoic origin of plectocretacicoid fossils (70–96.9 Ma) their inclusion/exclusion in the dataset has important implications for tetraodontiform ages (29, 32–34). Our divergence time estimates including plectocretacicoids place crown Tetraodontiformes within the late Cretaceous (92.21 Ma, 95% highest posterior density [HPD]: 86.78 to 113.16 Ma) and the stem age at 98.62 Ma (95% HPD: 96.07 to 114.42 Ma; Figs. 1*A*, S4). Excluding the plectocretacicoids from the analysis shifts the age of crown tetraodontiforms forward to 62.45 Ma (95% HPD: 60.52–87.30; *SI Appendix*, Figs. S5, S6*A*), which is consistent with other estimations that exclude this extinct superfamily (32). We also assessed the sensitivity of divergence times estimates to root prior choice, finding no strong effects (Tables S1-S4; Figs. S7-S8).

Evolutionary Model Fitting

To test the Cope-Bergmann hypothesis, using tetraodontiform fishes as a model system, we fit a series of models of continuous trait evolution using mean maximum standard length as a proxy for body size (see *Materials and Methods*). To account for uncertainty in tree topology and divergence time estimates, each model was tested on 500 trees evenly selected from the posterior distribution of five independent subsets that reached convergence in the Bayesian Markov Chain Monte Carlo (MCMC) runs using the two fossil schemes (i.e., including and excluding Plectocretacicoidea). Models tested included the simple Brownian motion (BM; random walk) model, an early-burst (EB) model, an Ornstein-Uhlenbeck (OU) model, two variants of the trend model (mean trend and rate trend), and a pair of climate-dependent OU models, each fitted using a separate temperature curve as input. While the trend models tested are assumed to explicitly model Cope's rule, where time is the sole factor responsible for an increase in clade's body size, the OU climate model allows for tests of both Cope's and Bergmann's rules, where both time and temperature can influence the underlying trait (see *Materials and Methods*). Paleotemperatures curves used for the OU climate model include sea surface temperatures at tropical latitudes (15°N–15°S) and the global average temperature (GAT) for the past 100 Ma (2). Given the broad circumtropical/subtropical distributions and habitat preferences (mainly marine shallow-water dwellers) of tetraodontiforms, these two contrasting temperature curves most accurately capture the spectrum of past environmental affinities in this group. Nonetheless, we also tested an additional curve based on deep-sea temperature data (35), which yielded similar results (Fig. S9; Table S5).

Highest support went to the climate-driven model of evolution using a global average sea surface temperature curve (Fig. 1*B*), with an Akaike weight (AICw) of 0.999 (Table 1). All other models received substantially less support. These results are robust to the inclusion or exclusion of the superfamily Plectocretacicoidea (Fig. 1*B*, Table 1; *SI Appendix*, Fig. S6*B*, Table S6). The climate OU GAT model resulted in the best fitting model for 329/500 (65.8%) trees analyzed (Table S7). To visualize the scale and directionality of body size over time, we reconstructed and plotted ancestral body size as a traitgram. On average, extinct species and families are 2–3 times smaller in size compared to extant species and families (Fig. 2). When plectocretacicoids are excluded, similar patterns are observed (*SI Appendix*, Fig. S10).

Table 1. Evolutionary model results, including superfamily Plectocretacicoidea. Model fitting results for the seven macro-evolutionary models tested on 500 trees selected from the combined posterior distributions of five genomic subsets. Number of model parameters, mean values for the corrected Akaike information criterion (AICc), mean log likelihood (lnL), and weighted AIC (AICw) are reported. The strongest support went to the climate-driven OU model using the global average sea surface temperature curve.

Evolutionary Model	Parameters	AICc	lnL	AICw
OU climate GAT	5	98.25	-43.99	0.999
OU climate tropical	5	112.42	-51.08	8.35e-04
Mean trend	3	126.91	-60.40	3.60e-06
OU	3	137.88	-65.89	2.47e-09
Rate trend	3	146.78	-70.34	2.89e-11
BM	2	148.01	-71.98	1.56e-11
EB	3	148.03	-70.96	1.54e-11

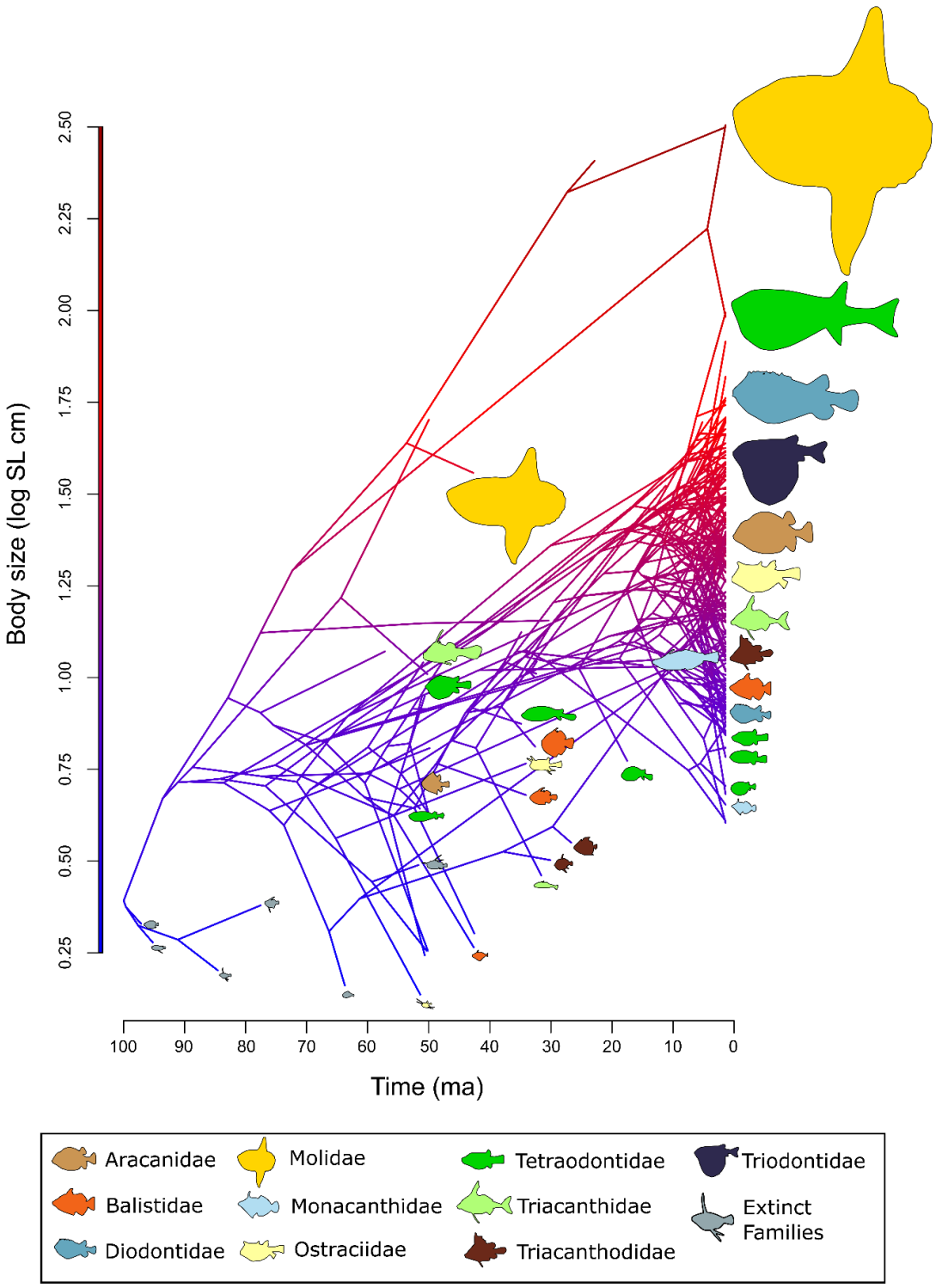


Figure 2. Evolution of tetraodontiform body size over time, including superfamily Plectocretacicoidea. Ancestral reconstruction of body size in tetraodontiforms, as estimated using the 'contMap' function in the R package 'phytools' (76). The log-transformed mean maximum standard length for each species is plotted as a traitgram on the Y axis, with time on the X axis. Fish silhouettes are scaled to represent proportional log body size and colored by family, with extinct families in grey. The estimated ancestral body size of tetraodontiforms is 2-3 times smaller than the mean of present-day taxa.

In addition, we examined the potential effect of tree age on model selection in our subsets inclusive of Plectocretacicoidea. Our 500 trees range in age from 96.25 to 130.86 Ma, with an average root age of 103.18 Ma (Fig. S11). It appears that slightly younger trees are favored by the Mean Trend model (mean age = 99.44 Ma; Fig. S12) compared with the two climate OU models (GAT mean age = 103.47 Ma, Tropical mean age = 104.49 Ma). However, this could be an effect of the fewer number of trees being favored by the Mean Trend model overall (n=64), compared with the GAT (n=329) and Tropical (n=107) curve climate OU models. While model selection on a small sample of trees can produce biased estimates, using a large number of trees obtained from the posterior distributions of independently assembled gene subsets provides a powerful approach to account for tree uncertainty in macroevolutionary inferences (e.g. 29, 30).

Because there is a global trend towards declining temperatures over the evolutionary history of Tetraodontiformes (i.e. from the late Cretaceous to present day), it can be difficult to decouple the effects of temperature (Bergmann's rule) from other processes that may be correlated with increased body sizes (Cope's rule), as the patterns generated by either rule can be indistinguishable from one another. To further assess the role of the overall trend, we decomposed the temperature curve into two distinct components: the smoothed overall trend and the fluctuations around this trend (Fig. S13). To identify the model parameters with the strongest weights, we ran three separate analyses on both temperature curves (GAT and tropical latitudes). The first analysis modeled the two independent parameters (overall trend + fluctuations) together, the second modeled only the overall trend, and the third modeled only the fluctuations. If the model with only the overall trend shows the best fit for the data, this would imply that the trend is more important compared with the temperature fluctuations (i.e., some evidence for Cope's rule, but inconclusive for Bergmann's). In our analyses using the GAT curve, we find that most support went to the overall trend + fluctuations model (AICw=0.769), followed by overall trend (AICw=0.152) and fluctuations (AICw=0.077; Table S8, Fig S14 *A*) models. Although we observe a different pattern in the tropical latitude curve, with most support going to the fluctuations model (AICw=0.852 vs. AICw=0.147 for trend + fluctuations model, and AICw=2.77e-06 for trend model; Table S9, Fig. S14 *B*), given that the original model fitting results indicate stronger support for a GAT curve over a tropical latitude curve (Table 1), we place more weight onto the decomposed GAT curve model results. All in all, our decomposed model analyses suggest that both the overall trend and the fluctuations around this trend are important to the OU climate model fit, providing support for the Cope-Bergmann rule (8) as an explanation of body size evolution in tetraodontiforms.

Ecomorphological Correlations

To test whether past ocean temperatures are correlated with tetraodontiform body size, we performed phylogenetic generalized least squares (PGLS) regression analyses under a best-fit model between reconstructed ancestral node body sizes and paleo-ocean temperatures at the age of each node using the maximum clade credibility (MCC) tree (see details in *SI Appendix*: Supplemental Materials and Methods). While phylogenetically informed statistical tests (such as PGLS) are traditionally used to compare two continuous species traits, we argue that for the purposes of our analyses, ocean temperature can be categorized as a species trait, following the approach of Garland et al. (36), where environmental traits can be used as long as these traits can be passed on from ancestor to descendent species. While many studies use a temperature curve based on deep-sea data for these types of analyses (6, 37), we account for temperature variation relating to the actual habitat and ecology of tetraodontiforms by testing two temperature curves

based on sea surface temperatures, as described above (2). From the late Cretaceous (~100 Ma) to the present day, sea surface temperatures (both at tropical latitudes and the global average) have been gradually decreasing, a trend that correlates negatively with tetraodontiform body size (Fig. 3). These PGLS regressions under a best fit OU model are statistically significant for the global average temperature (p = 5.571e-03), but not the tropical latitude temperature (p = 0.0653) curves (*SI Appendix*, Fig. S15; Table S10). These results reflect the model fitting analyses, where the GAT curve received substantially more model support than the tropical temperature curve (Table 1). Even when the superfamily Plectocretacicoidea is excluded from the analysis, shifting tetraodontiform divergence time estimations forward ~30 Ma, the PGLS analysis remains statistically significant for the GAT curve (p =4.84e-05, but not the tropical curve: p = 0.447; *SI Appendix*, Figs. S16, S17; Table S11). Examining trends among only fossil species reveals a similar pattern to the analyses where fossil and extant species are combined (*SI Appendix*, Fig. S18). Fossils-only analyses show that body size is strongly correlated with ocean temperatures for the global average curve (p = 0.0311), but not the tropical curve (p =0.1804; *SI Appendix*, Fig. S19; Table S12).

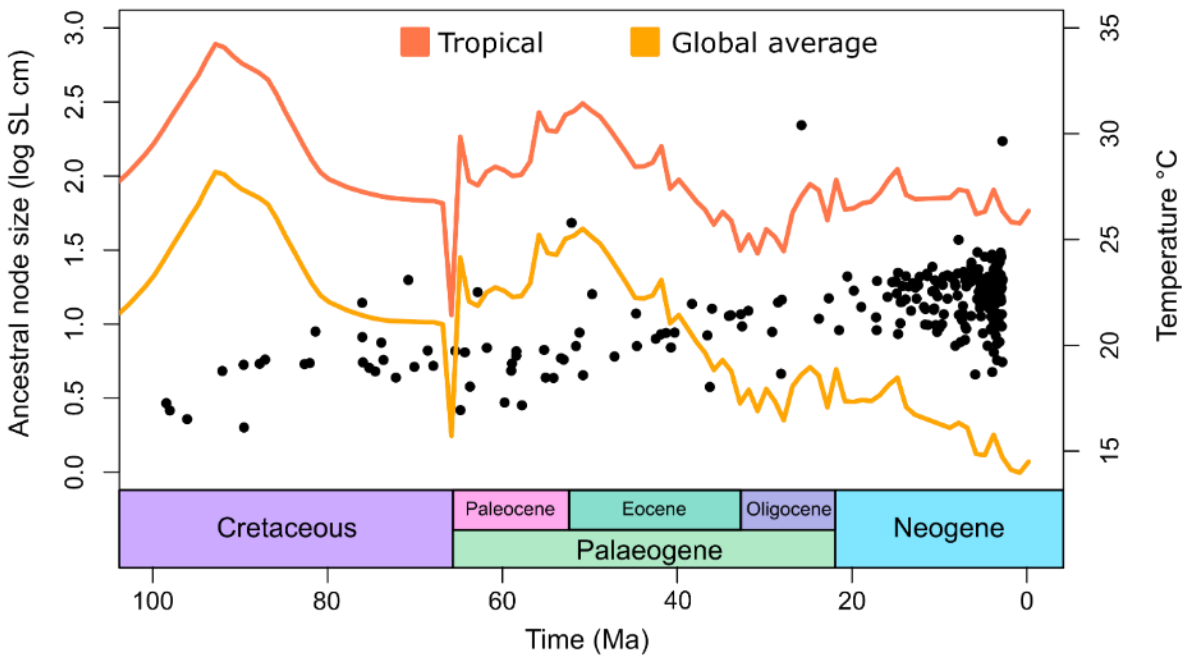


Figure 3. Tetraodontiform body size and temperature over time. Sea surface temperature for tropical latitudes (15°N–15°S; orange line) and a global average sea temperature (yellow line) are plotted for the past 100 Ma. The reconstructed ancestral node body size (log mean maximum standard length in cm) for tetraodontiforms is also plotted against time. Sea surface temperatures have been slowly cooling since the late Cretaceous, while tetraodontiform body size has gradually increased. See *SI Appendix*, Fig. S20 for a version of this plot colored by family.

Discussion

By integrating morphological and genome-wide sequence data in a total-evidence dating framework, we infer the most complete phylogeny for both living and extinct tetraodontiform fishes. Using this robust phylogenetic framework, which accounts for uncertainty in topology and divergence time estimates, we conduct a suite of comparative approaches to test for Cope's and Bergmann's rules. Our model-fitting analyses that incorporate paleoclimatic information provide strong support for the Cope-Bergmann rule (8), where tetraodontiform body size is strongly correlated with sea surface temperatures over the past 100 Ma. Furthermore, by decomposing the temperature curve into an overall trend (Cope's rule), the fluctuations around that trend (Bergmann's rule) and a combination of both (Cope-Bergmann rule) (8) we find strong support for the latter. Our results align with previous studies that examine paleoclimatic effects on species evolutionary trends, providing support for a correlation between temperature and body size. Studies of birds and mammals have found support for both rules (13, 38, 39). However, among ectotherms, findings have been mixed, with studies identifying support for Cope's rule in arthropods (19, 40) and reptiles (41); support for Bergmann's rule in fishes (10, 42), arthropods (8, 11, 19), reptiles (43), and amphibians (5, 12); an inverse or no support for Cope's rule in arthropods (44) and reptiles (23, 45); and an inverse or no support for Bergmann's rule in arthropods (46), reptiles (43), and amphibians (47, 48). Bergmann's rule, originally hypothesized for endothermic vertebrates (49–51), describes how larger endothermic species might conserve heat better in higher latitudes with cooler temperatures due to an increased surface area to volume ratio. The heat conservation hypothesis behind this rule would likely not apply to ectotherms at all (52, 53; but see 54), thus other explanations or hypotheses for this pattern are necessary (see below).

Studies examining paleoclimatic effects on evolutionary trends benefit from a robust fossil record. However, variables such as habitat composition, sampling effort, and specimen morphology can influence fossil preservation. Hard-bodied organisms inhabiting shallow marine environments are more likely to fossilize compared to soft-bodied, deep-sea species (55), and there are additional taphonomic biases related to body size. Among fishes, larger fossils are more likely to become disarticulated and then scattered by various hydrological processes, resulting in a lower probability of discovery (56). In contrast, smaller specimens have fragile bones, and thus a higher potential of being destroyed compared to larger specimens, thus they may not be as common in the fossil record (57, 58). All analyses that incorporate fossil data will have taphonomic biases, but a goal should be to minimize potential biases whenever possible. Tetraodontiform fishes are well represented in the fossil record, likely owing to their hardened external anatomy and habitat preferences for shallow marine waters. Their fossil record is rich, with extinct representatives in all 10 living families, as well as 12 exclusively fossil families (4). Among the tetraodontiform fossils in our dataset, most (51 specimens, representing 14 species in 10 families) come from the Monte Bolca Lagerstätten (50.5–48.5 Ma). Within this Eocene locality, tetraodontiform fossils are exceptionally preserved (59) and body sizes range from 8 mm SL (Eolactoria sorbinii†) to 521 mm SL (Protobalistum imperialis†) (Dataset S3). Given the large range in sizes that are found in Monte Bolca and their overall completeness, it is unlikely that preservation and size-related taphonomic biases have major effects on our analyses.

Cope's rule is often explained as a selective advantage towards a larger body size. Despite the fact that overall body size increased over time, we find that support for a mean trend model (i.e., Cope's rule), while higher than that of the OU, rate trend, EB, and BM models, was not the

strongest, implying that a selective advantage to large body sizes over time alone does not provide enough explanatory power. Indeed, we find support for a Cope-Bergmann rule (8) in our climate OU models where body size increases over time as sea surface temperatures gradually cool. Cope's rule has previously been thought to be a manifestation of Bergmann's rule during periods of climate cooling (Cope-Bergmann rule) (8), and any trend in increased body size during these cooler periods would just be seen as coincidental with Cope's rule, thus the two rules might be combined into a single, Cope-Bergmann rule. Contrary to one of the three predictions of the Cope-Bergmann rule postulated by Hunt and Roy (8), which suggests that increased body sizes are attributable to evolutionary changes within lineages/species, the evolutionary trajectories observed in tetraodontiforms may be indicative of lineage turnover resulting from species selection. While unlike Hunt and Roy (8) we lack a high-resolution temporal dataset to assess intraspecific evolution, a clear general pattern emerges from our dataset: as smaller and older tetraodontiforms species become extinct they are replaced by larger and younger ones. Ultimately, these patterns may be indicative of Bergmann's rule operating at interspecific taxonomic scales (9, 60).

In other marine ectotherms, temperature is a driving factor for increased body size. Hunt et al. (19) examined trends in 19 species of deep-sea ostracods, and found body size increased in 84% of examined species from the Eocene (40 Ma) to the present, during which global deep-sea temperatures gradually cooled. They also identify strongest support for a temperature-tracking model (a simple model where expected change in body size corresponds to change in temperature), finding significant body size increases only during periods of sustained cooling. Taken together, these results highlight the importance of incorporating environmental variables into macro-evolutionary studies that examine trends over time.

It is unlikely that a global explanation for body size evolution exists for all ectotherms in response to paleoclimate change and by extension, Bergmann's rule. However, temperature seems to be an important driver of their evolutionary patterns, as ectothermic species have a reduced capacity for heat conservation compared to endotherms. An increase in body size in colder environments may be related to the temperature-size rule (TSR), which states that ectotherms reared in colder temperatures in an experimental setting tend to grow more slowly and mature with a larger body size relative to ectotherms reared under warm temperatures (61). Paleotemperature changes over large timescales are correlated with changes in ectotherm body size (8, 37). Additionally, because many environmental variables are temperature-dependent, temperature may play an indirect role in determining ectotherm body size, with additional confounding factors coming into play. Amongst aquatic ectotherms, oxygen may be a more limiting factor for body size evolution. Reduced dissolved oxygen in warmer waters is limiting for ectotherms dependent on aquatic respiration. This temperature-dependent oxygen limitation has been proposed to relate to Bergmann's rule in aquatic ectotherms (14) and these reductions in body size are greater in aquatic taxa compared with terrestrial taxa (62). Other studies have linked warming-induced anoxia as a driver of decreased body size in early Jurassic marine invertebrates (20). Temperature-dependent oxygen limitation may also explain why past studies of terrestrial ectotherms either find an inverse Bergmann's rule (47, 48) or a non-existent trend (43, 46), meaning that increased body size in relation to temperature cooling may be much stronger in aquatic-respiring ectotherms (14).

Past climatic changes strongly shaped the trajectory of phenotypic evolution across many clades. On a global scale, large paleoclimate changes are associated with extinction events, such as those that occurred during the end-Permian (ca. 252 Ma) and Cretaceous-Paleogene (ca. 66 Ma) events (63). These events can be size-selective and are often thought to favor small-bodied taxa (64, 65). In general, small-bodied species are presumed to be at a lower risk of extinction,

potentially owing to their faster generation times and increased fecundity (24). But they can also be at increased risk of extinction due to factors such as geographic range, which is often smaller compared to that of their large-bodied counterparts (65). Although extinction events favoring small-bodied taxa have been documented (24, 65, 66), the opposite pattern has been observed as well, implying that this may simply be a clade-specific effect (65). When extinction risk is examined at higher taxonomic levels, these events tend to favor large-bodied taxa (64) suggesting that additional biological factors play a role in extinction risk (e.g. lesser predation of large-bodied tetraodontiforms due to increased toxicity of flesh (67)) and scale is an important factor to consider. Among ectotherms, those in marine environments may be most affected by global extinction events related to climate warming compared to terrestrial habitats (68). This was the case for the largest mass extinction event, the end-Permian, where temperature-induced hypoxia drove a majority of marine species to extinction (63). Examining these past patterns of ectotherm evolution may provide insights into how species will respond during the next chapter of global climate change.

In conclusion, we find a strong link between the evolution of body size in tetraodontiforms and past climate and paleotemperatures. Gradual climate cooling over the past 100 Ma (especially pronounced during the past 50 Ma) was associated with increases in average tetraodontiform body length. Our results are robust to a number of factors driving uncertainty in macroevolutionary inferences, including the use of different genomic subsets and root priors for time tree inference, the inclusion or exclusion of the controversial plectocretacicoid fossils, the utilization of different paleotemperature curves, and the implementation of alternative comparative approaches. While the evolution of body size in tetraodontiforms appears to conform to the Cope-Bergmann rule, other factors (e.g., ocean acidification, dissolved oxygen concentrations) could affect this trait and thus deserve further investigation.

Acknowledgements

We are thankful to L. Smith and A. Bentley (University of Kansas) for providing tissue samples. A. Graham (CSIRO Australian National Fish Collection, Hobart), C. Huddleston, and D. Pitassy (Smithsonian Institution National Museum of Natural History) assisted with preparing and shipping samples. R. Peterson and V. Rodriguez (The George Washington University) helped by conducting DNA extractions. We also thank G. Soreghan (University of Oklahoma) and C. Lear (Cardiff University) for their advice on paleoclimate datasets and G. Slater (University of Chicago) for advice on modeling. The computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma. OSCER Director H. Neeman and OSCER Senior System Administrators J. Speckman and H. Severini provided valuable technical expertise.

Author Contributions

E.M.T., R.B.R., and D.A. planned and oversaw the project; L.C.H. and D.A. carried out data assembly and alignment of the genomic datasets; E.M.T. analyzed data; G.C. and J.C.T. curated the ages and fish fossil species included in divergence time estimation analyses; D.A., R.B.R., M.W., W.T.W., J.J.P., G.O., and C.C.B. collected, identified, and curated the fish materials examined; E.M.T., J.C., and A.B. carried out analyses associated with the OU climate model; E.M.T., D.A., and R.B.R. wrote the paper and all other authors contributed to the writing.

Data Availability Statement

Raw sequence reads are available at National Center for Biotechnology Information Sequence Read Archive BioProject (number PRJNA767646). All other supplementary data and code have been deposited in Dryad (https://doi.org/10.5061/dryad.z34tmpgfw). All other study data are included in the article and supporting information.

Funding

This project was supported by National Science Foundation (NSF) grants to D.A (DEB-2015404, DEB-2144325, and DBI- 2131464), R.B.R. (DEB-1932759 and DEB-1929248), G.O. (DEB-1457426 and DEB-1541554), and C.B. (DEB-1541552).

References

1. J.-C. Svenning, W. L. Eiserhardt, S. Normand, A. Ordonez, B. Sandel, The Influence of Paleoclimate on Present-Day Patterns in Biodiversity and Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* **46**, 551–572 (2015).
2. C. R. Scotese, H. Song, B. J. W. Mills, D. G. van der Meer, Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. *Earth-Science Reviews* **215**, 103503 (2021).
3. J. Chen, Y. Xu, Establishing the link between Permian volcanism and biodiversity changes: Insights from geochemical proxies. *Gondwana Research* **75**, 68–96 (2019).
4. D. Arcila, J. C. Tyler, Mass extinction in tetraodontiform fishes linked to the Palaeocene–Eocene thermal maximum. *Proc. R. Soc. B.* **284**, 20171771 (2017).
5. A. Martínez-Monzón, H.-A. Blain, G. Cuenca-Bescós, M. Á. Rodríguez, Climate and amphibian body size: a new perspective gained from the fossil record. *Ecography* **41**, 1307–1318 (2018).
6. J. Clavel, H. Morlon, Accelerated body size evolution during cold climatic periods in the Cenozoic. *Proc Natl Acad Sci USA* **114**, 4183–4188 (2017).
7. M. Hardman, L. M. Hardman, The Relative Importance of Body Size and Paleoclimatic Change as Explanatory Variables Influencing Lineage Diversification Rate: An Evolutionary Analysis of Bullhead Catfishes (Siluriformes: Ictaluridae). *Systematic Biology* **57**, 116–130 (2008).
8. G. Hunt, K. Roy, Climate change, body size evolution, and Cope's Rule in deep-sea ostracodes. *Proceedings of the National Academy of Sciences* **103**, 1347–1352 (2006).
9. S. Meiri, Bergmann's Rule – what's in a name? *Global Ecology and Biogeography* **20**, 203–207 (2011).
10. F. Fernández-Torres, P. A. Martínez, M. Á. Olalla-Tárraga, Shallow water ray-finned marine fishes follow Bergmann's rule. *Basic and Applied Ecology* **33**, 99–110 (2018).
11. D. S. Johnson, *et al.*, The fiddler crab, Minuca pugnax, follows Bergmann's rule. *Ecology and Evolution* **9**, 14489–14497 (2019).

12. T. L. Yu, D. L. Wang, M. Busam, Y. H. Deng, Altitudinal variation in body size in Bufo minshanicus supports Bergmann's rule. *Evol Ecol* **33**, 449–460 (2019).
13. V. Salewski, C. Watt, Bergmann's rule: a biophysiological rule examined in birds. *Oikos* **126** (2017).
14. N. Rollinson, L. Rowe, Temperature-dependent oxygen limitation and the rise of Bergmann's rule in species with aquatic respiration. *Evolution* **72**, 977–988 (2018).
15. S. M. Stanley, An Explanation for Cope's Rule. *Evolution* **27**, 1–26 (1973).
16. D. W. McShea, Mechanisms of Large-Scale Evolutionary Trends. *Evolution* **48**, 1747–1763 (1994).
17. L. H. Liow, P. D. Taylor, Cope's Rule in a modular organism: Directional evolution without an overarching macroevolutionary trend. *Evolution* **73**, 1863–1872 (2019).
18. J. Verheyen, R. Stoks, Temperature variation makes an ectotherm more sensitive to global warming unless thermal evolution occurs. *Journal of Animal Ecology* **88**, 624–636 (2019).
19. G. Hunt, S. A. Wicaksono, J. E. Brown, K. G. Macleod, Climate-driven body-size trends in the ostracod fauna of the deep Indian Ocean. *Palaeontology* **53**, 1255–1268 (2010).
20. V. Piazza, C. V. Ullmann, M. Aberhan, Temperature-related body size change of marine benthic macroinvertebrates across the Early Toarcian Anoxic Event. *Sci Rep* **10**, 4675 (2020).
21. D. Jablonski, Body-size evolution in Cretaceous molluscs and the status of Cope's rule. *Nature* **385**, 250–252 (1997).
22. M. Foote, J. J. Sepkoski, Absolute measures of the completeness of the fossil record. *Nature* **398**, 415–417 (1999).
23. P. L. Godoy, R. B. J. Benson, M. Bronzati, R. J. Butler, The multi-peak adaptive landscape of crocodylomorph body size evolution. *BMC Evol Biol* **19**, 167 (2019).
24. L. Sallan, A. K. Galimberti, Body-size reduction in vertebrates following the end-Devonian mass extinction. *Science* **350**, 812–815 (2015).
25. F. Santini, J. C. Tyler, A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to Recent. *Zoological Journal of the Linnean Society* **139**, 565–617 (2003).
26. L. C. Hughes, *et al.*, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci USA* **115**, 6249–6254 (2018).
27. D. Arcila, R. Alexander Pyron, J. C. Tyler, G. Ortí, R. Betancur-R., An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution* **82**, 131–145 (2015).
28. F. Santini, L. Sorenson, M. E. Alfaro, A new phylogeny of tetraodontiform fishes (Tetraodontiformes, Acanthomorpha) based on 22 loci. *Molecular Phylogenetics and Evolution* **69**, 177–187 (2013).
29. M. Rincon-Sandoval, *et al.*, Evolutionary determinism and convergence associated with water-column transitions in marine fishes. *Proc Natl Acad Sci USA* **117**, 33396–33403 (2020).
30. A. Santaquiteria, *et al.*, Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences. *Systematic Biology*, syab028 (2021).
31. M. E. Alfaro, *et al.*, Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nat Ecol Evol* **2**, 688–696 (2018).

32. T. J. Near, *et al.*, Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences* **110**, 12738–12743 (2013).
33. R. Betancur-R., *et al.*, The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr* **5** (2013).
34. M. Matschiner, A. Böhne, F. Ronco, W. Salzburger, The genomic timeline of cichlid fish diversification across continents. *Nat Commun* **11**, 5895 (2020).
35. B. S. Cramer, K. G. Miller, P. J. Barrett, J. D. Wright, Late Cretaceous–Neogene trends in deep ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal geochemistry (δ^{18} O and Mg/Ca) with sea level history. *J. Geophys. Res.* **116**, C12023 (2011).
36. T. Garland Jr., P. H. Harvey, A. R. Ives, Procedures for the Analysis of Comparative Data Using Phylogenetically Independent Contrasts. *Systematic Biology* **41**, 18–32 (1992).
37. J. A. Velasco, F. Villalobos, J. A. F. Diniz-Filho, S. Poe, O. Flores-Villela, Macroecology and macroevolution of body size in *Anolis* lizards. *Ecography* **43**, 812–822 (2020).
38. E. J. Torres-Romero, I. Morales-Castilla, M. Á. Olalla-Tárraga, Bergmann's rule in the oceans? Temperature strongly correlates with global interspecific patterns of body size in marine mammals. *Global Ecology and Biogeography* **25**, 1206–1215 (2016).
39. F. Bokma, *et al.*, Testing for Depéret's Rule (Body Size Increase) in Mammals using Combined Extinct and Extant Data. *Syst Biol* **65**, 98–108 (2016).
40. N. A. Heim, M. L. Knope, E. K. Schaal, S. C. Wang, J. L. Payne, Cope's rule in the evolution of marine animals. *Science* **347**, 867–870 (2015).
41. R. B. J. Benson, R. A. Frigot, A. Goswami, B. Andres, R. J. Butler, Competition and constraint drove Cope's rule in the evolution of giant flying reptiles. *Nat Commun* **5**, 3567 (2014).
42. R. A. Saunders, G. A. Tarling, Southern Ocean Mesopelagic Fish Comply with Bergmann's Rule. *The American Naturalist* **191**, 343–351 (2018).
43. K. G. Ashton, C. R. Feldman, Bergmann's rule in nonavian reptiles: turtles follow it, lizards and snakes reverse it. *Evolution* **57**, 1151–1163 (2003).
44. J. T. Waller, E. I. Svensson, Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution* **71**, 2178–2193 (2017).
45. D. S. Moen, Cope's rule in cryptodiran turtles: do the body sizes of extant species reflect a trend of phyletic size increase? *J Evolution Biol* **19**, 1210–1221 (2006).
46. S. Pallarés, M. Lai, P. Abellán, I. Ribera, D. Sánchez-Fernández, An interspecific test of Bergmann's rule reveals inconsistent body size patterns across several lineages of water beetles (Coleoptera: Dytiscidae). *Ecological Entomology* **44**, 249–254 (2019).
47. Q. Liu, *et al.*, Latitudinal variation in body size in Fejervarya limnocharis supports the inverse of Bergmann's rule. *Animal Biology* **68**, 113–128 (2018).
48. T. L. Yu, *et al.*, Altitudinal body size variation in Rana kukunoris: the effects of age and growth rate on the plateau brown frog from the eastern Tibetan Plateau. *Ethology Ecology & Evolution* **0**, 1–13 (2021).
49. B. Rensch, Histological Changes Correlated with Evolutionary Changes of Body Size. *Evolution* **2**, 218–230 (1948).
50. E. D. Cope, *The Primary Factors of Organic Evolution* (Open Court Publishing Company, 1904).
51. C. G. L. C. Bergmann, *Ueber die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grosse* ... (Vandenhoeck und Ruprecht, 1848).

52. C. Watt, S. Mitchell, V. Salewski, Bergmann's rule; a concept cluster? *Oikos* **119**, 89–100 (2010).
53. J. Santilli, N. Rollinson, Toward a general explanation for latitudinal clines in body size among chelonians. *Biological Journal of the Linnean Society* **124**, 381–393 (2018).
54. F. J. Zamora-Camacho, S. Reguera, G. Moreno-Rueda, Bergmann's Rule rules body size in an ectotherm: heat conservation in a lizard along a 2200-metre elevational gradient. *J Evol Biol* **27**, 2820–2828 (2014).
55. R. B. J. Benson, R. Butler, R. A. Close, E. Saupe, D. L. Rabosky, Biodiversity across space and time in the fossil record. *Current Biology* **31**, R1225–R1236 (2021).
56. J. S. Albert, D. M. Johnson, J. H. Knouft, Fossils provide better estimates of ancestral body size than do extant taxa in fishes. *Acta Zoologica* **90**, 357–384 (2009).
57. E. S. Stinnesbeck, J. Rust, F. Herder, Paleobiology and taphonomy of the pycnodont fish Nursallia gutturosum, based on material from the Latest-Cenomanian-middle Turonian Vallecillo platy limestone, Mexico. *PalZ* **93**, 659–668 (2019).
58. G. R. Smith, R. F. Stearley, C. E. Badgley, Taphonomic bias in fish diversity from cenozoic floodplain environments. *Palaeogeography, Palaeoclimatology, Palaeoecology* **63**, 263–273 (1988).
59. M. Friedman, G. Carnevale, The Bolca Lagerstätten: shallow marine life in the Eocene. *Journal of the Geological Society* **175**, 569–579 (2018).
60. T. M. Blackburn, K. J. Gaston, N. Loder, Geographic gradients in body size: a clarification of Bergmann's rule. *Diversity and Distributions* **5**, 165–174 (1999).
61. D. Atkinson, R. M. Sibly, Why are organisms usually bigger in colder environments? Making sense of a life history puzzle. *Trends in Ecology & Evolution* **12**, 235–239 (1997).
62. J. Forster, A. G. Hirst, D. Atkinson, Warming-induced reductions in body size are greater in aquatic than terrestrial species. *Proceedings of the National Academy of Sciences* **109**, 19310–19314 (2012).
63. D. P. G. Bond, S. E. Grasby, On the causes of mass extinctions. *Palaeogeography, Palaeoclimatology, Palaeoecology* **478**, 3–29 (2017).
64. J. L. Payne, N. A. Heim, Body size, sampling completeness, and extinction risk in the marine fossil record. *Paleobiology* **46**, 23–40 (2020).
65. W. J. Ripple, *et al.*, Extinction risk is most acute for the world's largest and smallest vertebrates. *PNAS* **114**, 10678–10683 (2017).
66. J. S. Berv, D. J. Field, Genomic Signature of an Avian Lilliput Effect across the K-Pg Extinction. *Systematic Biology* **67**, 1–13 (2018).
67. E. L. A. Malpezzi, J. De Freitas, F. T. Rantin, Occurrence of toxins, other than paralysing type, in the skin of tetraodontiformes fish. *Toxicon* **35**, 57–65 (1997).
68. M. L. Pinsky, A. M. Eikeset, D. J. McCauley, J. L. Payne, J. M. Sunday, Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature* **569**, 108–111 (2019).
69. L. C. Hughes, *et al.*, Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Molecular Ecology Resources* **21**, 816–833 (2021).
70. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).
71. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

72. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
73. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).
74. R Core Team, *R: A language and environment for statistical computing* (2020).
75. A. Brinkworth, A. Goswami, J. Clavel, A New Ornstein-Uhlenbeck process for modeling the evolution of traits with time-variable optima. *In prep*.
76. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).

Evolutionary innovation accelerates morphological diversification in pufferfishes and their relatives

Emily M. Troyer[1], Kory M. Evans[2], Christopher Goatley[3-5], Matt Friedman[6], Giorgio Carnevale[7], Benjamin Nicholas[8], Matthew Kolmann[9], Katherine E. Bemis[10,11], Dahiana Arcila[12*]

[1]*Department of Biology and Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, 730 Van Vleet Oval, Richards Hall, Norman, OK 73019, USA*
[2]*Rice University, Biosciences Department, 6100 Main Street, Houston, TX 77005, USA*
[3]*School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton, Hampshire SO14 3ZH, UK*
[4]*Australian Museum Research Institute, Australian Museum, 1 William Street, Sydney, NSW 2010, Australia*
[5]*Function, Evolution and Anatomy Research (FEAR) Lab, School of Environmental and Rural Science, University of New England, Armidale, NSW 235, Australia*
[6]*Museum of Paleontology and Department of Earth and Environmental Sciences, University of Michigan, 1105 North University Avenue, Ann Arbor, Michigan 48109, USA*
[7]*Dipartimento di Scienze della Terra, Università degli Studi di Torino, Via Valperga Caluso 35, 10125 Turin, Italy*
[8]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA*
[9]*Department of Biology, University of Louisville, Louisville, KY 40292, USA*
[10]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, USA*
[11]*National Systematics Laboratory, Office of Science and Technology, NOAA Fisheries, Washington, D.C. 20560, USA*
[12]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA*

Abstract

Evolutionary innovations have played an important role in shaping the diversity of life that we see today. However, the way in which these innovations arise and their downstream effects on patterns of morphological diversification remain poorly understood. Here, we examine the effects of evolutionary innovation on trait diversification using tetraodontiform fishes (pufferfishes, boxfishes, ocean sunfishes, and allies) as a model system owing to their range of habitats and divergent morphologies, such as the fusion of the teeth into a beak in several families. Using three-dimensional geometric morphometric data for 176 fossil and extant species, we examine the effect of phenotypic integration and habitat on the evolution of innovative traits. Strong phenotypic integration is thought to be a requirement for rapid evolution of highly divergent structures, like the tetraodontiform beak. Our results show that this innovation arose in the presence of highly conserved patterns of integration across the skull, suggesting that integration did not limit the range of available phenotypes to tetraodontiforms. Furthermore, we find that irrespective of habitat, beaks have allowed tetraodontiforms to diversify into novel ecological niches and general rules pertaining to evolutionary innovation may be more nuanced than previously thought.

Introduction

One of the most fascinating natural phenomena is the ability of organisms to evolve and diversify. Innovations have shaped the process of evolution and can account for the evolutionary success of many clades (Heard and Hauser 1995). Innovations are novel traits which directly enhance diversification by allowing species access to previously unattainable ecological niches and resources (Miller 1949; Heard and Hauser 1995). These adaptive breakthroughs can be found throughout the Tree of Life and include such innovations as flight in birds (Mayr 1963; Heard and Hauser 1995), adhesive toepads in tree-dwelling lizards (Miller and Stroud 2022), and pharyngeal jaws in cichlids and other fishes (Liem 1973). However, the process by which these innovations arise and their downstream effects on patterns of morphological diversification remain poorly understood. Thus, comprehending which factors contribute to evolutionary innovation is critical to our understanding of the generation of biological diversity.

Phenotypic modularity, the concept that certain traits evolve semi-independently of each other, is thought to facilitate evolutionary innovation (Yang 2001; Jablonski 2022). Highly modular traits will exhibit a strong degree of covariation within individual structural regions (modules) but much lower covariation between modules, allowing organisms to evolve semi-independent adaptations and promote the evolution of complexity and diversity (Wagner and Altenberg 1996; Goswami et al. 2014; Evans et al. 2017). Alternatively, more recent attention has been given to phenotypic integration, a potential facilitator of innovation. More integrated traits covary strongly with each other and evolve together in a coordinated fashion (Olson and Miller 1999). Although modular traits can occupy a larger region of trait space (Goswami et al. 2014), highly integrated traits occur in more constrained trait space and along a specific direction. In this way, highly integrated traits can explore more of the outer edges of trait space, producing more extreme morphologies such as the asymmetrically flattened skulls of flatfishes and the avian beak (Navalón et al. 2020; Evans et al. 2021). Current studies on phenotypic integration and modularity mainly focus on extant taxa and, when fossils are involved, tend to concentrate on short timescales. Thus, more comprehensive analyses covering broader taxonomic scope and deeper time scales are needed to fully understand these patterns.

Innovations may also enable access to novel habitats, which can further shape evolution by providing ecological release from competition and predation, resulting in increased lineage and morphological diversification rates (Yoder et al. 2010). Certain habitats may promote more morphological diversification than others. Coral reefs are highly diverse ecosystems containing 25% of all marine life despite occupying less than 0.2% of the ocean floor (Souter et al. 2021). Reefs host complex interactions between coral and fish species, facilitating fine-scale niche partitioning and associated morphological and functional specialization, leading to increased ecological opportunities (Alfaro et al. 2007; Cowman and Bellwood 2011; Price et al. 2011; Brandl et al. 2018; Evans et al. 2019*b*). Additionally, many reef-associated species bear novel innovations, such as the intramandibular jaw joint of butterflyfishes (Chaetodontidae), bristletooth surgeonfishes (Acanthuridae), rabbitfishes (Siganidae), angelfishes (Pomacanthidae), and combtooth blennies (Blenniidae) which expands the gape, assisting in grazing algae and invertebrates from reef surfaces (Gibb et al. 2015), and beak-like jaws of parrotfishes (Labridae; Scarini) which facilitates durophagous feeding modes (Price et al. 2010; Evans et al. 2023*a*).

Tetraodontiform fishes represent an excellent system to explore mechanisms related to evolutionary innovation. These fishes possess a host of unique features, including the cuboidal body of boxfishes, erectable body spines in porcupinefishes, and the fusion of the teeth into a beak in several families, including the most speciose family, the pufferfishes (Tetraodontidae) (Tyler 1980; Santini and Tyler 2003). This novel dentition develops from elongated tooth bands which continuously fuse together during ontogeny to form a mineralized beak (Fraser et al. 2012; Thiery et al. 2017). Beaks are a relatively rare morphology in fishes, and most macroevolutionary studies of non-avian beaks focus on parrotfishes (Price et al. 2010; Evans et al. 2023*a*). The tetraodontiform beak represents an extreme phenotypic trait, and because of this, has potentially evolved in a highly integrated fashion, similar to the avian beak. In addition to displaying novel phenotypes, tetraodontiforms permit tests for links between morphological diversity and habitat association. These fishes are circumglobally distributed and occupy a wide variety of ecosystems, from coral reefs to open oceans and even freshwater rivers (Tyler 1980). Lastly, their fossil record extends to the Late Cretaceous with many well-preserved specimens, allowing for evolutionary tests over deep time scales (Santini and Tyler 2003).

Here, we investigate evolutionary mechanisms of innovation and reveal general patterns of morphological diversification in tetraodontiform fishes. Past work shows that reef association promotes lineage diversification in tetraodontiforms (Alfaro et al. 2007); however, the impact of habitat on patterns of phenotypic diversification remain unclear. At the same time, the evolution of the beak in many tetraodontiform families could represent a significant innovation facilitating evolution in beak-bearing clades. Tetraodontiform beaks represent a divergent morphology which may promote rapid evolution due to strong phenotypic integration of the skull. Here, we use a three-dimensional (3D) geometric morphometrical approach to quantify tetraodontiform skull morphology using micro-computed tomography (CT) scans of 173 extant and three extinct species, representing some of the only known three-dimensionally preserved fossil fish skulls. We use phylogenetic comparative methods to analyze rates of skull shape evolution and disparity over time and quantify patterns of modularity and integration between beaked and non-beaked species. We hypothesize that high skull integration in beaked species and coral reef association will promote faster rates of morphological evolution and increases in morphological disparity, contributing to the overall evolution of the clade.

Materials and Methods

Taxonomic sampling and CT-scan data acquisition

We analyzed the skull shape of 176 species of Tetraodontiformes, including 173 extant and three fossil species. This sampling encompasses all ten living families, with fossil representatives from Tetraodontidae and Triodontidae. A comprehensive list of the scanned species, scanning locations, and specimen voucher information is provided in Appendix 1 (Supplementary Material). We included three of the only known catalogued three-dimensional fossil tetraodontiform skulls, *Sphoeroides hyperostosus*† (Tetraodontidae), *Triodon antiquus*† (Triodontidae) and *Ctenoplectus williamsi*† (stem Triodontidae). Each species was represented by a single adult specimen that underwent micro-CT scanning either at the University of Washington Friday Harbor Laboratories (Bruker Skyscan 1173; 40 species), Rice University (Bruker Skyscan 1273; 92 species), the University of New England, Australia (General Electric phoenix v|tome|x s; 13 species), Cornell University (General Electric 120 micro-CT; 1 species), and the University of Michigan (Nikon XT H 225 ST; 1 fossil species), totaling 147 new scans. Previously scanned fossil specimens were also acquired from Close et al. (2016) and an unpublished scan of *Triodon antiquus*†, both scanned on a Nikon XT H 225 ST at the Natural History Museum, London. Finally, we also sourced scans for 27 additional species from MorphoSource (http://morphosource.org).

Segmentation, digitization, reproducibility, and fossil landmarks

Scans were segmented in 3D Slicer (Kikinis et al. 2014) to isolate the skull bones from the rest of the body. Within 3D Slicer, digitization of the specimens was conducted using a landmark scheme of 170 points (48 fixed landmarks and 122 sliding semilandmarks) as detailed in Figure S1 and Table S1. This scheme represents an extended version of the general Eupercaria scheme from Evans et al. (2023*b*), ensuring a comprehensive representation of the diversity of tetraodontiform skull shapes. To ensure consistency of landmark placement, all landmarking was conducted by the same person (EMT). After landmarking all 176 specimens, each scan was inspected again for verification, with slight adjustments made when necessary. A reproducibility test was performed for quality control by re-landmarking two randomly selected specimens. The original landmarks were then juxtaposed with the re-landmarks by plotting both sets into a phylomorphospace (Figure S2).

All landmarks were placed on the left side of the skull. However, for one fossil specimen, *Triodon antiquus*†, the left side could not be landmarked due to taphonomic degradation. To address this, the CT scan was converted to a three-dimensional mesh and then inverted for landmarking using the *MeshLab* software (Cignoni et al. 2008). Additional taphonomic processes affecting our fossil specimens occasionally rendered some landmarks unplaceable. Instead of proceeding with a significantly reduced subset of landmarks shared across all extant and fossil specimens, we chose to estimate the missing landmarks for the fossil specimens using the 'MissingGeoMorph' function in the R package *LOST* (Arbour and Brown 2020) for this purpose. We applied a Bayesian principal component analysis (BPCA) method to estimate missing landmark data, leveraging principal component regressions and Bayesian estimations to determine the position of absent landmarks (Oba et al. 2003). Empirical data set analyses by Arbour and Brown (2014) have shown this method to be highly reliable for landmark estimations. Moreover, these types of estimates produce a better fit to the original data than exclusion of specimens with incomplete landmarks (Arbour and Brown 2014).

Skull shape analyses

After digitization, we imported the landmark coordinates into *R* version 4.2.3 (R Core Team 2023) using a custom script from Buser et al. (2023) and processed them with the *geomorph* package version 4.0.5 (Baken et al. 2021). To remove the effect of non-shape variation, such as scale, size, and orientation across specimens, we performed a generalized Procrustes superimposition between specimens (Rohlf and Slice 1990). Semilandmarks were slid along their tangent directions using the Procrustes distance criterion, because sliding using bending energy may cause spurious correlations among landmarks which can bias modularity analyses (Zelditch and Swiderski 2023). Given the biomechanical complexity of fish skulls, which contain multiple moving elements, analyzing shape can be challenging due to preservation artefacts affecting jaw positions (Evans et al. 2019*a*). We account for these biases in rotation and translation of these mobile structures by performing a local superimposition, ensuring standardized positioning of the different skull elements (Rhoda et al. 2021*a*, 2021*b*). Following local superimposition, we then conducted a principal components analysis (PCA) to assess the main axes of skull shape variation. The first two principal components (PC1 and PC2) were visualized as a phylomorphospace using the pruned, time-calibrated phylogeny of Troyer et. al (2022). Additionally, we employed the 'plotRefToTarget' function in *geomorph* to plot the primary and secondary axes of skull shape variation as ball and stick plots (Figure S3).

Phylogenetic estimation, trait coding, and ancestral trait reconstructions

To investigate skull evolution across Tetraodontiformes, we used the time-calibrated phylogeny proposed by Troyer et al. (2022), which encompasses 239 extant and fossil species of Tetraodontiformes. Using the *ape* package (Paradis and Schliep 2019), we pruned the tree to retain only the 176 taxa for which morphological data was available. Habitat associations for extant species were obtained from FishBase (Froese and Pauly 2023) and Fishes of Australia (Bray and Gomon 2023). Each species was categorized as being coral reef-associated or non-reef-associated (Appendix 1). For fossil species, categorization was based on the paleobiotope where they were discovered, with reef-association being determined by the presence of hermatypic corals (Marramà et al. 2016; Friedman and Carnevale 2018). Dental morphology for each species was also characterized. Species were defined as beaked if they possessed highly modified and fused teeth, a characterization based on the criteria of Tyler (1980) (Appendix 1). The beaked group consists of all species from the families Molidae, Diodontidae, Tetraodontidae, and Triodontidae. Non-beaked species include those from the families Balistidae, Monacanthidae, Triacanthidae Triacanthodidae, Ostraciidae, and Aracanidae. Non-beaked species possess teeth that are discrete units and protrude out of the jaw sockets, while beaked species possess teeth that do not protrude and are incorporated into the matrix of the jaw bones. For the fossil species in our analysis, their classification as beaked or non-beaked was based on characters 68-70 from the morphological matrix by Santini and Tyler (2003) defining beaked species as possessing teeth fused into a parrot-like beak and non-beaked species having teeth as discrete units, either slender caniniform, stoutly conical, incisiform-molariform, or thick caniniform teeth.

For each trait of interest in our analyses (habitat association, mouth type), we reconstructed ancestral states in *phytools* (Revell 2012) using stochastic character mapping (Huelsenbeck et al. 2003) under an equal rates model with the 'make.simmap' function on the complete tree from Troyer et al. (2022), containing seven outgroup taxa, 52 fossil tetraodontiforms, and 187 extant

tetraodontiforms. Empirical Bayesian posterior probabilities for estimated ancestral states were plotted for each node of the phylogeny.

Rates of skull shape evolution and morphological disparity

We quantified the rate of skull shape evolution between reef and non-reef-associated species, as well as between beaked and non-beaked species using the *'*compare.evol.rates' function in *geomorph.* Significance was assessed using the phylogenetic simulation approach run with 1,000 iterations (Adams and Collyer 2018). Similarly, to compare skull morphological disparity, we employed the 'morphol.disparity' function in *geomorph* for both reef vs. non-reef and beaked vs. non-beaked. This function calculates morphological disparity by estimating the Procrustes variance for each group using the residuals of a linear model fit. Additionally, we used the 'compare.multi.evol.rates' function in *geomorph* to evaluate rates of bone module evolution between beaked and non-beaked species. This method calculates evolutionary rate parameters of multivariate traits (σ^2) from predefined modules. Significance is assessed by comparing the observed rate to a null rate matrix derived from a random simulation using 1,000 permutations.

To quantify rates of tetraodontiform skull shape evolution across the pruned phylogeny, we used BayesTraitsV4 (Pagel and Meade 2022). To reduce dimensionality, we employed only the first 24 principal components, which account for 95% of the total shape variation. PCs were multiplied by 1,000, following Evans et al. (2023*b*), because BayesTraits has computational issues with very small numbers. We account for evolutionary correlations in trait variation using the "TestCorrel" function in BayesTraits, which sets trait correlation to zero. We used a reversible-jump Markov chain Monte Carlo method with uninformative priors and ran two independent chains for 200 million generations, sampling every 10,000 generations, with the first 60 million discarded as burn-in. Convergence was visually assessed using Tracer v1.7.1 (Rambaut et al. 2018), with all ESS (Effective Sample Size) values exceeding 200. We tested two models of trait evolution: a single rate Brownian motion model that assumes one rate across the phylogeny, and a variable rates model that allows for changes in rates throughout the phylogeny and identifies where rates differ (Venditti et al. 2011). Model fitting was performed by calculating Bayes factors based on the marginal likelihoods from both models. A Bayes factor greater than 10 is regarded as strong support for that particular model (Pagel and Meade 2022).

To assess changes in subclade morphological disparity for Tetraodontiformes, we implemented disparity-through-time (DTT) analyses under a Brownian motion model using the *'geiger'* package. We also compare DTT for reef vs. non-reef associated species as well as beaked vs non-beaked species. We used the average squared Euclidean distance among all pairs of data points as our disparity index. The DTT method calculates changes in relative subclade disparity through time across nodes in the phylogeny. We compared the observed disparity to that under a simulated null Brownian motion model iterated over 1,000 generations. We used the observed and simulated disparities to calculate a morphological disparity index (MDI), which measures the deviation from expectations for relative within-clade disparities under a model of Brownian motion. A negative MDI indicates that disparity is distributed among subclades, and is commonly interpreted as evidence for an early burst, characteristic of an adaptive radiation (Harmon et al. 2010; Slater and Pennell 2014). A positive MDI indicates that disparity is distributed within subclades. Because MDI estimations at multiple time points suffer from a high false-positive rate, we use the two-tailed rank envelope method of Murrell (2018) to assess significance. This method provides an overall p-value as well as a p-interval because the ranks will almost always result in some ties. Because this is a two-tailed test, p-values below 0.025 are considered significant.

Evolutionary modularity and integration

To test for patterns of evolutionary modularity between beaked and non-beaked species, we defined eight *a priori* hypotheses of modularity that encompass a range of functional, embryological, and sensory hypotheses from previous literature (Westneat 2005; Helfman et al. 2009; Kague et al. 2012; Evans et al. 2019*a*), as well as an eleven-module individual bone hypothesis (Figure S4). We evaluated modularity using the 'phylo.modularity' function in the *geomorph* package. This function uses the covariance ratio (CR) method, which is a measure of the relative strength of covariation between modules compared to the strength within modules (Adams 2016). A CR less than 1 indicates a more modular system, while a CR greater than 1 indicates less modularity. Then, under the best-supported hypothesis, we compared the effect sizes (strength of the modular signal) for beaked and non-beaked species using the 'compare.CR' function in *geomorph*. The best-supported model is indicated by the lowest effect size. Additionally, we evaluated our eight hypotheses of modularity with 'phyloEMMLi' (Goswami and Finarelli 2016), which applies maximum likelihood to compare different modularity hypotheses while also accounting for phylogenetic non-independence. We visualized the results of the modularity analyses by creating network plots showing the magnitude of integration between each module.

Using our best-supported modular hypothesis, we tested evolutionary integration among modules using the 'phylo.integration' function in *geomorph*. This method uses partial least squares (PLS) analysis to quantify the degree of covariation between our hypothesized modules (Rohlf and Corti 2000). PLS values closer to 1 indicate higher integration. Because this method can be sensitive to sample size between groups (Adams and Collyer 2016), we first randomly removed 42 non-beaked species until both groups contained 67 species. Finally, we compared effect sizes between groups using 'compare.PLS' in the *geomorph* package. All scripts and data produced for this study can be found on the Dryad repository (doi:10.5061/dryad.4f4qrfjjx).

Results

Skull shape evolution in Tetraodontiformes

Tetraodontiformes display a wide diversity of skull shapes (Figure 1 A). The primary axis of shape variation (principal component 1, PC1) accounts for 37% of the overall variance and overwhelmingly corresponds to mouth type. Along PC1, beaked species occupy a separate area of the morphospace from their non-beaked counterparts (Figure 1 A). In addition to possessing an elongated distal margin of the first tooth, which corresponds to a beak, skulls of beaked species are characterized by an anterior orbit, a long supraoccipital crest, and a wider preopercle (Figure S3). Additionally, we find that beaked species exhibit significantly greater skull morphological disparity than non-beaked species ($p = 0.002$). By contrast, reef and non-reef-associated species do not occupy different areas of the morphospace, and display no significant difference in skull disparity ($p = 0.476$; Figure S5 A).

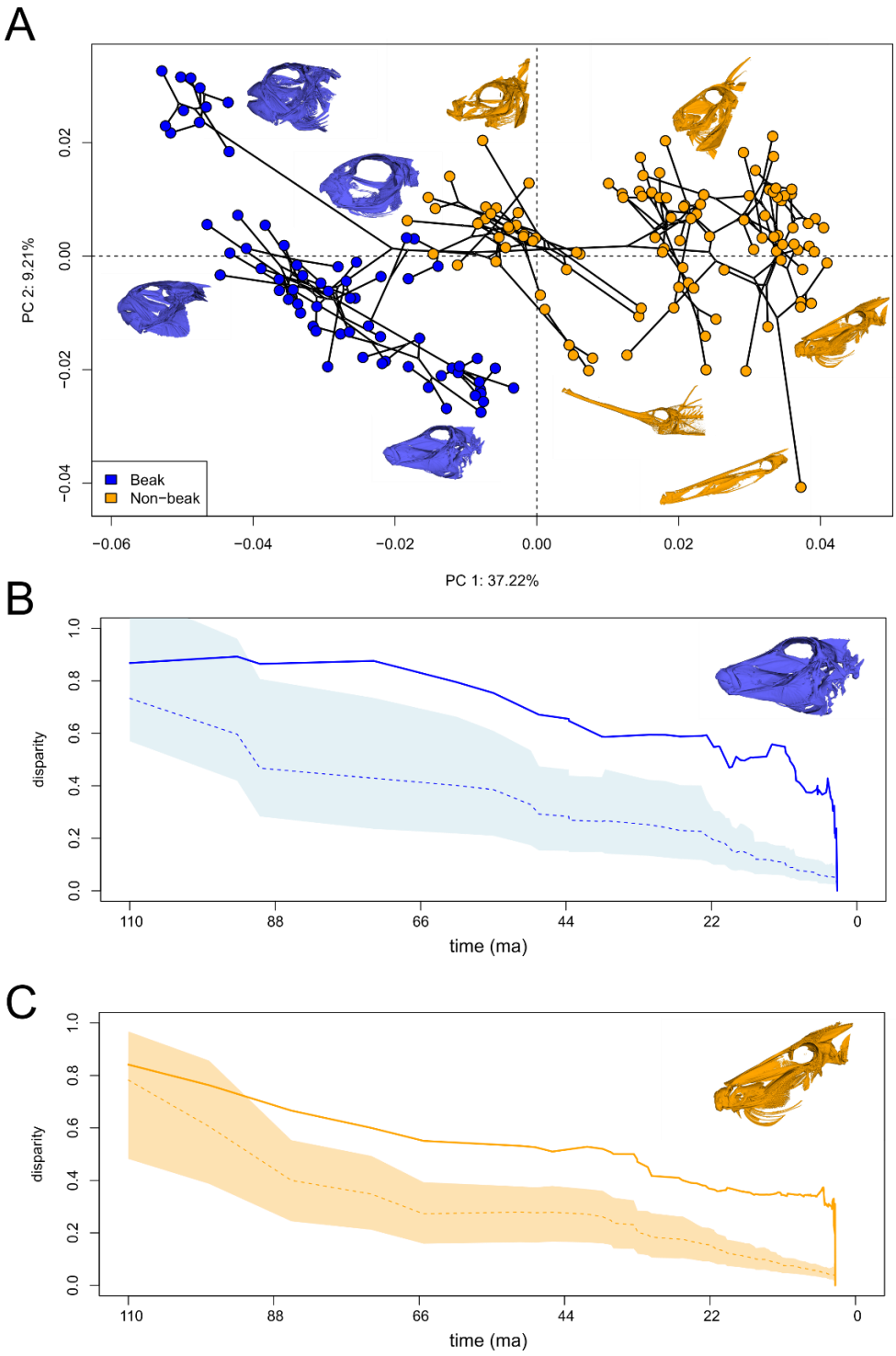


Figure 1. Skull shape disparity and evolution across 176 tetraodontiform species. A) Phylomorphospace analysis of skull shape in tetraodontiform fishes. Beaked species are indicated in blue while non-beaked species are in orange. Insets depict CT scans of tetraodontiform skulls showing a wide degree of morphological disparity. B) Disparity through time plots of beaked and C) non-beaked tetraodontiform species showing disparity rate over scaled time. Dashed lines indicate the Brownian motion expectation and shaded regions represent the 95% confidence interval. Solid lines indicate the actual measured disparity.

Disparity-through-time (DTT) analyses suggest that skull shape disparity for Tetraodontiformes is principally distributed within subclades based on a low, yet positive morphological disparity index (MDI = 0.179, rank envelope test: $p = 0.007$, p-interval = 0.0009-0.0139; Figure S6). When comparing DTT between beaked and non-beaked species, both sets of species deviate significantly from the null Brownian distribution beginning around 90 million years ago (Ma), and continue to exceed Brownian expected disparity until present day (Figure 1 B-C). However, beaked species exhibit higher disparity over this time interval (MDI = 0.35; rank envelope test: $p = 0.007$; p-interval = 0.0009-0.0139) compared with non-beaked species (MDI = 0.22; rank envelope test: $p = 0.005$; p-interval = 0.0009-0.0109). Additionally, beaked species display a sharp upturn in diversity from 20-10 Ma. Reef-associated and non-reef-associated species also exhibit differences in disparity-through-time (Figure S5 B-C). Reef species display a lower MDI (MDI = 0.131; rank envelope test: $p = 0.007$; p-interval = 0.0009-0.0149) compared with overall Tetraodontiformes, while non-reef species display a slightly higher MDI (MDI = 0.181; rank envelope test: $p = 0.009$; p-interval = 0.0009-0.0189).

Tempo of skull evolution

In examining overall rates of morphological evolution, we find significant differences in the tempo of beaked and non-beaked tetraodontiforms. Beaked species display significantly higher rates of skull shape evolution ($p = 0.001$; rate ratio = 1.42). When comparing by habitat, interestingly, non-reef-associated species show an evolutionary pace almost twice as fast as their reef-associated counterparts ($p = 0.001$; rate ratio = 1.71).

On a branch-specific level, tetraodontiform fishes exhibit variable rates of skull shape evolution. Our BayesTraits analysis yielded strong support for a variable rates model of trait evolution over a single rate Brownian motion model (Table S2). The most significant rate increases occur on the two branches leading to beaked families: the clade of Molidae, Tetraodontidae, Diodontidae; and the clade Triodontidae (Figure 2). Within the pufferfishes (Tetraodontidae), we find a distinct rate increase leading to the sharpnose pufferfishes (*Canthigaster* spp.), coinciding with a shift towards more elongated skulls compared to other pufferfish species. Beyond the beaked species, notable rate increases are evident in the spikefishes (*Halimochirurgus* spp.) and the filefish *Anacanthus barbatus*, both of which exhibit remarkable snout elongations.

To determine the timing of evolutionary transitions to coral reef and non-reef habitats as well as the evolution of the beak, we performed ancestral character reconstructions for each trait using the time-calibrated phylogeny from Troyer et al. (2022). We recovered the ancestral state of all Tetraodontiformes as non-reef-associated, with multiple transitions to reef habitats occurring over their evolutionary history (Figure S7). Additionally, we identify the ancestral mouth state as non-beaked with two independent transitions to a beak (Figure S8). Remarkably, the oldest tetraodontiform fossils exhibiting beak-like structures are estimated to be approximately 50 Ma. However, phylogenetic analyses reveal long stems for these groups extending back over 80 Ma, suggesting that the evolutionary origins of beak development in this group may precede the current fossil record (Figure S7, S8).

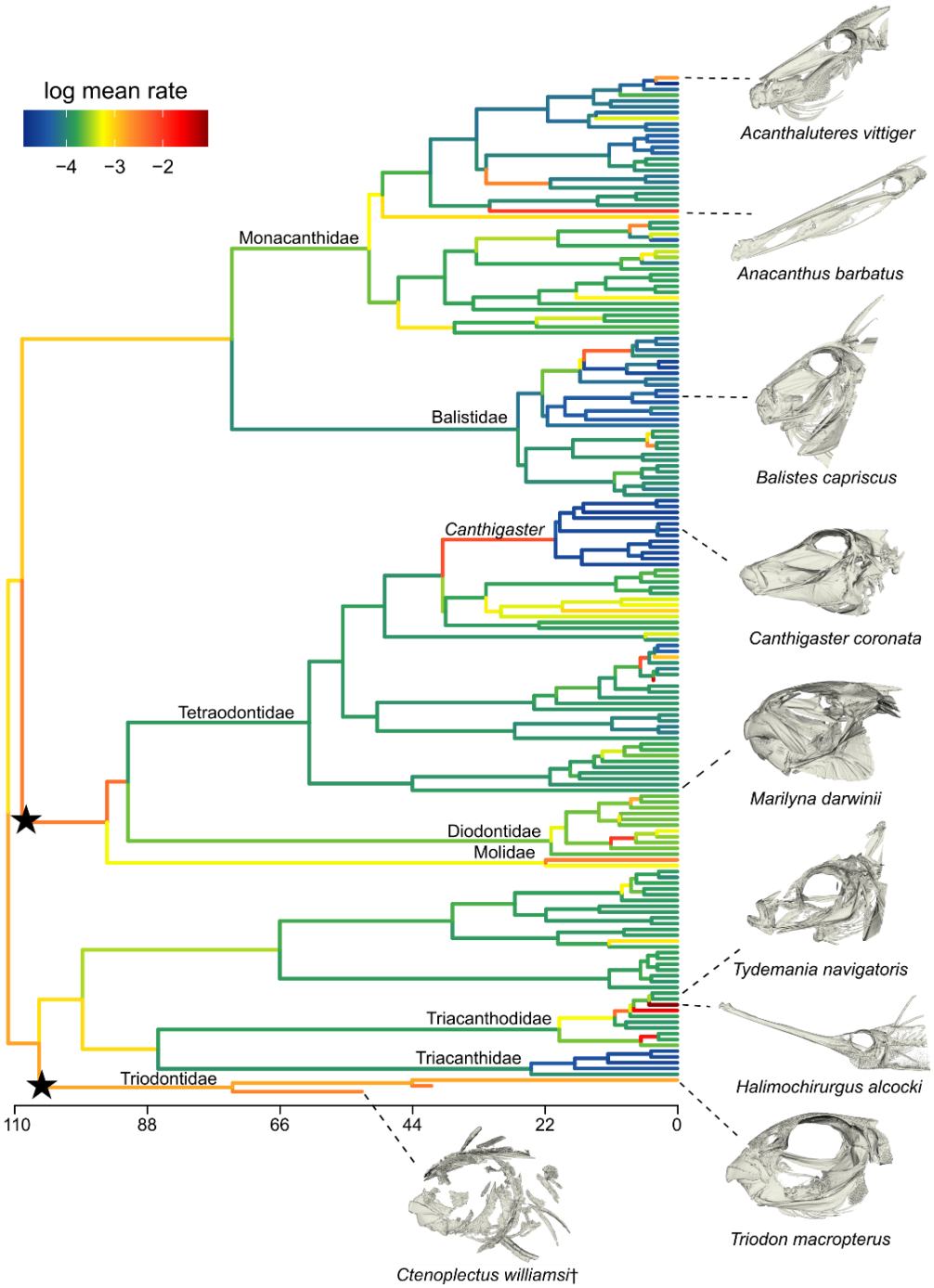


Figure 2. Tempo of skull shape evolution in Tetraodontiformes. Estimated rate shifts of skull shape evolution mapped onto the time-calibrated Tetraodontiformes phylogeny from Troyer et al. (2022) under the best-supported model of trait evolution. Color gradient on branches indicates the rate of shape evolution (log mean rate: warmer colors are higher rates, while cooler colors are slower rates). Representative CT scans of tetraodontiform skulls are shown for branches of interest. Black stars indicate branches leading to beaked families, which display faster evolutionary rates. Rates estimated by using BayesTraitsV4 (Pagel and Meade 2022).

Evolutionary modularity and integration

To assess if patterns of skull modularity and integration differ between beaked and non-beaked tetraodontiform fishes, we compared multiple *a priori* hypotheses of modularity (Figure S4) based on previous studies (Westneat 2005; Helfman et al. 2009; Kague et al. 2012; Evans et al. 2019*a*) using the phylogenetically informed analyses phyloEMMLi (Goswami and Finarelli 2016) and a covariance ratio (CR) analysis (Adams 2016). Our analyses find strong model support for an eleven-module hypothesis of modularity where each bone partition is a module (Table S3). Both beaked and non-beaked species display similar levels of skull modularity (CR beaked = 0.813; CR non-beaked = 0.865; $p = 0.66$) and skull integration (partial least squares (PLS) beaked = 0.768; PLS non-beaked = 0.727; $p = 0.87$).

Despite no significant difference in the degree of skull modularity and integration between beaked and non-beaked species, we do observe a substantial difference in rates of bone module evolution between groups. Notably, beaked species demonstrate higher rates overall than their non-beaked relatives (Figure 3). Among beaked species, the highest rates are found in the premaxilla, maxilla, frontal, and dentary. Non-beaked species, however, exhibit the highest rates in the frontal and maxilla. For bones implicated in jaw function (e.g., premaxilla, maxilla, dentary), beaked species exhibit evolutionary rates that are approximately two times faster than non-beaked taxa.

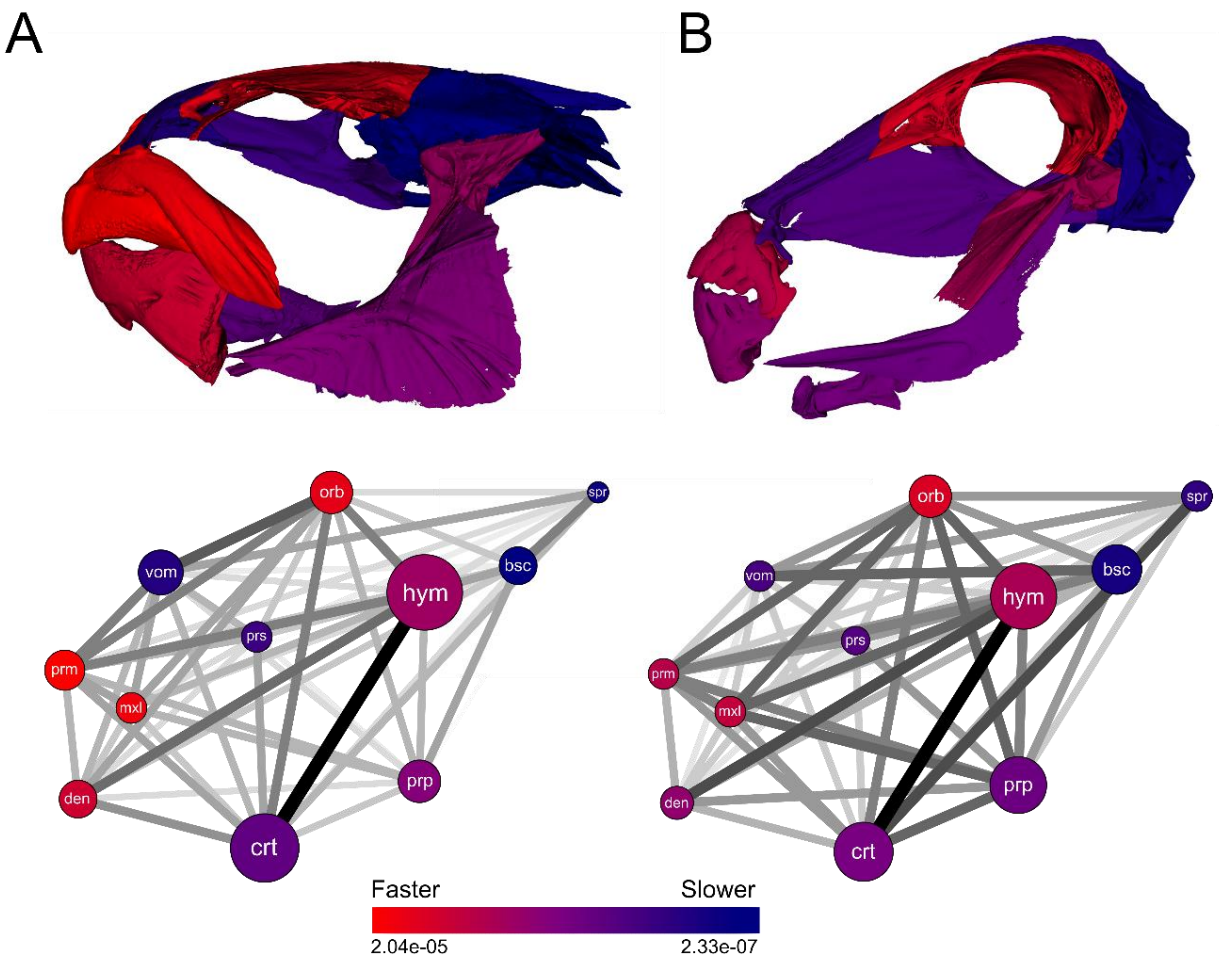


Figure 3. Rates of bone module evolution in tetraodontiform skulls. Representative skulls from a **A)** beaked (*Marilyna darwinii*) and a **B)** non-beaked (*Balistes capriscus*) tetraodontiform depicting bones modules colored by rate of shape evolution under the best fit modularity hypothesis. Network plots show the magnitude of integration between each bone module for both groups. Larger module circles indicate higher modularity, while darker lines between modules indicate higher integration. Bone modules are colored by rate of evolution. Bone abbreviations are: prm (premaxilla), den (dentary), mxl (maxilla), vom (vomer), prs (parasphenoid), crt (ceratohyal), prp (preopercle), hym (hyomandibula), orb (orbit/frontal), bsc (basioccipital), spr (supraoccipital).

Discussion

In this study, we used a three-dimensional geometric morphometric approach for extant and fossil species to examine factors promoting morphological diversification and evolutionary innovation in tetraodontiform fishes. Despite displaying conserved patterns of modularity and integration, skulls of beaked tetraodontiforms evolve at much faster rates and show higher levels of morphological disparity when compared to non-beaked tetraodontiforms, especially in bones contributing to the jaws and beak. Furthermore, contrary to long-held beliefs, we find that coral reef association does not increase the rate of skull evolution or promote morphological disparity. Instead, the presence of a beak is likely a crucial component of tetraodontiform evolutionary success.

The role of integration on morphological diversification is currently debated. Traditionally, integration has been thought to constrain phenotypic diversification, while modularity is viewed as a prerequisite to facilitate innovation, with the ability to increase morphological diversification (Yang 2001; Marroig et al. 2009; Goswami et al. 2014). More recently, integration has been suggested to aid in the evolution of innovation by promoting evolution along specific trajectories and facilitating rapid evolution within a constrained region of traitspace (Goswami et al. 2014; Navalón et al. 2020; Evans et al. 2021). In this way, integration can promote large responses to selective pressures by aiding in the evolution of maximally disparate phenotypes. However, our results show that despite beaked tetraodontiforms displaying rapid morphological diversification, there are no significant differences in patterns of modularity and integration between beaked and non-beaked species, suggesting covariation is highly conserved throughout the entire clade. This conserved pattern of skull integration and modularity is seen in other major vertebrate clades as well, including mammals, caecilians, and squamates (Porto et al. 2009; Marshall et al. 2019; Watanabe et al. 2019), and highlights the fact that rapid evolution of maximally disparate phenotypes, such as the tetraodontiform beak, can still arise while maintaining a conserved pattern and magnitude of trait integration across an entire clade.

Coral reefs have been previously linked to increases in rates of morphological evolution in fishes (Price et al. 2011, 2013). However, our results suggest a contrasting pattern, with non-reef-associated tetraodontiforms showing elevated rates of skull evolution, suggesting reef association alone may not be enough to promote large changes in their morphological diversification, and in this case, might even impede phenotypic divergence. Recent studies lend support to these findings. Evans et al. (2019*b*) examined pharyngeal jaw morphology across reef and non-reef-associated wrasses only to find no difference in rates of morphological evolution between groups. However, higher rates were found in specialized reef-associated clades, such as cleaner wrasses and parrotfishes (Evans et al. 2019*b*). We observe a similar result in Tetraodontiformes, where species in the genus *Canthigaster* display a rapid increase in the rates of skull shape evolution (Figure 2). These are small, reef-associated pufferfishes that have evolved a pointed snout, perhaps allowing them to easily maneuver into tiny crevices and feed on small benthic organisms like sponges, corals, and invertebrates (Allen and Randall 1977; Santini et al. 2013*a*). Our results indicate that a broad rule describing reefs as promoters of morphological diversity may be more nuanced than previously thought. Instead, other factors, such as specialized trophic ecologies found on reefs may play a more important role in governing morphological adaptation.

Regardless of habitat, the beak in certain tetraodontiform families is likely an important factor in their high rates of morphological diversification. This innovation enabled tetraodontiforms to enter novel dietary niches and access hard-bodied prey, such as gastropods,

crabs, and bivalves (Turingan 1994; Thiery et al. 2017). While many studies define key innovations as those that also lead to increases in lineage diversification, the original term, first coined by Miller (1949), defines key innovations as phenotypic traits that allow a lineage to exploit its environment in a novel way. Our results follow Miller's original definition of key innovation, as we observe a variable species richness among beaked tetraodontiforms. For example, while the most speciose family, Tetraodontidae, contains over 200 species, the deep-sea family, Triodontidae, represents an ancient monotypic lineage. In addition to a beak, other tetraodontiform clades possess unique morphologies, such as body inflation (Wainwright and Turingan 1997; Bemis et al. 2023), which may contribute to their diversification, and future studies should examine these aspects of their anatomy.

The evolutionary origins of the tetraodontiform beak are still debated. Previous phylogenetic hypotheses based solely on morphological data placed all four extant beaked families (Tetraodontidae, Diodontidae, Molidae, Triodontidae) into a clade (Winterbottom 1974; Santini and Tyler 2003). However, with the advent of molecular analyses, the placement of Triodontidae has become unclear, with contrasting results depending upon the type and number of loci used. Early estimations based on a handful of nuclear and mitochondrial loci resolve Triodontidae as sister to a clade containing the non-beaked boxfishes (Ostraciidae), trunkfishes (Aracanidae), spikefishes (Triacanthodidae) and the triplespines (Triacanthidae) (Santini et al. 2013*b*). More recently, ultraconserved element (UCE) data from 989 loci resolves Triodontidae as sister to Ostraciidae+Aracanidae (Ghezelayagh et al. 2022). Even incorporating morphological and molecular data under a total evidence approach fails to yield consistent results. For instance, using 15 nuclear and one mitochondrial marker in conjunction with 210 morphological characters Arcila et al. (2015) resolved Triodontidae as the sister lineage of all other beaked families. More recently, Troyer et al. (2022) expanded upon this dataset to include an additional 1,103 loci based on exon capture, resulting in Triodontidae being sister to the clade of Ostraciidae+Aracanidae and Triacanthidae+Triacanthodidae. With difficulties of placing Triodontidae into the tetraodontiform phylogeny, this raises the question of whether the tetraodontiform beak has a single evolutionary origin or represents convergent or parallel morphologies arising multiple independent times.

Altogether, our results suggest that despite a conserved pattern and magnitude of trait integration, we still see increased morphological rates in the jaws of beaked tetraodontiforms, suggesting that the beak is an important structure promoting their overall morphological diversification. Additionally, rules that were previously thought to be broadly applicable, such as reef associations driving morphological diversity, are perhaps not applicable across large taxonomic scales and more nuanced situations, such as trophic specialization, may better explain this phenomenon.

Acknowledgments

We thank the following people and collections that provided specimens for CT scanning, Jessica Nakano, Amanda Millhouse, Diane Pitassy, and Abigail Reft (Smithsonian Institution National Museum of Natural History), Caleb McMahan (Field Museum of Natural History), Alastair Graham and John Pogonoski (Commonwealth Scientific and Industrial Research Organisation [CSIRO] Australian National Fish Collection, Hobart), Katherine Maslenikov (University of Washington Fish Collection), Mariangeles Arce H. and Mark Sabaj (The Academy of Natural Sciences), Dave Catania (California Academy of Sciences), and Amanda Hay, Sally Reader, and Kerryn Parkinson (Australian Museum). We also thank Grace Breitenbeck and Eric Hilton for

specimen transportation, Adam Summers and the oVert and ScanAllFishes initiatives for allowing us to micro-CT scan fishes at Friday Harbor Labs, Richard Flavel at UNE Armidale for assistance with micro-CT scanning Australian Museum specimens, Teresa Porri of the Biotechnology Resource Center Multiscale Imaging Facility at Cornell University, and Bryan Juarez for assistance plotting disparity through time.

Author Contributions

E.M.T and D.A. designed research; E.M.T. performed research and analyzed data; E.M.T., K.M.E., C.G., M.F., B.N., M.K., K.E.B. collected CT scan data; G.C. assisted with fossil habitat categorizations; and E.M.T. and D.A. wrote the paper with contributions from all authors.

Data Availability Statement

All scripts and data produced for this study will be deposited in the Dryad repository (doi:10.5061/dryad.4f4qrfjjx).

Funding

This project was supported by National Science Foundation (NSF) grants to D.A. (DEB-2015404, DEB-2144325) and K.M.E. (DEB-2237278). G.C. was supported by funds (ex-60% 2023) from the Università degli Studi di Torino. Scans of fossils at the Natural History Museum, London were made with support of Leverhulme Trust grant number RPG-2012-658. This study includes data produced in the CTEES facility at University of Michigan, supported by the Department of Earth and Environmental Sciences and College of Literature, Science, and the Arts.

References

Adams, D. C. 2016. Evaluating modularity in morphometric data: challenges with the RV coefficient and a new test measure. (P. Peres-Neto, ed.)Methods in Ecology and Evolution 7:565–572.

Adams, D. C., and M. L. Collyer. 2016. On the comparison of the strength of morphological integration across morphometric datasets. Evolution 70:2623–2631.

———. 2018. Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations. Systematic Biology 67:14–31.

Alfaro, M. E., F. Santini, and C. D. Brock. 2007. Do Reefs Drive Diversification in Marine Teleosts? Evidence from the Pufferfish and Their Allies (order Tetraodontiformes). Evolution 61:2104–2126.

Allen, G. R., and J. E. Randall. 1977. Review of the sharpnose pufferfishes (subfamily Canthigasterinae) of the Indo-Pacific. Records of the Australian Museum 30:475–517.

Arbour, J., and C. M. Brown. 2020. LOST: Missing Morphometric Data Simulation and Estimation.

Arbour, J. H., and C. M. Brown. 2014. Incomplete specimens in geometric morphometric analyses. (S. Nakagawa, ed.)Methods in Ecology and Evolution 5:16–26.

Arcila, D., R. Alexander Pyron, J. C. Tyler, G. Ortí, and R. Betancur-R. 2015. An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae). Molecular Phylogenetics and Evolution 82:131–145.

Baken, E. K., M. L. Collyer, A. Kaliontzopoulou, and D. C. Adams. 2021. geomorph v4.0 and gmShiny: Enhanced analytics and a new graphical interface for a comprehensive morphometric experience. Methods in Ecology and Evolution 12:2355–2363.

Bemis, K. E., J. C. Tyler, A. Kaneko, K. Matsuura, K. Murakumo, V. C. Espíndola, J.-L. Justine, et al. 2023. Pelvic-Fan Flaring and Inflation in the Three-Tooth Puffer, Triodon macropterus (Tetraodontiformes: Triodontidae), with Additional Observations on Their Behavior in Captivity. Ichthyology & Herpetology 111:222–240.

Brandl, S. J., C. H. R. Goatley, D. R. Bellwood, and L. Tornabene. 2018. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. Biological Reviews 93:1846–1873.

Bray, D., and M. Gomon. 2023. Fishes of Australia. Museums Victoria and OzFishNet.

Buser, T. J., V. E. Kee, R. C. Terry, A. P. Summers, and B. L. Sidlauskas. 2023. Taurus of the Tidepool? Inferring the Function of Cranial Weapons in Intertidal Sculpins (Pisces: Cottoidea: Oligocottinae). Ichthyology & Herpetology 111:98–108.

Cignoni, P., M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. 2008. MeshLab: an Open-Source Mesh Processing Tool.

Close, R. A., Z. Johanson, J. C. Tyler, R. C. Harrington, and M. Friedman. 2016. Mosaicism in a new Eocene pufferfish highlights rapid morphological innovation near the origin of crown tetraodontiforms. Palaeontology 59:499–514.

Cowman, P. F., and D. R. Bellwood. 2011. Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots: Cladogenesis on coral reefs. Journal of Evolutionary Biology 24:2543–2562.

Evans, K. M., O. Larouche, S. M. Gartner, R. E. Faucher, S. G. Dee, and M. W. Westneat. 2023*a*. Beaks promote rapid morphological diversification along distinct evolutionary trajectories in labrid fishes (Eupercaria: Labridae). (T. Stayton & T. Connallon, eds.)Evolution qpad115.

Evans, K. M., O. Larouche, S.-J. Watson, S. Farina, M. L. Habegger, and M. Friedman. 2021. Integration drives rapid phenotypic evolution in flatfishes. Proceedings of the National Academy of Sciences 118:e2101330118.

Evans, K. M., O. Larouche, J. L. West, S. M. Gartner, and M. W. Westneat. 2023*b*. Burrowing constrains patterns of skull shape evolution in wrasses. Evolution & Development 25:73–84.

Evans, K. M., M. Vidal-García, V. A. Tagliacollo, S. J. Taylor, and D. B. Fenolio. 2019*a*. Bony Patchwork: Mosaic Patterns of Evolution in the Skull of Electric Fishes (Apteronotidae: Gymnotiformes). Integrative and Comparative Biology 59:420–431.

Evans, K. M., B. T. Waltz, V. A. Tagliacollo, B. L. Sidlauskas, and J. S. Albert. 2017. Fluctuations in Evolutionary Integration Allow for Big Brains and Disparate Faces. Scientific Reports 7:40431.

Evans, K. M., K. L. Williams, and M. W. Westneat. 2019*b*. Do Coral Reefs Promote Morphological Diversification? Exploration of Habitat Effects on Labrid Pharyngeal Jaw Evolution in the Era of Big Data. Integrative and Comparative Biology 59:696–704.

Fraser, G. J., R. Britz, A. Hall, Z. Johanson, and M. M. Smith. 2012. Replacing the first-generation dentition in pufferfish with a unique beak. Proceedings of the National Academy of Sciences 109:8179–8184.
Friedman, M., and G. Carnevale. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. Journal of the Geological Society 175:569–579.
Froese, R., and D. Pauly. 2023. FishBase.
Ghezelayagh, A., R. C. Harrington, E. D. Burress, M. A. Campbell, J. C. Buckner, P. Chakrabarty, J. R. Glass, et al. 2022. Prolonged morphological expansion of spiny-rayed fishes following the end-Cretaceous. Nature Ecology & Evolution 6:1211–1220.
Gibb, A. C., K. Staab, C. Moran, and L. A. Ferry. 2015. The Teleost Intramandibular Joint: A mechanism That Allows Fish to Obtain Prey Unavailable to Suction Feeders. Integrative and Comparative Biology 55:85–96.
Goswami, A., and J. A. Finarelli. 2016. EMMLi: A maximum likelihood approach to the analysis of modularity. Evolution 70:1622–1637.
Goswami, A., J. B. Smaers, C. Soligo, and P. D. Polly. 2014. The macroevolutionary consequences of phenotypic integration: from development to deep time. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130254.
Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.
Heard, S. B., and D. L. Hauser. 1995. Key evolutionary innovations and their ecological mechanisms. Historical Biology 10:151–173.
Helfman, G. S., B. B. Collette, D. E. Facey, and B. W. Bowen, eds. 2009. The diversity of fishes: biology, evolution, and ecology (2nd ed.). Blackwell, Chichester, UK ; Hoboken, NJ.
Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback. 2003. Stochastic Mapping of Morphological Characters. Systematic Biology 52:131–158.
Jablonski, D. 2022. Evolvability and Macroevolution: Overview and Synthesis. Evolutionary Biology 49:265–291.
Kague, E., M. Gallagher, S. Burke, M. Parsons, T. Franz-Odendaal, and S. Fisher. 2012. Skeletogenic Fate of Zebrafish Cranial and Trunk Neural Crest. (H. H. Roehl, ed.)PLoS ONE 7:e47394.
Kikinis, R., S. D. Pieper, and K. G. Vosburgh. 2014. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. Pages 277–289 *in* F. A. Jolesz, ed. Intraoperative Imaging and Image-Guided Therapy. Springer, New York, NY.
Liem, K. F. 1973. Evolutionary Strategies and Morphological Innovations: Cichlid Pharyngeal Jaws. Systematic Biology 22:425–441.
Marramà, G., A. F. Bannikov, J. C. Tyler, R. Zorzin, and G. Carnevale. 2016. Controlled excavations in the Pesciara and Monte Postale sites provide new insights about the palaeoecology and taphonomy of the fish assemblages of the Eocene Bolca Konservat-Lagerstätte, Italy. Palaeogeography, Palaeoclimatology, Palaeoecology 454:228–245.
Marroig, G., L. T. Shirai, A. Porto, F. B. de Oliveira, and V. De Conto. 2009. The Evolution of Modularity in the Mammalian Skull II: Evolutionary Consequences. Evolutionary Biology 36:136–148.
Marshall, A. F., C. Bardua, D. J. Gower, M. Wilkinson, E. Sherratt, and A. Goswami. 2019. High-density three-dimensional morphometric analyses support conserved static

(intraspecific) modularity in caecilian (Amphibia: Gymnophiona) crania. Biological Journal of the Linnean Society 126:721–742.
Mayr, E. 1963. Animal Species and Evolution. Cambridge, MA: Belknap.
Miller, A. H. 1949. Some ecologic and morphologic considerations in the evolution of higher taxonomic categories. Ornithologie als biologische Wissenschaft 84.
Miller, A. H., and J. T. Stroud. 2022. Novel Tests of the Key Innovation Hypothesis: Adhesive Toepads in Arboreal Lizards. Systematic Biology 71:139–152.
Murrell, D. J. 2018. A global envelope test to detect non-random bursts of trait evolution. Methods in Ecology and Evolution 9:1739–1748.
Navalón, G., J. Marugán-Lobón, J. A. Bright, C. R. Cooney, and E. J. Rayfield. 2020. The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. Nature Ecology & Evolution 4:270–278.
Oba, S., M. Sato, I. Takemasa, M. Monden, K. Matsubara, and S. Ishii. 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics 19:2088–2096.
Olson, E. C., and R. L. Miller. 1999. Morphological Integration. University of Chicago Press, Chicago, IL.
Pagel, M., and A. Meade. 2022. BayesTraits, version 4. University of Reading, Berkshire, UK.
Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics (Oxford, England) 35:526–528.
Porto, A., F. B. de Oliveira, L. T. Shirai, V. De Conto, and G. Marroig. 2009. The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. Evolutionary Biology 36:118–135.
Price, S. A., R. Holzman, T. J. Near, and P. C. Wainwright. 2011. Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes: Reefs promote fish morphological diversity. Ecology Letters 14:462–469.
Price, S. A., J. J. Tavera, T. J. Near, and Peter. C. Wainwright. 2013. Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. Evolution 67:417–428.
Price, S. A., P. C. Wainwright, D. R. Bellwood, E. Kazancioglu, D. C. Collar, and T. J. Near. 2010. Functional innovations and morphological diversification in parrotfish. Evolution 64:3057–3068.
R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67:901–904.
Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things).
Rhoda, D., P. D. Polly, C. Raxworthy, and M. Segall. 2021*a*. Morphological integration and modularity in the hyperkinetic feeding system of aquatic-foraging snakes. Evolution 75:56–72.
Rhoda, D., M. Segall, O. Larouche, K. Evans, and K. D. Angielczyk. 2021*b*. Local Superimpositions Facilitate Morphometric Analysis of Complex Articulating Structures. Integrative and Comparative Biology 61:1892–1904.

Rohlf, F. J., and M. Corti. 2000. Use of Two-Block Partial Least-Squares to Study Covariation in Shape. Systematic Biology 49:740–753.
Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. Systematic Zoology 39:40.
Santini, F., M. T. T. Nguyen, L. Sorenson, T. B. Waltzek, J. W. Lynch Alfaro, J. M. Eastman, and M. E. Alfaro. 2013*a*. Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). Journal of Evolutionary Biology 26:1003–1018.
Santini, F., L. Sorenson, and M. E. Alfaro. 2013*b*. A new phylogeny of tetraodontiform fishes (Tetraodontiformes, Acanthomorpha) based on 22 loci. Molecular Phylogenetics and Evolution 69:177–187.
Santini, F., and J. C. Tyler. 2003. A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to Recent. Zoological Journal of the Linnean Society 139:565–617.
Slater, G. J., and M. W. Pennell. 2014. Robust Regression and Posterior Predictive Simulation Increase Power to Detect Early Bursts of Trait Evolution. Systematic Biology 63:293–308.
Souter, D., S. Planes, J. Wicquart, D. Obura, and F. Staub. 2021. *Status of Coral Reefs of the World: 2020*. International Coral Reef Initiative.
Thiery, A. P., T. Shono, D. Kurokawa, R. Britz, Z. Johanson, and G. J. Fraser. 2017. Spatially restricted dental regeneration drives pufferfish beak development. Proceedings of the National Academy of Sciences 114.
Troyer, E. M., R. Betancur-R, L. C. Hughes, M. Westneat, G. Carnevale, W. T. White, J. J. Pogonoski, et al. 2022. The impact of paleoclimatic changes on body size evolution in marine fishes. Proceedings of the National Academy of Sciences 119:e2122486119.
Turingan, R. G. 1994. Ecomorphological relationships among Caribbean tetraodontiform fishes. Journal of Zoology 233:493–521.
Tyler, J. C. 1980. Osteology, phylogeny, and higher classification of the fishes of the order Plectognathi (Tetraodontiformes). U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, [Seattle, Wash.].
Venditti, C., A. Meade, and M. Pagel. 2011. Multiple routes to mammalian diversity. Nature 479:393–396.
Wagner, G. P., and L. Altenberg. 1996. Complex adaptations and the evolution of evolvability. Evolution; International Journal of Organic Evolution 50:967–976.
Wainwright, P. C., and R. G. Turingan. 1997. Evolution of pufferfish inflatoin behavior. Evolution 51:506–518.
Watanabe, A., A.-C. Fabre, R. N. Felice, J. A. Maisano, J. Müller, A. Herrel, and A. Goswami. 2019. Ecomorphological diversification in squamates from conserved pattern of cranial integration. Proceedings of the National Academy of Sciences 116:14688–14697.
Westneat, M. W. 2005. Skull Biomechanics and Suction Feeding in Fishes. Pages 29–75 *in*Fish Physiology (Vol. 23). Elsevier.
Winterbottom, R. 1974. The familial phylogeny of the Tetraodontiformes (Acanthopterygii: Pisces) as evidenced by their comparative myology. Smithsonian Contributions to Zoology 1–201.
Yang, A. S. 2001. Modularity, evolvability, and adaptive radiations: a comparison of the hemi- and holometabolous insects. Evolution and Development 3:59–72.

Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, T. J. Hagey, et al. 2010. Ecological opportunity and the origin of adaptive radiations. Journal of Evolutionary Biology 23:1581–1596.

Zelditch, M. L., and D. L. Swiderski. 2023. Effects of Procrustes Superimposition and Semilandmark Sliding on Modularity and Integration: An Investigation Using Simulations of Biological Data. Evolutionary Biology 50:147–169.

They might not be giants: Genetic convergence on miniaturization in gobiid fishes

Emily M. Troyer[1], Dahiana Arcila[2]

[1]*Department of Biology and Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, 730 Van Vleet Oval, Richards Hall, Norman, OK 73019, USA*

[2]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA*

Abstract

Organism size is a key biological property, associated with numerous trade-offs between various life history traits. Large bodies may be advantageous due to their association with longevity, higher fecundity, and wider geographic ranges. However, smaller bodies display higher developmental rates and earlier reproduction times. On the far end of the size spectrum, miniaturization, the extreme reduction of adult body size, has evolved numerous times across the Tree of Life. Miniaturization is especially common among fishes, with species in the family Gobiidae (gobies) being an exceptional case. Despite how common miniaturization appears, its genetic mechanisms are poorly understood. There are multiple, independent transitions to miniaturization within Gobiidae, allowing for tests of genetic convergence within a comparative macroevolutionary framework. Here, we assemble the first *de novo* transcriptomes for six species of gobiid fishes, which represent three clades, each containing a closely related large-bodied and small-bodied species. We identify sets of statistically significant orthologs which are differentially expressed between large-bodied and small-bodied species in each clade. From these, we identify several candidate genes potentially involved in miniaturization, including *ybx1* and *bzw2*, both known to affect cell growth and development. These candidate genes offer insight into the genetic convergence on miniature body size and provide a framework for future studies.

Introduction

Organism size is one of the most fundamental species traits, correlated with numerous biological processes from metabolism to reproduction[1]. In general, larger bodied species will inhabit wider geographic ranges, possess longer lifespans, and display higher fecundity levels compared to their smaller counterparts[1–3]. However, small-bodied species often have faster developmental rates and earlier reproduction times, demonstrating the trade-offs between size and life history strategies. On the very far end of the size spectrum, the extreme reduction of adult body size, or miniaturization, occurs. This phenomenon has evolved numerous times across the Tree of Life, and has various consequences which extend to different aspects of a species' life history.

A large number of morphological, behavioral, and ecological consequences arise from miniaturization. Miniaturized taxa are commonly characterized by having greatly reduced or simplified morphologies which can range from underdeveloped to the complete loss of certain structures[4–6]. For instance, some of the world's smallest vertebrates, cyprinid fishes in the genus *Paedocypris*, have highly simplified anatomies which include narrow frontal bones which leave a portion of the brain exposed[7,8]. Miniaturization is also highly associated with morphological novelties due to structural reorganization of the body plan[4]. One of the more striking examples of this is the increased flight ability associated with miniaturization of avian dinosaurs[9]. Ecological and behavioral changes can often accompany the evolution of miniaturization, where these changes lead to the evolution of parasitism[4]. Such examples include the miniaturized male ceratioid anglerfishes which parasitize females of their own species[10] and numerous invertebrates such as parasitoid fig wasps[11] and the world's smallest fly, *Euryplatea nanaknihali*, which parasitizes ants[12].

Miniaturization is found in throughout the Tree of Life with examples in mammals[13], birds[14], amphibians[5], reptiles[15], and insects[16], and fishes[8,17,18]. Among fishes, the family Gobiidae (gobies) is extremely diverse, containing *ca*. 2,000 known species[19]. Most gobiids are relatively small (<100 mm), but there are several instances of miniaturization that occur across several genera, such as the pygmygobies (*Trimma* spp.) the dwarfgobies (*Eviota* spp.), and the paedomorphic infantfishes (*Schindleria* spp.), all of which rarely exceed 25 mm in length[18,20,21]. Due to these multiple, independent instances of miniaturization which appear across their phylogeny, gobiids offer a unique model system in which to test hypothesis relating to the evolution of miniaturization. While we understand much about the morphological characteristics of miniaturization, its genomic basis is not well understood. There have been several studies investigating candidate genes associated with body size evolution in fishes[22–24], however none have explicitly analyzed miniaturized taxa across a macroevolutionary scale.

Here, we investigate the genomic basis of miniaturization by assembling the first transcriptomes for six species of gobiid and comparing gene expression between pairs of closely related large- and small-bodied species within three separate clades. We test for genetic convergence on small body size by analyzing differentially expressed one-to-one orthologs. Orthologs that are expressed at statistically higher levels in small-bodied species as compared to large-bodied species will be considered candidate genes for miniaturization. In addition, we infer a new phylogeny for 134 gobiid species using genome-wide data and reconstruct ancestral body size along the phylogeny. This is the first study utilizing a comparative transcriptome approach to analyze the evolutionary basis of miniaturization in gobiid fishes, which are often the poster child for small-bodied fishes on coral reefs.

Materials and Methods

Sampling and sequencing

We sourced three individual gobiids of six species (*Amblyeleotris guttata*, *Coryphopterus personatus*, *Eviota atriventris*, *Gobiodon citrinus*, *Trimma hollemani*, and *Valenciennea puellaris*) from the aquarium trade. These species represent three pairs of closely related taxa where a large-bodied and a small-bodied species are in close phylogenetic relationship and are also readily available through the aquarium trade. Fishes were held for 24-48 hours in aquaria before being euthanized with MS-222 in accordance with IACUC guidelines (approval #R19-018). To

minimize gene expression differences due to developmental differences, we sampled only adult fishes. Additionally, all individuals were sampled between 11 am and 4 pm to minimize differences from circadian rhythms. Individuals were sampled in eight batches according to their availability from the aquarium trade (Table S1). We sampled from tissues from the liver and skeletal muscle, as these are important in growth and metabolism[24]. Tissues were extracted and flash-frozen in liquid nitrogen then stored at -80 °C. Tissues samples were shipped to Novogene (https://www.novogene.com/us-en/) for outsourcing of RNA extraction and reference-based sequencing. Total mRNA was extracted from each tissue sample using the poly A enrichment method, where poly-T oligo-attached magnetic beads are used to pull down the polyadenylated (poly-A) tail of mRNA. RNA Integrity Numbers (RIN) were calculated for each sample and only samples with RIN > 4.0 were used. One sample (*Trimma hollemani* liver) failed and was excluded from the analysis, leaving 35 remaining samples. Sequencing was conducted on an Illumina platform using 150 bp paired end reads. Raw reads were filtered to remove low-quality reads and adapter contamination and submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA1088833. After quality control, sequences were aligned using HISAT2 v. 2.0.5[25] to the closest available reference genome, *Boleophthalmus pectinirostris*, the great blue-spotted mudskipper (belonging to the sister family Oxudercidae), obtained from NCBI. Once aligned, sequences were mapped back to the reference genome. However, because of the low percentage of mapped reads to the reference genome (Table S1), likely due to the phylogenetic distance between the reference and the samples, we chose to conduct a *de novo* assembly in Trinity, for which all subsequent analyses are based (see *De novo transcriptome assembly).*

Phylogenomic estimations

We generated new genomic data from tissue samples extracted from museum specimens for 40 species of gobiids and 27 outgroup taxa of families in Gobioidei. DNA was extracted and shipped to Arbor Biosciences for library preparation and target enrichment. Libraries for all samples were processed at the sequencing facilities at the University of Chicago. Sequencing of pair end 100 bp reads was completed on a HiSeq 4000 with a total of 192 samples multiplexed per lane, including samples for other projects not listed here. Raw reads will be deposited to NCBI Sequence Read Archive under BioProject number PRJNA1088833. Target capture probes are based on a set of 1,103 single-copy nuclear exon markers for teleost fishes (Hughes et al. 2018). To increase our taxonomic scope, we mined these same exon markers from genomes and transcriptomes of 97 gobiid species and 8 outgroups from the NCBI database using the pipeline from Hughes et al.[26]. We also mined these exons from the 6 species-specific *de novo* transcriptome assemblies (see *De novo transcriptome assembly*).

Exons were aligned using MACSE v. 2.03[27] after cleaning potentially non-homologous fragments with the -cleanNoHomologousSequences flag. After alignment, an additional filtering step was run to clean up any single-taxon insertions caused by assembly errors, remove edges containing more than 60% gaps and short sequences below 50% of the alignment length using the 'AlignmentCleanerCodons.py' script from Hughes et al.[26]. After alignment, an initial phylogeny was estimated with FastTree v. 2.1.1[28] to assess quality. We pruned the alignment to include one individual per species and we removed 13 individuals due to potential contamination and misidentification. Additionally, we removed exons with low pairwise identity (<80%), leaving a total of 980 exon markers for phylogenomic inference. After these quality control steps, we ran

another FastTree to confirm. Our final taxonomic sampling includes 165 species (134 gobiids, 31 outgroup taxa from three Gobiodei families [Oxudercidae, Eleotridae, and Odontobutidae]).

We estimated the Gobiodei phylogeny under a maximum likelihood framework using a concatenated RAxML-NG v. 1.1.0[29] analysis, partitioned by codon position. Non-parametric bootstrap support values were estimated using 1,000 replicates. To account for discordance from incomplete lineage sorting, we inferred individual gene trees in IQTREE v. 1.6.12[30] which were used to estimate a species tree in ASTRAL-III[31]. Branch support was estimated using an ultra-fast bootstrapping approach with 1,000 bootstrap replicates.

Body size data

We recorded available standard length (SL) and total length (TL) data from FishBase[19] for each species in the phylogeny. Length data was unavailable on FishBase for six species (*Egglestonichthys bombylios, Kraemeria bryani, Lophogobius cristulatus, Lythrypnus crocodilus, Sicydium gymnogaster, Rhinogobius lindbergi*) so we recorded total and standard length from other sources, such as published papers. Total length was the most commonly available measurement. For those species missing TL sizes, we calculated a linear relationship (R^2=0.9967) between TL and SL based on measurements from the 18 individual gobiids obtained from the aquarium trade (see *Sampling and Sequencing*). We use the equation (TL=1.257 *SL – 0.0818cm) to calculate TL of the missing FishBase species, so that every species in our dataset has an associated total length for use in downstream analysis (Table S2). To infer body size across the Gobioidei phylogeny, we ran a maximum likelihood ancestral character state reconstruction for total length using the "fastAnc" function in the *phytools* package[32] in R v. 4.2.3[33].

De novo transcriptome assembly

Raw read quality was assessed using FastQC v. 0.11.5[34]. Low quality reads and adaptor sequences were removed using Trimmomatic v. 0.38[35] using the following parameters: ILLUMINACLIP:2:30:10, LEADING:5, TRAILING:5, SLIDINGWINDOW:4:15, MINLEN:31. Paired reads were retained, and quality was assessed again with FastQC. Following pipeline 4 from White et al.[36], we generated species specific *de novo* assemblies for each of the six gobiids using Trinity v. 2.15.1[37]. Read normalization within species to the default k-mer maximum coverage of 200 was carried out prior to assembly with the –normalize_reads flag and the strand-specific parameter included with the –SS_lib_type RF flag to indicate paired reads. Transcripts for each species-specific transcriptome assembly were indexed using Salmon v. 1.10.1[38] and then reads were mapped back to their respective assembly and transcript abundances were estimated using salmon quant with the '-l A -p 8 --validateMappings –gcBias' options. The percentage of reads which mapped back to the respective species-specific assembly were recorded (Table S3).

We assessed transcriptome quality in terms of completeness of gene content with BUSCO v. 5.2.2 (Benchmarking Universal Single Copy Orthologs)[39] using both the Vertebrata and the Actinopterygii (ray-finned fishes) databases as reference (Table S4). We used BowTie2 v. 2.3.4.2[40] to assess assembly quality in terms of read representation where normalized reads are mapped back onto the transcriptome. In general, a mapping rate greater than 80% is ideal. Additional assembly summary statistics, including percent GC, N50, and average contig length were calculated using the 'TrinityStats.pl' script contained within Trinity (Table S4).

For each species-specific assembly, peptide open read frames (ORFs) were predicted using TransDecoder v. 5.7.1[41] and the longest ORF was retained (> 100 amino acids in length). To reduce redundancy, we clustered highly similar peptide sequences to select only one transcript per gene

using CD-HIT v. 4.8.1[42] under a 0.995 similarity threshold. These files were prepared for OrthoFinder v. 2.5.4[43] by initially running the '-op -S blast -f' option in OrthoFinder to run Basic Local Alignment Search Tool (BLAST)[44] searches separately. Pairwise query-database protein BLAST searches were then run with blastp using the options 'outfmt 6 -evalue 0.001'[45]. Orthologs were identified in OrthoFinder using the '-S blast -M msa -a 10' options. OrthoFinder summary results, including number and percentage of genes per orthogroup, were recorded (Table S5). One-to-one orthologs between each species pair per clade (i.e. Clade 1: *Amblyeleotris guttata* and *Coryphopterus personatus;* Clade 2: *Eviota atriventris* and *Gobiodon citrinus;* and Clade 3: *Trimma hollemani* and *Valenciennea puellaris*) were extracted from the OrthoFinder Gene Count output, as well as the one-to-one orthologs for all six species together, and all orthologs (regardless of being one-to-one) for all species, for use in all downstream analyses.

Gene expression at the level of the one-to-one orthologs was quantified from the Salmon abundances using the 'abundance_estimates_to_matrix.pl' script in the Trinity program, creating an un-normalized counts matrix for use in the differential expression analyses. Hierarchical cluster analyses were performed using the Pearson method to ensure tissue samples clustered according to their biological replicate.

Differential gene expression analyses

For each of the three clades of large-small species using the one-to-one orthologs, as well as for comparisons between all large species and all small species using all orthologs, we ran differential expression analyses using the DESeq2 method[46] implemented in R. This method uses the un-normalized counts matrix as input and detects differentially expressed genes (orthologs in this case) by fitting a generalized linear model of raw gene counts which follow a negative binomial (Gamma-Poisson) distribution[46]. Results tables containing the log2 fold changes (LFC), Wald test *p*-values, and Benjamini and Hochberg adjusted *p*-values of the False Discovery Rate (FDR) for each differentially expressed one-to-one ortholog were generated with the 'results' function. One-to-one orthologs were considered to be differentially expressed if the LFC value was greater than 2 and adjusted FDR *p*-value less than 0.05. These results are summarized in Table S6. To visualize differentially expressed orthologs, we constructed cluster heatmaps and volcano plots in R.

To identify potential genes involved in miniaturization, we extracted the differentially expressed orthologs which were upregulated (positive LFC) in all small-bodied species and down-regulated (negative LFC) in all large-bodied species, and vice-versa, for both tissue types individually. We blasted the ortholog protein sequences against the UniProtKB/Swiss-Prot reference proteome databases[47,48] under an expectation value threshold of 0.001 and recorded the gene names and associated biological process and function.

Functional annotation of orthologs

To infer protein function of the upregulated differentially expressed orthologs for both the small-bodied species and the large-bodied species we used the Protein Analysis Through Evolutionary Relationships (PANTHER) database[49] and the 'pantherScore2.2.pl' script to score protein sequences against the PANTHER HMM library, and generated a generic mapping file. We used this generic mapping file as input at https://pantherdb.org/ to generate a functional classification list containing the gene ontology (GO) terms for each ortholog. We summarized the GO terms associated with biological processes using REVIGO[50] with the UniProt database[48] as reference. REVIGO summarizes a list of GO terms using a semantic clustering algorithm which detects

representative subsets of the GO terms and plots these terms as a scatterplot, clustering highly similar GO terms together.

Results

Phylogenetic estimations and body size analysis

Under a maximum likelihood framework which utilizes a phylogenomic dataset based on 980 exon markers, we estimate a phylogenetic tree for 134 species of gobiid fishes and 31 species from three outgroup families within Gobioidei (Figure S1). Our approach combines both newly generated genomic sequences as well as previously published sequences. Additionally, we generated a species-tree with bootstrap support values based on a summary multispecies coalescent approach (Figure S2). Both trees are largely similar, with the only incongruence being where *Larsonella pumila* is included in Gobiidae in the RAxML tree, but within Oxudercidae in the ASTRAL species tree. This is likely due to the very long branch length in the RAxML tree (Figure S1).

Our phylogeny is largely congruent compared to the most recently published extensive gobiid phylogeny from McCraney et al.[51], which is based 23 loci for 827 taxa (367 of which are gobiids). We recovered expected topology regarding the placement of the four Gobioidei families sampled, with Oxudercidae as sister to Gobiidae, and Eleotridae and Odonotobutidae as outgroups to Oxudercidae+Gobiidae. One notable exception lies in the placement of the paedomorphic infantfishes in the genus *Schindleria*, where we recover this group as sister to all other gobiids. However, in the recent McCraney phylogeny, *Schindleria* is nested within the gobiid tree as sister to the Gobiodon lineage (consisting of the genera *Gobiodon, Paragobiodon, Eviota, Pleurosicya,*and *Bryaninops*).

Our ancestral character reconstruction reveals that the family Gobiidae evolved from a larger-bodied ancestor, transiting from a log total length of 3.01 cm to 2.15 cm (Figure 1). Along with this initial transition to smaller body size, there exists more multiple independent transitions to miniaturization with the gobiids. These transitions are especially prominent within the dwarfgobies (*Eviota* spp.) and the paedomorphic infantfishes (*Schindleria* spp.).

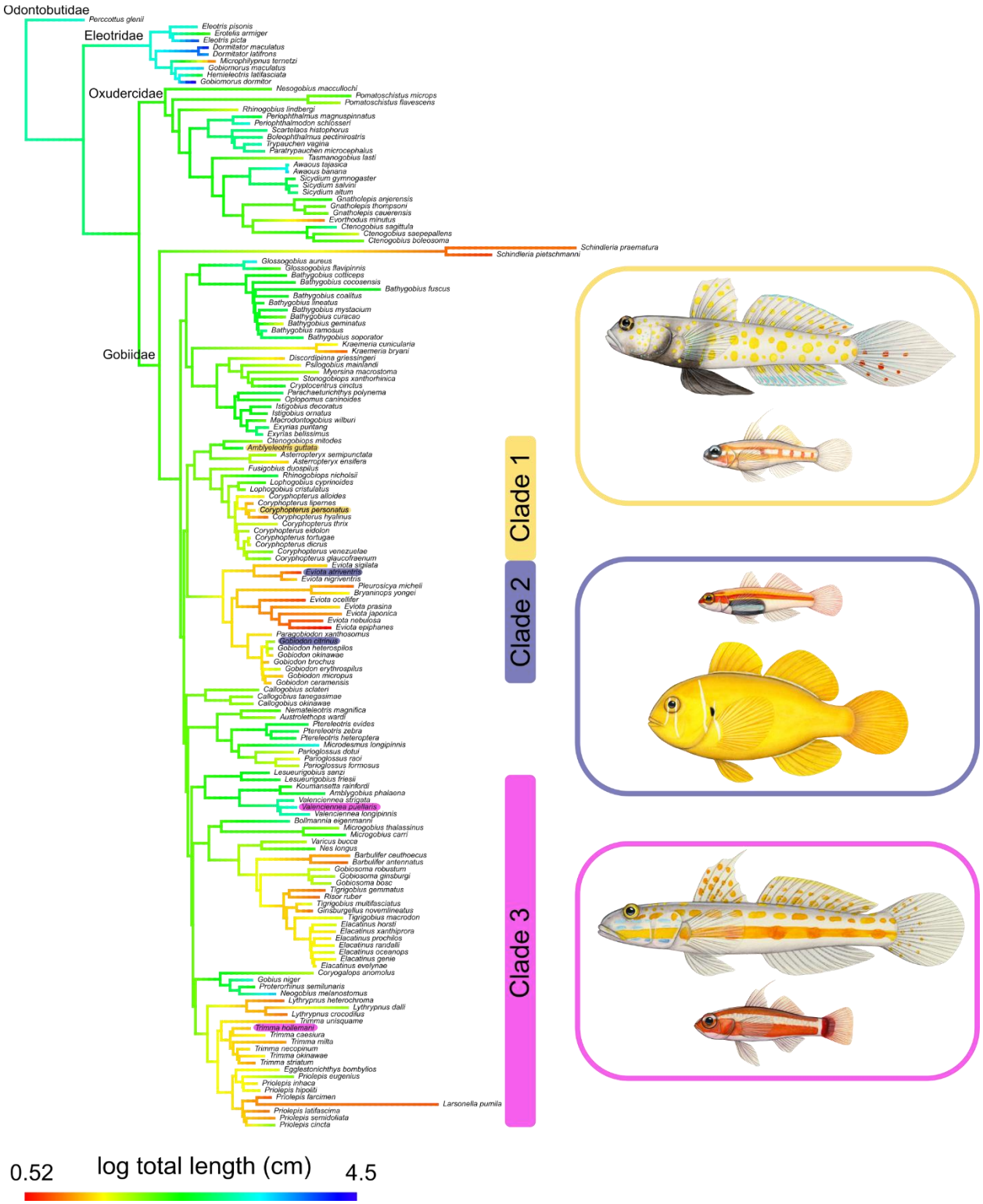


Figure 1. Gobiid phylogeny estimated in RaxML-ng showing the ancestral character reconstruction for total body length. Large sizes are cooler colors and smaller sizes are warmer colors on the phylogeny. Three closely related clades, each containing a large-bodied and a small-bodied gobiid, have been selected for comparative transcriptomic analyses.

RNA-seq processing and de novo assembly

RNA-seq from two tissue types from three individuals of six gobiid species produced a total of 841 M raw paired reads, averaging 24 M per sample (Table S3). Preliminary analyses mapping the raw reads back to the closest available reference genome, *Boleophthalmus pectinirostris*, revealed extremely low percentages of mapped reads for each tissue sample (0.76-11.4%; Table S1). Instead, we utilized a *de novo* assembly approach to assemble the transcriptomes for each of the six gobiid species in our analysis. Filtering of low-quality reads and adaptor sequences removed an average of 3.3% of reads per sample. Normalization within species retained on average 42 M reads (15.6% of input; Table S4).

Separate assemblies were generated for each of the six gobiid species. These species-specific assemblies contained an average of 361,435 genes and 486,134 transcripts per species with an average contig length of 637 bp and N50 of 1150 bp (Table S4). BUSCO scores for the species-specific transcriptome assemblies ranged from 49.3–89.5% with an average of 74.6% using the Vertebrata database and ranged from 47.7–86.1% (71.2% average) using the Actinopterygii database (Table S4). Overall alignment mapping rates from BowTie2 were very high, with an average of 96.71% (Table S4).

Sample correlation matrices based on hierarchical cluster analyses for each tissue sample show high correlation among shared tissue types and species for each of the three clades as well as for the large-bodied and small-bodied species comparisons (Figures S3-S4).

Ortholog identification

The majority of genes were assigned to orthogroups, ranging from 92.7-94.6% with an average of 93.7% (Table S5). From these orthogroups, OrthoFinder identified 970 one-to-one orthologs present in only the three large-bodied species and 665 one-to-one orthologs present in only the three small-bodied species. Of these, only 19 one-to-one orthologs were shared between all six species. For protein identification and functional annotation, these 19 one-to-one orthologs were blasted against the UniProt database[48] and associated genes, biological processes, and molecular functions were recorded, if known (Table S7). These 19 orthologs are associated with various wide-ranging biological processes, such as transcription regulation, signal transduction, protein transport, and mRNA processing (Table S7).

On a cladewise basis, OrthoFinder identified 3,236 one-to-one orthologs in Clade 1, 4,162 one-to-one orthologs in Clade 2, and 2,991 one-to-one orthologs in Clade 3 (Table S6). Because of the small number of one-to-one orthologs present across all six species, we also opted to include all orthogroups, regardless of being a one-to-one ortholog, when comparing the large-bodied species to the small-bodied species in the differential expression analyses comparing all large-bodied to all small-bodied species. This resulted in 7,871 orthogroups with at least one copy shared among all six species for all tissue types. On a tissue-specific basis, 7,292 orthogroups were shared in the muscle tissue of all species and 6,595 were shared in the liver tissues.

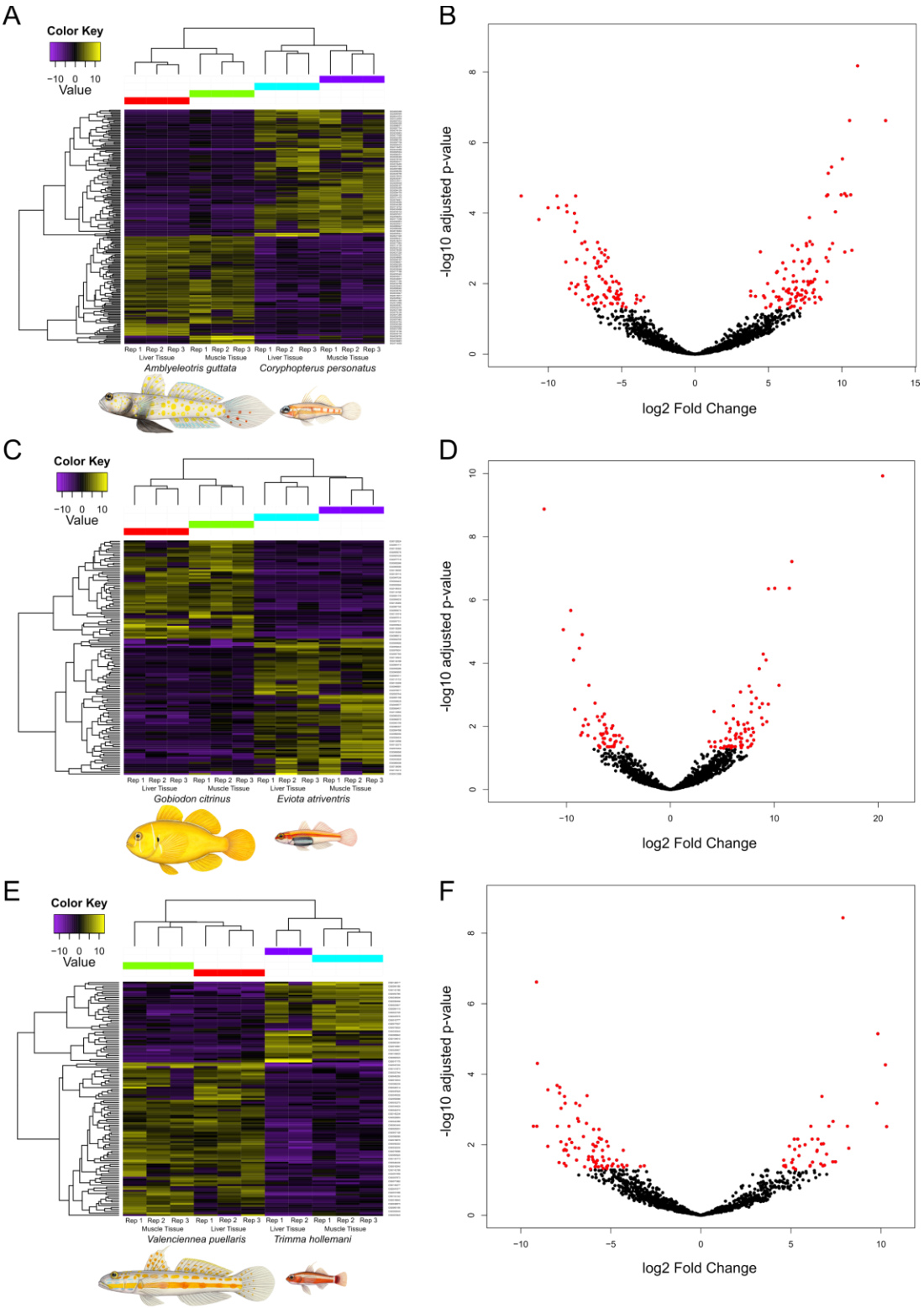


Figure 2. Heatmaps and volcano plots showing differentially expressed one-to-one orthologs for three gobiid clades containing a closely-related large-bodied and small-bodied species. Both upregulated (yellow) and downregulated (purple) orthologs with log fold change values greater than 2 were clustered. Heatmap rows represent individual orthologs and columns represent tissue replicates for each species. Orthologs were clustered according to expression similarity based on Euclidean distances. Illustrations by Julie Johnson.

Identification of differentially expressed orthologs

One-to-one orthologs and orthogroups were used in downstream differential expression analyses using DESeq2[46]. We first analyzed each large and small species pair on a clade-by-clade basis, using all tissue samples. In Clade 1, 196 orthologs were found to be differentially expressed (6.05%), 129 orthologs in Clade 2 were differentially expressed (3.09%), and 125 orthologs in Clade 3 were differentially expressed (4.17%; Table S6). Further analyses on a tissue specific basis for each clade found between 8.85-14.63% of orthogroups differentially expressed among liver tissue and 5.03-7.31% in muscle tissue (Table S6). When comparing all three large species to all three small species, we observe 42.12% of orthologs to be differentially expressed among all tissue types, 15.66% in liver tissue, and 20.64% in muscle tissue (Table S6). Heatmaps and volcano plots showing the upregulated and downregulated differentially expressed orthologs were constructed for all three clade comparisons (Figure 2) and for the large-bodied vs. small-bodied species comparisons (Figure 3).

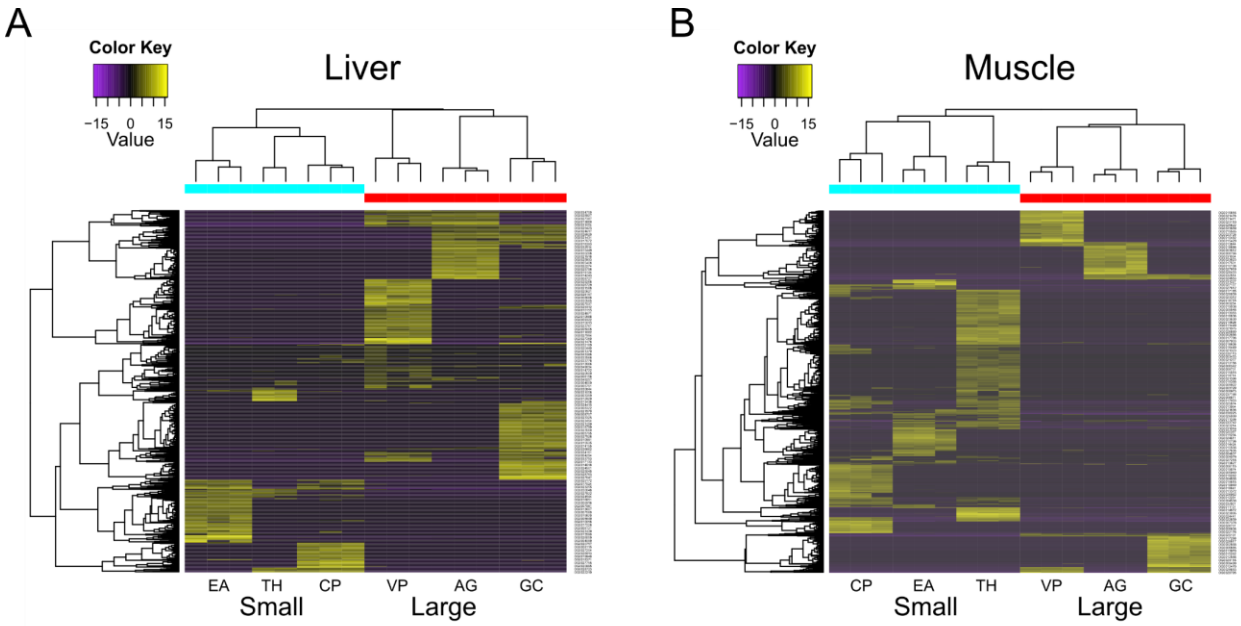


Figure 3. Heatmaps showing differentially expressed one-to-one orthologs between three large- and three small-bodied gobiid species for A) liver tissue samples and B) skeletal muscle tissue samples. Both upregulated (yellow) and downregulated (purple) orthologs with log fold change values greater than 2 were clustered. Heatmap rows represent individual orthologs and columns represent tissue replicates for each species. Orthologs were clustered according to expression similarity based on Euclidean distances. Species abbreviations are AG: *Amblyeleotris guttata*; CP: *Coryphopterus personatus*; EA: *Eviota atriventris*; GC: *Gobiodon citrinus*; TH: *Trimma hollemani*; and VP: *Valenciennea puellaris*.

Biological processes associated with differentially expressed orthologs

To test for convergence on a functional level, we analyzed the differentially expressed one-to-one orthologs for all three small-bodied species compared to all three large-bodied species. Across all three clades, 224 one-to-one orthologs were found to be upregulated in all three small-bodied species and 226 one-to-one orthologs were upregulated in all three large-bodied species (Figure 3). Using the PANTHER database, we linked these differentially expressed orthologs with their respective biological processes. PANTHER identified 196 GO terms associated with biological processes from the upregulated orthologs in large-bodied species, and 179 from the upregulated genes in the small-bodied species. We summarized these GO terms using REVIGO[50] to remove redundant GO terms and found 49 terms associated with biological processes to be shared between all species, 147 unique terms associated with large-bodied species, and 130 unique GO terms associated with small-bodied species (Figure 4). The GO terms of the large and small-bodied species largely overlap with each other, with the majority being associated with metabolism and metabolic processes.

We identified several candidate genes potentially involved in body size evolution and miniaturization. Of the differentially expressed orthologs between the large-bodied and small-bodied species, three orthologs were found to be upregulated in the liver tissues of all small-bodied species and downregulated in all large-bodied species, and ten orthologs are upregulated in the muscle tissue of all small-bodied species and downregulated in all large-bodied species. On the opposite side of the spectrum, we identified seven orthologs upregulated in the muscle tissue and two in the liver tissue of all large-bodied species and subsequently downregulated in all small-bodied species (Table S8). Several of these genes (i.e., *irfd1*, *pcnp*, *mrps9*, *pdcd5*, and *sat1*) are upregulated in the livers of large-bodied species and downregulated in the livers of small-bodied species, but upregulated in the muscle of small-bodied species and downregulated in large-bodied species. However, there are a few uniquely expressed genes in the small-bodied species, which are not expressed in the large-bodied species. These include *zgc:136908*, *NIPSNAP2*, *bzw2*, *chrnd*, and *rab6a* in the muscle tissue and *ybx1* in the liver tissue.

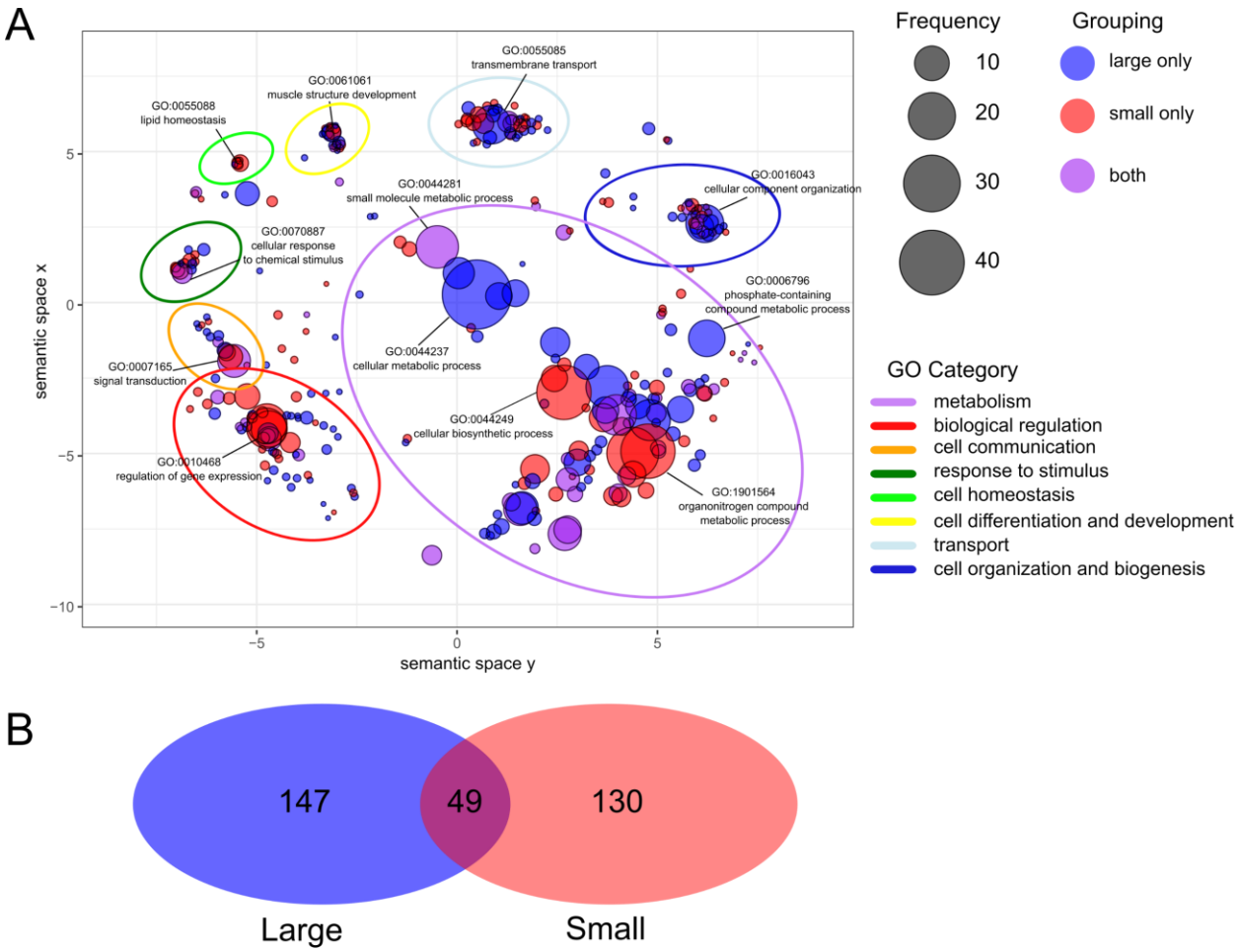


Figure 4. A) Gene ontology (GO) terms grouped using REVIGO semantic clustering. GO terms are the biological processes associated with differentially expressed one-to-one-orthologs. Bubble color indicates which group the GO term is represented in (large-bodied species, small-bodied species, or both) and bubble size indicates the relative frequency of the GO term in the reference Gene Ontology Annotation database[52], where larger bubbles are more general terms and smaller bubbles are more specific. Colored ovals representing broader GO term categories have been drawn around the scatterplot clusters. Names of specific GO terms have been plotted as representatives for each larger cluster. **B)** Venn diagram depicting the number of common and unique GO terms for large-bodied and small-bodied species.

Discussion

This study represents the first comparative approach into identifying genes involved in miniaturization in gobiid fishes, a clade in which body size reduction has evolved multiple independent times. We identified several potential candidate genes involved in size evolution, many of which have been previously associated with cell growth and development. *Ybx1* was significantly upregulated in liver tissues of all three small-bodied species compared to all three large-bodied species. This is a multifunctional protein-coding gene, involved in a variety of biological processes such as differentiation, stress response, embryogenesis, and development[53]. Gene knockdown of long noncoding RNAs in *ybx1* in zebrafish (*Danio rerio*) is shown to disrupt embryological development, leading to larval morphological deformities[53]. A second gene, *bzw2*, which was found to be upregulated in all small-bodied species and downregulated in all large-bodied species, is involved in biological processes such as cell-cell adhesion[54] and may be involved in cell differentiation and nervous system development[55]. Downregulation of *bzw2* has been shown to inhibit tumor cell growth in humans[54]. While there can be no substitute for demonstrating gene function using CRISPR-Cas9 in knockout studies, these candidate genes offer insight into starting points for exploring how function links to phenotype.

We see a number of shared orthologs across a large phylogenetic scale in our analyses of six gobiid species. The most recent common ancestor of these species originated during the Eocene approximately 49.8 million years ago (Ma)[56]. Because they originated from a common ancestor gene and separated via speciation events, it is unsurprising that we find a higher number of one-to-one orthologs present in the more closely related species. For instance, species in Clade 2 (*Eviota atriventris*. and *Gobiodon citrinus*) share 4,162 one-to-one orthologs and diverged approximately 34.9 Ma, while those in Clade 1 (*Amblyeleotris guttata* and *Coryphopterus personatus*) and Clade 3 (*Trimma hollemani* and *Valenciennea puellaris*) both diverged much earlier (46.6 Ma and 45.3 Ma, respectively)[56] and share fewer one-to-one orthologs (3,236 and 2,991, respectively; Table S6). Overall, we find that all six species share 19 one-to-one orthologs which are involved in a variety of biological processes from transcription, DNA replication, and cellular transport (Table S8). These single copy orthologs are highly conserved across a timescale of nearly 50 million years, and therefore may be of great functional importance.

Various explanations for why miniaturization evolved have been put forth, such as smaller species having higher speciation rates or being able to occupy novel niches[4,57]. Gobiid fishes are part of a larger ecological grouping termed "cryptobenthic reef fishes", which includes other small-bodied species (typically < 50 mm) which are visually or behaviorally cryptic, such as blennies (Blennidae), triplefins (Tripterygiidae), basslets (Grammatidae), and cardinalfishes (Apogonidae)[58,59]. Although they are tiny, cryptobenthic fishes represent over half of the fish biodiversity found on coral reefs[58]. This abundance is very likely associated with their small size, as cryptobenthic fishes readily make use of microhabitats throughout the reefs. Gobiids are also one of the most rapidly diversifying clade of fishes, exhibiting high speciation rates throughout their phylogeny[60,61]. Within Gobiidae, dwarfgobies in the genus *Eviota* are the most speciose (with over 100 recognized species) and especially miniscule in body size (10-20 mm in standard length)[62]. This exceptional species richness may be explained by increased opportunities for repeated shifts into various microhabitats[58], which could in turn, drive evolution of small body size within these restricted habitats[58].

Our phylogenetic inference of Gobiidae is notably similar to that of the previous comprehensive phylogeny from McCraney et al.[51]. However, we do observe a large incongruence

with the placement of the genus *Schindleria* being the outgroup to all Gobiidae in this study. *Schindleria* has been previously identified as a potential rogue taxon and difficult to classify morphologically due to its reduced paedomorphic features[18,51,63]. Species in the genus *Schindleria* were previously placed into their own family, Schindleriidae, by Akihito et al.[64] and Thacker[65], but recently reclassified into Gobiidae by Thacker[66] based on molecular phylogenetic data. It is worth noting that the three species of *Schindleria* in the McCraney phylogeny are based on very few markers (four mitochondrial, two nuclear), whereas the two *Schindleria* species in this study were estimated based on 94 and 138 exon markers mined from genomic sequence data. Furthermore, we observe very long branch lengths leading to species in the genus *Schindleria*. This has been previously documented in other phylogenetic estimations[51,63,67], and possibly indicative of high rates of molecular evolution and potentially hindering accurate phylogenetic placement. Thus, further comprehensive phylogenomic estimations will be required to successfully resolve its placement in the Gobioidei phylogeny.

This study generates the first *de novo* transcriptomes for six species of gobiid fishes, including three miniaturized species. Using a comparative transcriptome approach and differential expression analyses, our findings suggest that there is a conserved genetic component in the evolution of body size reduction. While our study identifies several candidate genes associated with miniaturization in gobiid fishes, body size is a highly complex, polygenic trait with multiple genes coordinating together to act on size evolution. As such, we are far from knowing the exact mechanisms involved and more studies are needed to illuminate the underpinnings of this complex trait.

Acknowledgements

The computing for this project was performed with the University of Oklahoma Supercomputing Center for Education & Research (OSCER). We thank OSCER team members Horst Severini and Thang Ha for their valuable technical expertise.

Author Contributions

E.M.T. performed research, analyzed data, and wrote the manuscript; E.M.T. and D.A. designed research

Data Availability

Raw sequence reads will be deposited at the National Center for Biotechnology Information Sequence Read Archive BioProject (number PRJNA1088833). All other supplementary data, including all code and scripts and datasets, will be deposited to Dryad. All other study data are included in the article and supporting information.

References

1. Speakman, J. R. Body size, energy metabolism and lifespan. *J Exp Biol* **208**, 1717–1730 (2005).
2. Brown, J. H. & Maurer, B. A. Macroecology: The Division of Food and Space Among Species on Continents. *Science* **243**, 1145–1150 (1989).
3. Baker, J., Meade, A., Pagel, M. & Venditti, C. Adaptive evolution toward larger size in mammals. *Proceedings of the National Academy of Sciences* **112**, 5093–5098 (2015).
4. Hanken, J. & Wake, D. B. Miniaturization of Body Size: Organismal Consequences and Evolutionary Significance. *Annual Review of Ecology and Systematics* **24**, 501–519 (1993).
5. Yeh, J. The effect of miniaturized body size on skeletal morphology in frogs. *Evolution* **56**, 628–641 (2002).
6. Niven, J. E. & Farris, S. M. Miniaturization of Nervous Systems and Neurons. *Current Biology* **22**, R323–R329 (2012).
7. Kottelat, M., Britz, R., Hui, T. H. & Witte, K.-E. Paedocypris, a new genus of Southeast Asian cyprinid fish with a remarkable sexual dimorphism, comprises the world's smallest vertebrate. *Proceedings of the Royal Society B: Biological Sciences* **273**, 895–899 (2006).
8. Rüber, L., Kottelat, M., Tan, H. H., Ng, P. K. & Britz, R. Evolution of miniaturization and the phylogenetic position of Paedocypris, comprising the world's smallest vertebrate. *BMC Evolutionary Biology* **7**, 38 (2007).
9. Lee, M. S. Y., Cau, A., Naish, D. & Dyke, G. J. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014).
10. Pietsch, T. W. Dimorphism, Parasitism and Sex: Reproductive Strategies among Deepsea Ceratioid Anglerfishes. *Copeia* **1976**, 781–793 (1976).
11. Xu, H. *et al.* Comparative Genomics Sheds Light on the Convergent Evolution of Miniaturized Wasps. *Molecular Biology and Evolution* **38**, 5539–5554 (2021).
12. Brown, B. V. Small Size No Protection for Acrobat Ants: World's Smallest Fly is a Parasitic Phorid (Diptera: Phoridae). *Annals of the Entomological Society of America* **105**, 550–554 (2012).
13. Rowe, T. B. 2.01 - The Emergence of Mammals. in *Evolution of Nervous Systems (Second Edition)* (ed. Kaas, J. H.) 1–52 (Academic Press, Oxford, 2017). doi:10.1016/B978-0-12-804042-3.00029-4.
14. Ocampo, D., Barrantes, G. & Uy, J. A. C. Morphological adaptations for relatively larger brains in hummingbird skulls. *Ecology and Evolution* **8**, 10482–10488 (2018).
15. Glaw, F. *et al.* Extreme miniaturization of a new amniote vertebrate and insights into the evolution of genital size in chameleons. *Sci Rep* **11**, 2522 (2021).
16. Polilov, A. A. Small Is Beautiful: Features of the Smallest Insects and Limits to Miniaturization. *Annual Review of Entomology* **60**, 103–121 (2015).
17. Weitzman, S. H. & Vari, R. P. Miniaturization in South American freshwater fishes; an overview and discussion. (1988).
18. Johnson, D. G. & Brothers, E. B. Schindleria: A Paedomorphic Goby (Teleostei: Aobioidei). *Bulletin of Marine Science* **52**, 441–471 (1993).
19. Froese, R. & Pauly, D. FishBase. (2023).
20. Goldsworthy, N. C. *et al.* Life-history constraints, short adult life span and reproductive strategies in coral reef gobies of the genus Trimma. *Journal of Fish Biology* **101**, 996–1007 (2022).

21. Greenfield, D. W. & Winterbottom, R. A key to the dwarfgoby species (Teleostei: Gobiidae: Eviota) described between 1871 and 2016. *Journal of the Ocean Science Foundation* **24**, 35–90 (2016).
22. Tian, C. *et al.* Identification of differentially expressed genes associated with differential body size in mandarin fish (Siniperca chuatsi). *Genetica* **144**, 445–455 (2016).
23. Wang, N., Wang, R., Wang, R. & Chen, S. Transcriptomics analysis revealing candidate networks and genes for the body size sexual dimorphism of Chinese tongue sole (Cynoglossus semilaevis). *Funct Integr Genomics* **18**, 327–339 (2018).
24. Jiang, W., Guo, Y., Yang, K., Shi, Q. & Yang, J. Insights into Body Size Evolution: A Comparative Transcriptome Study on Three Species of Asian Sisoridae Catfish. *International Journal of Molecular Sciences* **20**, 944 (2019).
25. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
26. Hughes, L. C. *et al.* Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Mol Ecol Resour* **21**, 816–833 (2021).
27. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution* **35**, 2582–2584 (2018).
28. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).
29. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
30. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
31. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
32. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00169.x (2012).
33. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2023).
34. Andrews, S. FastQC: a quality control tool for high throughput sequence data.
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. White, O. W., Reyes-Betancort, A., Carine, M. A. & Chapman, M. A. Comparative transcriptomics and gene expression divergence associated with homoploid hybrid speciation in *Argyranthemum*. *G3: Genes, Genomes, Genetics* **13**, jkad158 (2023).
37. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol* **29**, 644–652 (2011).
38. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
39. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
41. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* **8**, 10.1038/nprot.2013.084 (2013).
42. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
43. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
45. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
47. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M. & Bairoch, A. UniProtKB/Swiss-Prot. *Methods Mol Biol* **406**, 89–112 (2007).
48. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).
49. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* **31**, 8–22 (2022).
50. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).
51. McCraney, W. T., Thacker, C. E. & Alfaro, M. E. Supermatrix phylogeny resolves goby lineages and reveals unstable root of Gobiaria. *Molecular Phylogenetics and Evolution* **151**, 106862 (2020).
52. Camon, E. *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* **32**, D262–D266 (2004).
53. Huang, C., Zhu, B., Leng, D., Ge, W. & Zhang, X. D. Long noncoding RNAs implicated in embryonic development in Ybx1 knockout zebrafish. *FEBS Open Bio* **11**, 1259–1276 (2021).
54. Cheng, D.-D. *et al.* Downregulation of BZW2 inhibits osteosarcoma cell growth by inactivating the Akt/mTOR signaling pathway. *Oncology Reports* **38**, 2116–2122 (2017).
55. National Center for Biotechnology Information (NCBI).
56. Thacker, C. E. Biogeography of goby lineages (Gobiiformes: Gobioidei): origin, invasions and extinction throughout the Cenozoic. *Journal of Biogeography* **42**, 1615–1625 (2015).
57. Arbour, J. H. & Stanchak, K. E. The little fishes that could: smaller fishes demonstrate slow body size evolution but faster speciation in the family Percidae. *Biological Journal of the Linnean Society* **134**, 851–866 (2021).
58. Brandl, S. J., Goatley, C. H. R., Bellwood, D. R. & Tornabene, L. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. *Biological Reviews* **93**, 1846–1873 (2018).
59. Depczynski, M. & Bellwood, D. The role of cryptobenthic reef fishes in coral reef trophodynamics. *Mar. Ecol. Prog. Ser.* **256**, 183–191 (2003).
60. Thacker, C. E. Species and shape diversification are inversely correlated among gobies and cardinalfishes (Teleostei: Gobiiformes). *Org Divers Evol* **14**, 419–436 (2014).

61. Near, T. J. *et al.* Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences* **110**, 12738–12743 (2013).
62. Tornabene, L. *et al.* Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). *Molecular Phylogenetics and Evolution* **66**, 391–400 (2013).
63. Agorreta, A. *et al.* Molecular phylogenetics of Gobioidei and phylogenetic placement of European gobies. *Molecular Phylogenetics and Evolution* **69**, 619–633 (2013).
64. Akihito *et al.* Evolutionary aspects of gobioid fishes based upon a phylogenetic analysis of mitochondrial cytochrome *b* genes. *Gene* **259**, 5–15 (2000).
65. Thacker, C. E. Molecular phylogeny of the gobioid fishes (Teleostei: Perciformes: Gobioidei). *Molecular Phylogenetics and Evolution* **26**, 354–368 (2003).
66. Thacker, C. Phylogeny of Gobioidei and Placement within Acanthomorpha, with a New Classification and Investigation of Diversification and Character Evolution. *cope* **2009**, 93–104 (2009).
67. Tornabene, L., Deis, B. & Erdmann, M. V. Evaluating the phylogenetic position of the goby genus Kelloggella (Teleostei: Gobiidae), with notes on osteology of the genus and description of a new species from Niue in the South Central Pacific Ocean. *Zoological Journal of the Linnean Society* **183**, 143–162 (2018).

Appendix 1

Supplemental Materials to Chapter 1:

The impact of paleoclimatic changes on body size evolution in marine fishes

Published in: *Proceedings of the National Academy of Sciences*, 11 July 2022

https://doi.org/10.1073/pnas.2122486119

Emily M. Troyer[1,2], Ricardo Betancur-R.[1], Lily C. Hughes[3], Mark Westneat[3], Giorgio Carnevale[4], William T. White[5], John J. Pogonoski[5], James C. Tyler[6], Carole C. Baldwin[7], Guillermo Ortí[8], Andrew Brinkworth[9], Julien Clavel[10] & Dahiana Arcila[1,2*]

[1] *Department of Biology, University of Oklahoma, 730 Van Vleet Oval, Richards Hall Norman, OK 73019, USA.*

[2] *Department of Ichthyology, Sam Noble Oklahoma Museum of Natural History, 2401 Chautauqua Avenue, Norman, OK 73072, USA.*

[3] *University of Chicago, Department of Organismal Biology and Anatomy, and Committee on Evolutionary Biology, 1027 E. 57th St, Chicago IL, 60637, USA.*

[4] *Università degli Studi di Torino, Dipartimento di Scienze della Terra, Torino, Italy.*

[5] *CSIRO Australian National Fish Collection, National Research Collections Australia, Castray Esplanade, Hobart, Tasmania, 7001, Australia.*

[6] *Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, 10th & Constitution Ave. NW, Washington, DC 20560, USA.*

[7] *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th & Constitution Ave. NW, Washington, DC 20560, USA.*

[8] *The George Washington University, Department of Biological Sciences, Washington, D.C. 20052, USA.*

[9] *Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AZ, UK.*

[10] *Univ Lyon, Université Claude Bernard Lyon 1, CNRS, ENTPE, UMR 5023 LEHNA, F-69622, Villeurbanne, France.*

Appendix 1 Includes:

Supplemental Materials and Methods
Figures S1 to S22
Tables S1 to S14
SI References

Other supplementary materials for this manuscript include the following:

Raw sequence reads for >1100 exon markers for all newly generated taxa (145 individuals representing 135 species) in this analysis. Uploaded to NCBI Sequence Read Archive under BioProject number PRJNA767646:

Dataset S1

A compressed folder containing all data files is available from Dryad (https://doi.org/10.5061/dryad.z34tmpgfw).

These include:

Datasets S2-S13

Supplemental Materials and Methods:

Taxonomic sampling and genomic data. We generated new genomic data from tissue samples extracted from museum voucher specimens (Dataset S1, Table S13) for 131 species of the order Tetraodontiformes and four species in its living sister clade, Lophiiformes (1). We shipped DNA extractions to Arbor Biosciences for library preparation and target enrichment. Libraries for all samples were processed at the sequencing facilities at the University of Chicago (https://fgf.uchicago.edu). Sequencing of pair end 100 bp reads was completed on a HiSeq 4000 with a total of 192 samples multiplexed per lane, including samples for other projects not listed here. Raw reads for newly sequenced exon-capture data are archived on NCBI Sequence Read Archive under BioProject number PRJNA767646 (Dataset S1). Target capture probes were based on a set of 1,105 single-copy nuclear exon markers (2, 3). Several of these 1,105 loci include "legacy" markers, which are popular markers within fish phylogenetics (4). One of the 131 newly sequenced species (*Rhinecanthus verrucosus*) was excluded due to low capture efficiency, leaving 130 newly sequenced taxa for use in downstream analyses. To increase taxonomic sampling, exon markers from four additional tetraodontiform species were mined from previously published genomes and transcriptomes (2) (Dataset S2). In addition, a total of 51 tetraodontiform species and one outgroup (Perciformes: *Antigonia capros*) were downloaded from NCBI; these include sequences from one mitochondrial and 15 nuclear markers (1, 3–5) (Dataset S2). The final molecular matrix includes a total of 185 out of *ca.* 450 extant tetraodontiforms (~41.1%) representing all 10 extant families.

Alignment and quality control. Sequences for each exon were aligned with MACSE v 2.03 (6) after cleaning out potentially non-homologous fragments with the cleanNonHomologousSequences option. Separate alignments were conducted, including the target exon in MAFFT v. 7.427 (7). All alignments were visually inspected to adjust the reading frames, remove poor-quality reads, and correct misaligned sections in Geneious Prime v. 2020.1.1 (8). Alignment summary statistics, such as the percentage of missing data, GC content, proportion of variable sites, and alignment length was assessed using the python package AMAS (9) (Dataset S8).

Given the pervasiveness of contamination in phylogenomic datasets, quality control analyses consisted of visual inspection of individual gene trees estimated in IQTREE v.1.6.12 (10) and manual inspection of individual gene alignments, following the pipeline implemented by Arcila et al. (11). Additional steps of quality control included cross-validating species identifications of all newly sequenced tissue samples by blasting the mtDNA-COI barcode region against the Barcode of Life Data System (BOLD) repositories (12) using the 'bold_identification' python script (13) (Dataset S10). After these quality control steps, two of the newly sequenced exons were deleted from the analysis due to high levels of missing data, leaving a total of 1,103 loci.

Phylogenomic inference. We inferred phylogenetic trees and support values using maximum likelihood with IQTREE v.1.6.12 (10) and a multi-species coalescent approach in ASTRAL-III (14) based on IQTREE gene trees. To account for the effects of missing data in our dataset, we conducted two concatenation-based maximum likelihood (ML) analyses: one including all newly

sequenced taxa (134 tetraodontiform species, 47% missing data overall; Dataset S6), and a second analysis that excluded taxa with more than 65% missing data (102 tetraodontiform species, < 33% missing data overall; Dataset S7) for all 1,103 exon markers. Because the topology and branch lengths are largely in agreement between the two analyses (*SI Appendix:* Figs. S2, S3), all downstream phylogenetic analyses use the complete dataset (134 tetraodontiform species and 1,103 exon markers). We used PartitionFinder2 (15) to infer the best partitioning scheme for the concatenated datasets, beginning with a total of 3,309 *a priori* partitions (three codon positions for each of the 1,103 loci). From these initial partitions, a best-fit scheme of 719 partitions was identified and used as input for all ML phylogenetic analyses. Each of the 719 partitions used a substitution model of either GTR+G or GTR+I+G.

To account for discordance arising from incomplete lineage sorting (ILS), we conducted a multi-species coalescent analysis with multi-locus bootstrapping in ASTRAL-III (14) (Dataset S5) using 1,103 unrooted individual gene trees (partitioned by codon position) inferred from IQTREE (Dataset S4). We estimated branch support for all gene trees in IQTREE using the ultra-fast bootstrapping approach with 1,000 replicates implemented using the '-bnni' option (16). We used bootstrap replicates for each gene as input for ASTRAL, as well as all 1,103 maximum likelihood gene trees.

Integration of fossil and extant species. To combine the fossils and extant tetraodontiform species, we used the morphological matrix of Arcila and Tyler (17), which consists of 210 characters coded for 17 extant and 52 fossil tetraodontiform species plus two additional outgroup taxa (Lophiiformes: *Lophiodes monodi*, and Zeiformes: *Cyttus novaezelandiae*). We combined this morphological matrix with our genomic dataset for a total of 237 tetraodontiform species and seven outgroups. Our analyses use the GTRGAMMA and Mk models with four partitions; three for the molecular sequences (one for each codon position); and one for the morphological dataset.

Phylogenetic uncertainty and total evidence dating using the fossilized birth death process. In addition to the phylogenomic analyses described above, we conducted divergence time estimations under a total evidence, or tip-dating, framework using the Fossilized Birth Death (FBD) model in MrBayes v 3.2.7a (18). To account for topological uncertainty, we assembled 15 largely independent genomic subsets containing *ca.* 50 randomly selected loci (~20,000 bp) subsampled from the complete genomic dataset of 1,103 loci. All subsets overlap in only five "anchor" genes to maintain the same set of species for each subset (19, 20). In addition to genomic data, each subset contains the morphological dataset for fossil taxa. We ran all 15 subsets in MrBayes with tree sampling occurring every 10,000 generations. We used a relaxed clock model with the clock rate prior following a log normal distribution and independent gamma rate (IGR). After 6 months of total runtime, we found that only 5 (of the 15) subsets reached convergence based on estimated sample size (ESS) values close to or above 200. After filtering out analyses that fail to converge, we updated the root prior ages with different distributions and added certain phylogenetic constraints, described below.

Updated subset schemes (root priors and the inclusion/exclusion of Plectocretacicoidea). Because there is not yet consensus on whether the superfamily Plectocretacicoidea should be considered as stem tetraodontiforms and the exclusion of this superfamily has the potential to

drastically affect age estimations (17, 21), we used two different schemes including and excluding the plectocretacicoids for a total of 10 subsets. This superfamily, containing four extinct taxa, are some of the oldest acanthomorph fish fossils and share 14 morphological synapomorphies with Tetraodontiformes (17). The first scheme includes plectocretacicoids in the analysis and places the oldest fossil calibration age at 95 Ma, the minimum age of the oldest plectocretacicoid fossil (*Plectocretacicus clarae*†). The second analysis excludes plectocretacicoids and the oldest fossil calibration is 59 Ma, based on the minimum age of the oldest fossil (*Moclaybalistes danekrus*†) assigned to crown tetraodontiforms.

To test if the tree age prior influences the root age, we assigned three of the five subsets in each scheme a uniform distribution and the remaining two an offset exponential distribution for the tree age prior. For the subset scheme inclusive of Plectocretacicoidea, the uniform distribution has a minimum age of 95.9 Ma and maximum age of 110, and the offset exponential distribution has a minimum age of 95 Ma and mean age of 107.9. In the subset scheme exclusive of Plectocretacicoidea, the uniform distribution has a minimum age of 59 Ma and maximum age of 67.43 Ma, and the offset exponential distribution has a minimum age of 59 Ma and mean age of 61.8 Ma. To examine potential effects of the root prior distribution on the posterior estimations, we ran each distribution scheme (uniform distribution and offset exponential) without data for both fossil schemes (with and without Plectocretacicoidea; Tables S1-S4).

We placed constraints onto extant tetraodontiform families, as their monophyly is not debated, but rather how they are placed at the short inner branches at the base of the tree. In addition, we placed constraints for: Tetraodontoidea (Diodontidae+Tetraodontidae), Balistoidea (Balistidae+Monacanthidae), Triacanthoidea (Triacanthidae+Triacanthodidae), Ostracioidea (Ostraciidae+Aracanidae), and Plectocretacicoidea†. We also included constraints on fossil families and fossil genera for Bolcabalistidae†, Monacanthidae+ Bolcabalistidae†, *Zignoichthys*†+*Iraniplectus*†+Molidae,Triodontidae+Ctenoplectus†, Ostracioidea+*Protobalistum*†+*Spinacanthus*†, Tetraodontoidea+*Balkaria*†. Lastly, we assigned additional constraints for all Tetraodontiformes, all Lophiiformes, all Percomorpha, and a root constraint.

We ran the updated subset schemes in MrBayes until convergence (around 100-150 million generations, depending on the exact subset) and we combined all runs for each subset after discarding 10% of the first tree sampled as burn-in. We evenly sampled 100 trees (i.e. we took every nth tree until 100 trees were obtained) from the posterior distribution of each subset in a given scheme to take into account phylogenetic uncertainty for downstream comparative analyses, for a total of 500 trees. In addition, we constructed a maximum clade credibility (MCC) tree from 10,000 trees evenly sampled from the posterior of all five subsets using TreeAnnotator (22) (Dataset S11, S12).

Trait data. We compiled total length (TL) and standard length (SL) data for most fossil and extant tetraodontiform species in our dataset, targeting five individuals per species when possible (Dataset S3). We obtained length data for extinct species from museum collection databases (e.g., Smithsonian National Museum of Natural History) and published papers, and data for extant species from published papers, museum collections, and FishBase (23). In the case where TL was

the only measurement reported in a dataset, we estimated the equation for SL, for each tetraodontiform family, based on species measurements where both TL and SL were known (Dataset S13). Because of the bias for smaller specimens in museum collections, we omitted any measurements from individuals that were more than 20% smaller than the maximum recorded size listed on FishBase, leaving one to three individuals per species which were averaged to obtain a mean maximum SL per species. We chose maximum SL values as an indicator for how large a species could potentially reach. We excluded three of the extinct tetraodontiform species (*Archaeotetraodon cerrinaferoni*†, *Archaeotetraodon dicarloi*†, and *Ctenoplectus williamsi*†) from the body length analyses due to incomplete or fragmented fossils, which precluded SL from being accurately assessed. We performed all analyses using log transformed values.

Because many species of tetraodontiform fishes are irregularly shaped (i.e. boxfishes, pufferfishes, etc.), we assessed the robustness of using SL as a proxy for body size compared to other measurements such as volume and surface area (SA) in a subset of tetraodontiform species, representing all 10 extant families. Using publicly available computed-tomography (CT) scan data accessed from MorphoSource.org for 27 species of tetraodontiforms in addition to newly collected CT scan data for 14 species (Dataset S9), we created three-dimensional models using the image computing platform Slicer(24). To accurately calculate volume and surface area, we solidified models using the WrapSolidify option in the Segment Editor module and calculated volume (mm^3) and SA (mm^2) using the Segment Statistics module. We measured the SL of the models using the ruler tool. We log-transformed all measurements and performed phylogenetic generalized least squares (PGLS) analyses between SL and volume, and SL and SA. PGLS between SL and volume revealed a strong positive correlation ($p = 0.0005$; *SI Appendix,* Fig. S13). A similar pattern was found for SL and SA ($p = 0.0248$; *SI Appendix*, Fig. S14). Because the resulting measurements were highly correlated to SL across the majority of extant tetraodontiform families, we opted to simply use SL as a proxy for body size. In addition to the simplicity of calculating SL for each extant species in our dataset, SL is easily obtained from fossil taxa, whereas other measurements, such as volume or SA, are impractical to obtain for such species.

Paleotemperature data. We obtained temperature curves that span the nearly 100 Ma evolutionary history of tetraodontiforms from Scotese et al. (25). These authors used oxygen isotope data to reconstruct past global average ocean temperatures and sea surface temperatures between tropical latitudes (15°N-15°S). Given the broad distributions (circumtropical/subtropical latitudes) and habitat preferences (mainly marine shallow-water, but also deep-water, estuarine, and freshwater dwellers) of tetraodontiforms, these two contrasting temperature curves most accurately capture the spectrum of past environmental affinities in this group.

Evolutionary model fitting. We conducted model fitting analyses in R version 4.0.2 (26). We fitted models of continuous character evolution using the R package 'geiger' and the 'fitContinuous' function. We selected two standard models of evolution: a Brownian motion (BM) model, where body size is expected to be randomly dispersed through time and an Ornstein-Uhlenbeck (OU) model, which is similar to a BM model, but with a tendency for a trait to move towards a central optimum. In addition, we fitted evolutionary rate models which include an early burst (EB) model where the rate of trait evolution is initially high, then decreases exponentially

over time and a rate trend model which fits a linear trend in evolutionary rates through time (either towards a larger or smaller rate of evolution). To test Cope's rule, we fitted a mean trend (or drift) model, which fits a directional trend component (i.e. towards a larger or smaller body size through time). To more explicitly assess an increase in body size in response to paleoclimatic changes over time (Cope-Bergmann rule), we fitted a climate-dependent OU model (Eq. 1), which is an extension of the classical OU process and has a time-dependent optimum $\theta(t)$ (27).

$$dX(t) = \alpha[\theta(t) - X(t)]dt + \sigma dB(t) \qquad (1)$$

We tested this model using the two temperature curves mentioned above from Scotese et al. (25). We considered the best fitting model to be that with the lowest corrected Akaike information criteria (AICc) score. To account for tree uncertainty, we tested all models with the 500 trees selected from the posterior distribution. Additionally, we account for interspecific variation by incorporating measurement error into our comparative analyses by calculating the standard error of the mean for each species (standard deviation of SL/number of individuals) and incorporating this into the model fitting analyses.

Because there is a global trend towards declining temperatures over the evolutionary history of Tetraodontiformes (i.e. from the late Cretaceous to present day), it can be difficult to decouple the effects of temperature (Bergmann's Rule) from other processes that may be correlated to increased body sizes (Cope's Rule). To disentangle the effects of the long-term cooling trend from the main cyclic fluctuations of temperatures over the past 100 Ma, we decomposed the temperature curves and constructed an alternate "decomposed" OU climate model which uses two additional parameters: the overall trend and the variations around this trend (Fig. S13). To assess which parameter was most important to model fit, we ran three analyses on both temperature curves (GAT and tropical latitudes). The first analysis modeled the two independent components (overall trend and variations) together, the second modeled only the overall trend, and the third modeled only the variations (Fig. S14).

Ecomorphological correlations. To further determine if tetraodontiforms adhere to the Cope-Bergmann rule, we examined patterns of body size in relation to past ocean temperature changes for the two temperature curves mentioned above. We performed ancestral state reconstructions of body sizes for all nodes and mapped these onto the MCC tree using the 'contMap' function in the R package 'phytools' (28). This function maps estimated states at internal nodes using maximum likelihood and the 'fastAnc' function under a Brownian motion model. To take phylogenetic independence into account, we performed a phylogenetic generalized least squares (PGLS) analysis on each temperature curve from Scotese et al. (25). Because a PGLS requires tips as input, we modify our MCC tree by grafting branches with zero lengths onto each node of the MCC tree. These we now consider "tips" which can be assigned the estimated ancestral body size reconstructions for use in the analyses. Using the 'gls' function in the R package 'nlme', we first compared a set of evolutionary models including PGLS under a Brownian Motion (PGLS-BM) model, an Ornstein-Uhlenbeck (PGLS-OU) model, and a model where phylogeny is not considered (i.e. Ordinary Least Squares or OLS). Because the 'gls' function requires an input tree, the OLS regressions were run after transforming species covariances by a Pagel's λ value of 0 using the 'lambdaTree' function in the 'phytools' package .

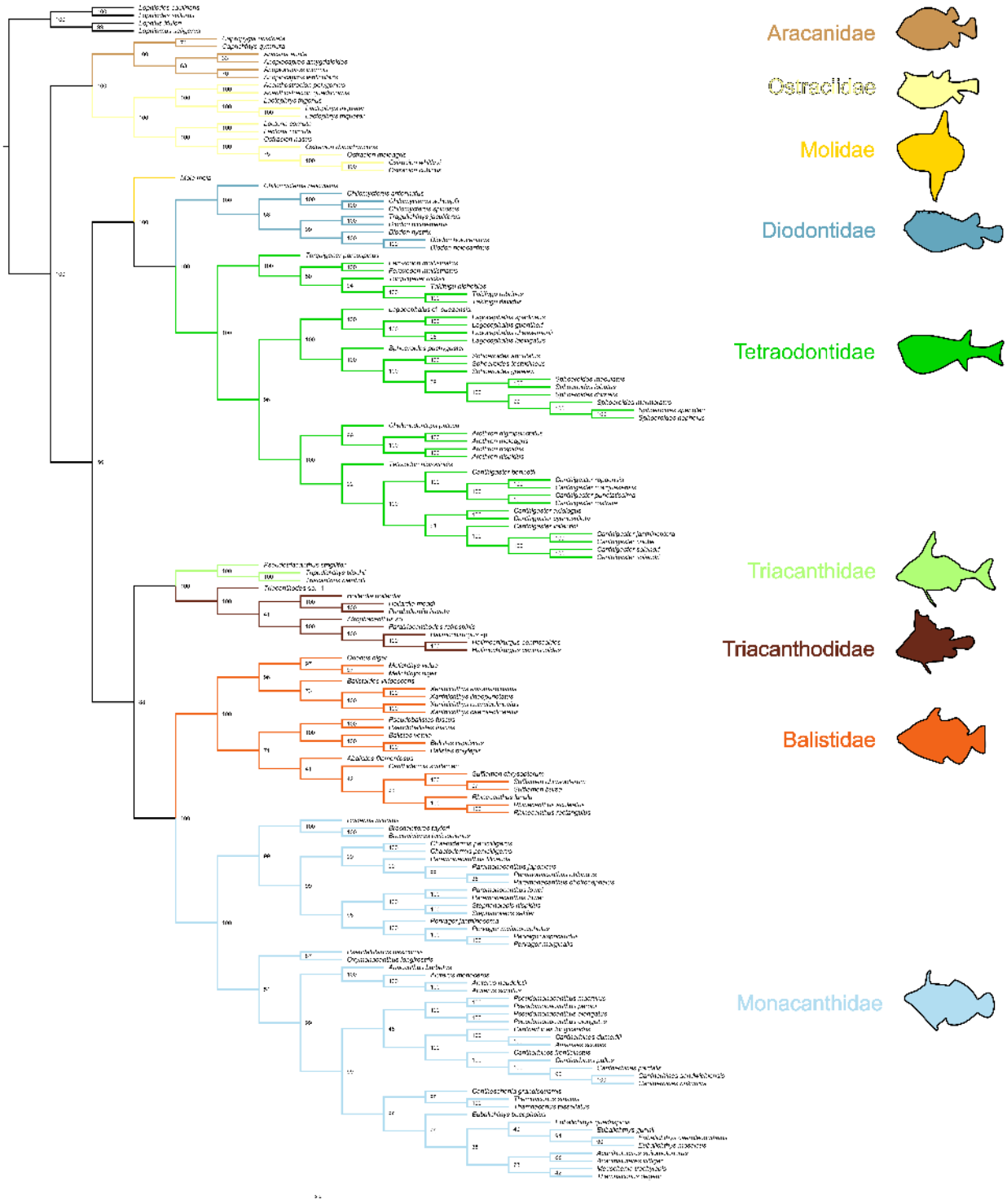


Fig. S1. Phylogeny of Tetraodontiformes based on multi-species coalescent analysis of 1,103 exons and 138 species (134 Tetraodontiformes, 4 outgroup Lophiiformes). Phylogenetic tree inferred with ASTRAL for all newly sequenced taxa and four previously published transcriptomes (3). Colors indicate families. Nodal values indication bootstrap support. Tree is a cladogram and branch lengths do not represent any evolutionary distance.

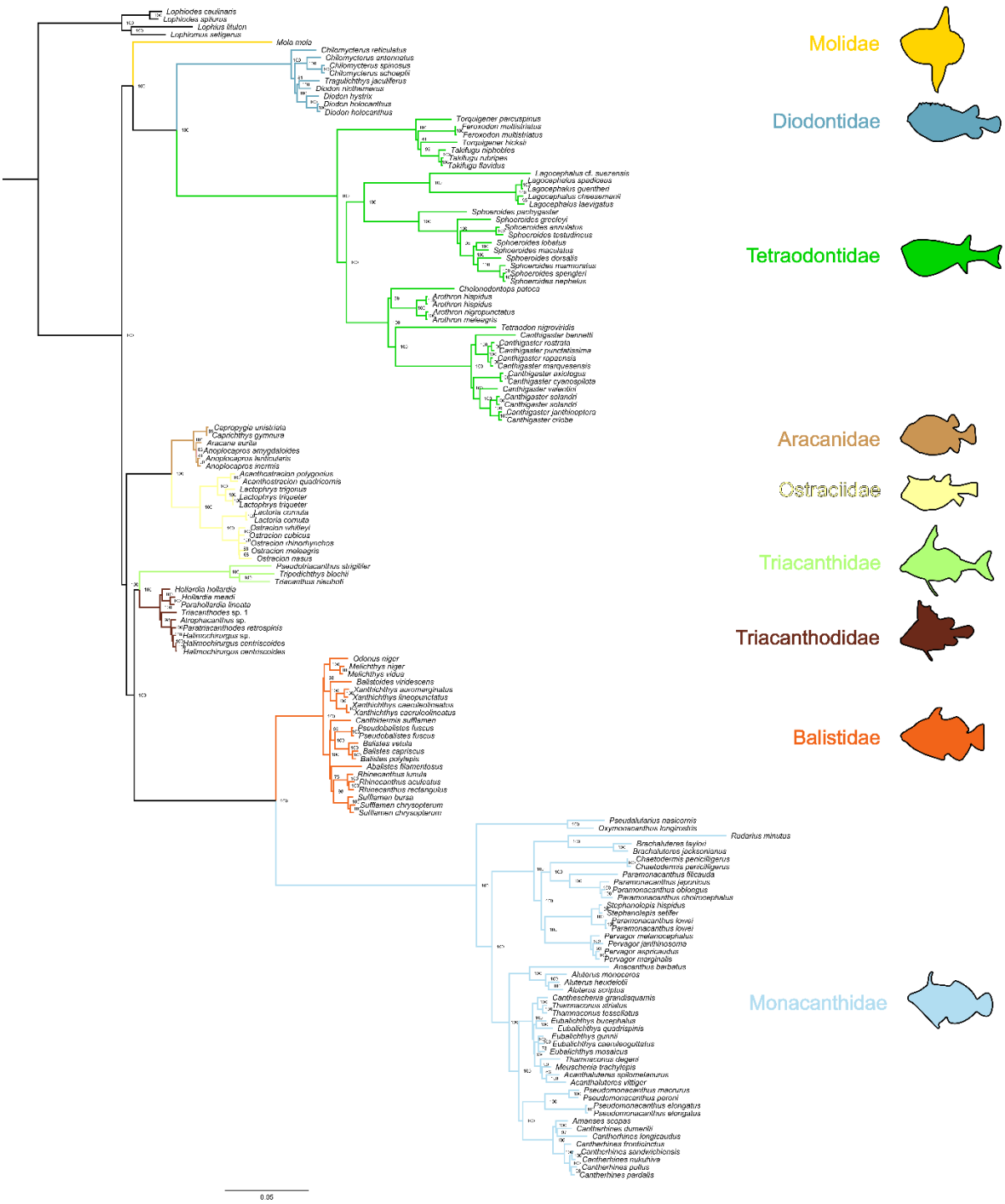


Fig. S2. Phylogeny of Tetraodontiformes based on concatenation analysis of 1,103 exons and 138 species (134 Tetraodontiformes, 4 outgroup Lophiiformes). Phylogenetic tree inferred with IQTREE using the best fit partition scheme identified with PartitionFinder for all newly sequenced taxa and four previously published transcriptomes (3). Colors indicate families. Nodal values indication bootstrap support. Branch lengths are the number of nucleotide substitutions per nucleotide site.

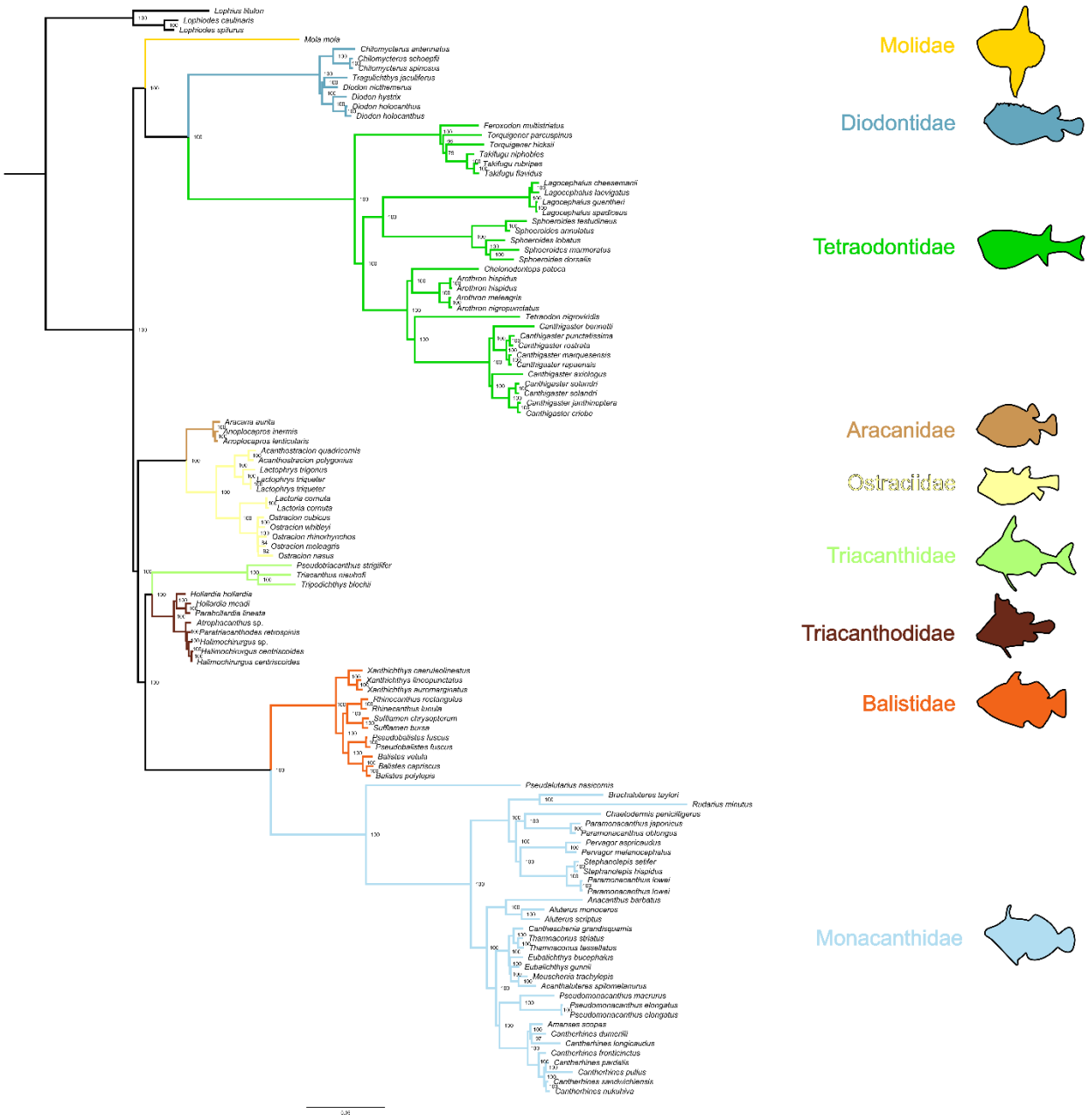


Fig. S3. Phylogeny of Tetraodontiformes, excluding taxa with more than 65% missing data, based on concatenation analysis of 1,103 exons and 105 species (102 Tetraodontiformes, 3 outgroup Lophiiformes). Phylogenetic tree inferred with IQTREE using the best fit partition scheme identified in PartitionFinder for all newly sequenced taxa and four previously published transcriptomes (3). Colors indicate families. Nodal values indication bootstrap support. Branch lengths are the number of nucleotide substitutions per nucleotide site.

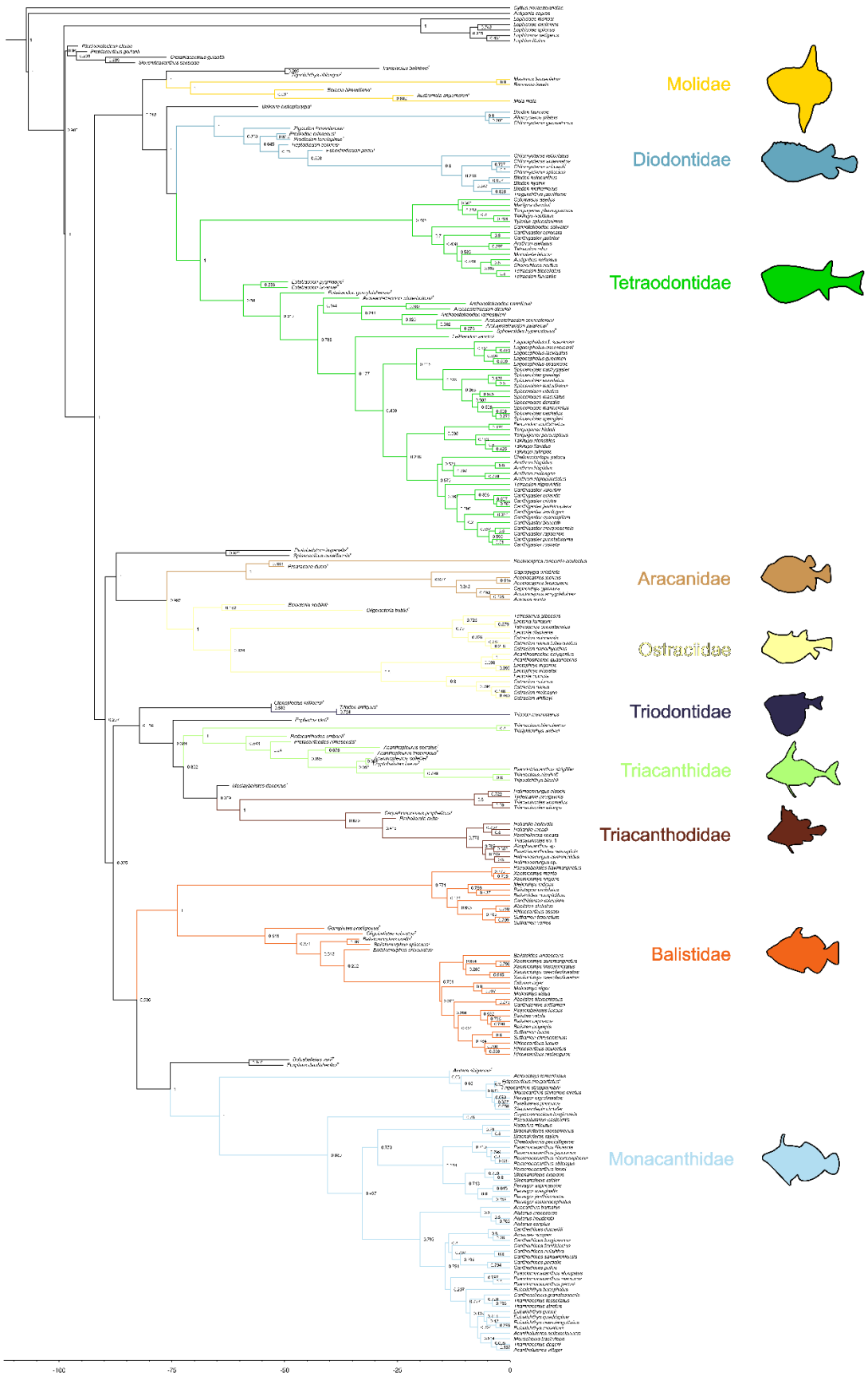

Molidae
Diodontidae
Tetraodontidae
Aracanidae
Ostraciidae
Triodontidae
Triacanthidae
Triacanthodidae
Balistidae
Monacanthidae
-100
-75
-50
-25
0

Fig. S4. Maximum clade credibility (MCC) tree of Tetraodontiformes. Time-calibrated phylogenetic tree using a total-evidence framework based on Bayesian inference of 1,103 exons and 237 tetraodontiforms (52 fossil, 185 extant) and seven outgroups. MCC tree generated from 10,000 trees evenly selected from the posterior distribution of five subsets. Colors indicate families. Scale bar represents millions of years. Posterior probability values are given for each node.

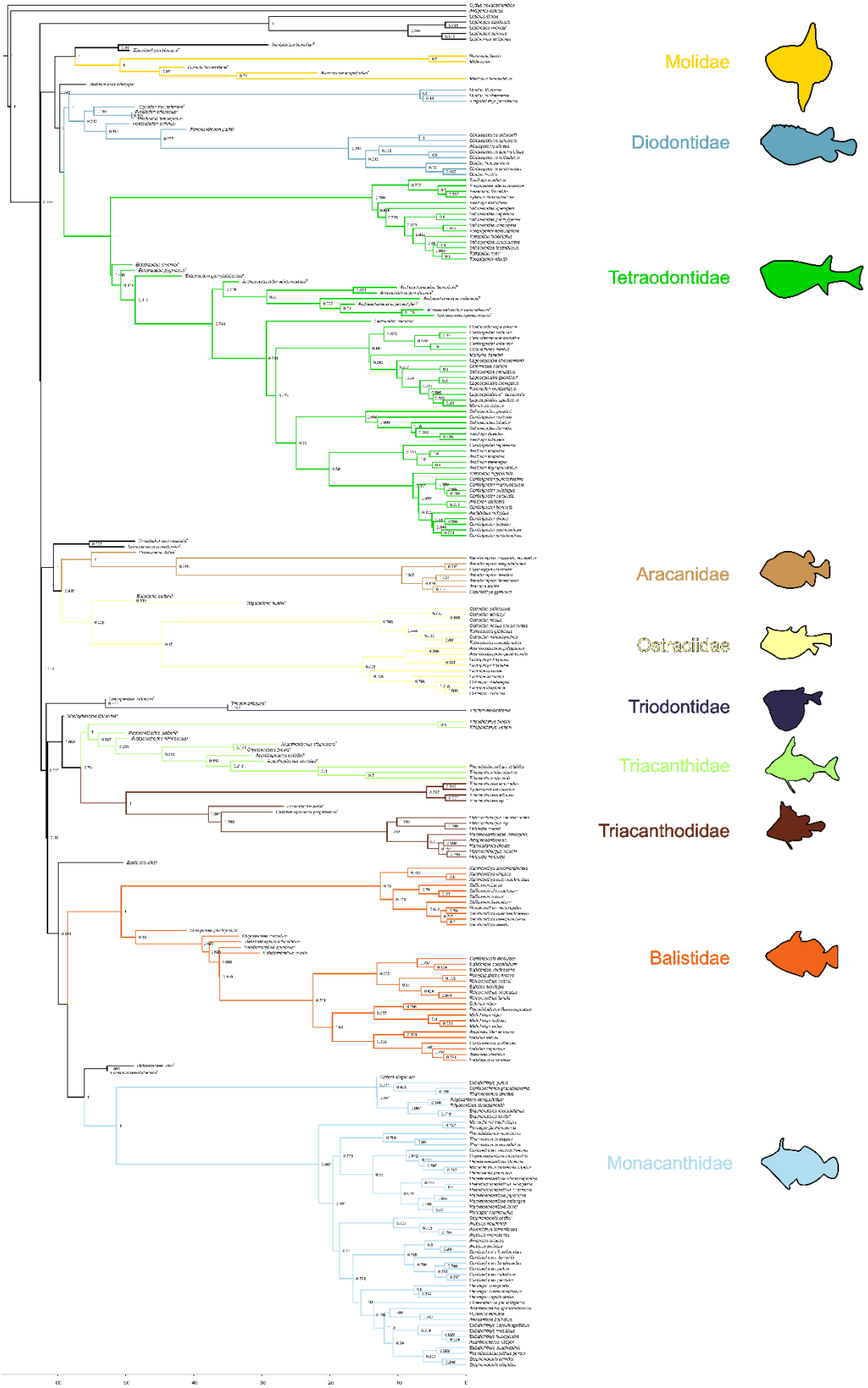

Molidae
Diodontidae
Tetraodontidae
Aracanidae
Ostraciidae
Triodontidae
Triacanthidae
Triacanthodidae
Balistidae
Monacanthidae

Fig. S5. Maximum clade credibility (MCC) tree of Tetraodontiformes, excluding the superfamily Plectocretacicoidea. Time-calibrated phylogenetic tree using a total-evidence framework based on Bayesian inference of 1,103 exons and 233 tetraodontiforms (48 fossil, 185 extant) and seven outgroups. MCC tree generated from 10,000 trees evenly selected from the posterior distribution of five subsets. Colors indicate families. Scale bar represents millions of years. Posterior probability values are given for each node.

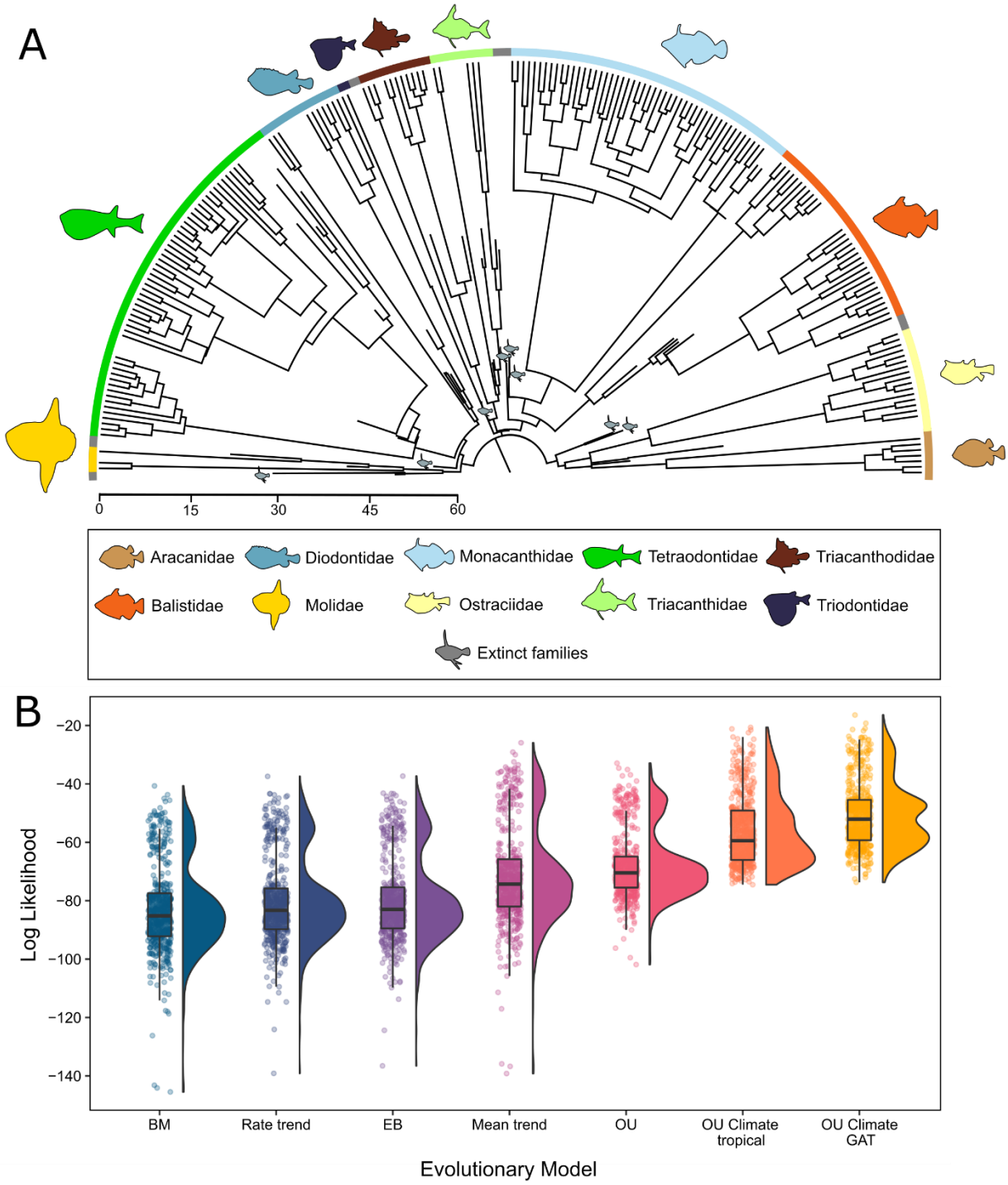

A
0
15
30
45
60
Aracanidae
Diodontidae
Monacanthidae
Tetraodontidae
Triacanthodidae
Balistidae
Molidae
Ostraciidae
Triacanthidae
Triodontidae
Extinct families
B
Log Likelihood
−20
−40
−60
−80
−100
−120
−140
BM
Rate trend
EB
Mean trend
OU
OU Climate tropical
OU Climate GAT
Evolutionary Model

Fig. S6. Tip-dating tree inferred for tetraodontiforms and evolutionary model fitting results, excluding superfamily Plectocretacicoidea. ***A,*** Maximum clade credibility (MCC) tree derived from a total-evidence dating analysis using the fossilized birth-death model in MrBayes (excluding plectocretacicoids). MCC tree is derived from 10,000 trees evenly sampled from the posterior distribution of five independent subsets. See *SI Appendix*, Fig. S5 for an expanded version of this tree. ***B,*** Raincloud plots (half-violin plots and boxplots) for each model of body size evolution tested, representing the distribution of likelihood scores from 500 trees evenly selected from the posterior distribution of five gene subsets in the Bayesian analysis (see also Table S6). Dots represent raw likelihood score for each of the 500 trees analyzed, for each model. Evolutionary models tested include Early-burst (EB), Brownian motion (BM), rate trend, Ornstein-Uhlenbeck (OU), mean trend, a climate OU model using tropical ocean temperatures, and a climate OU model using global average ocean temperatures (GAT).

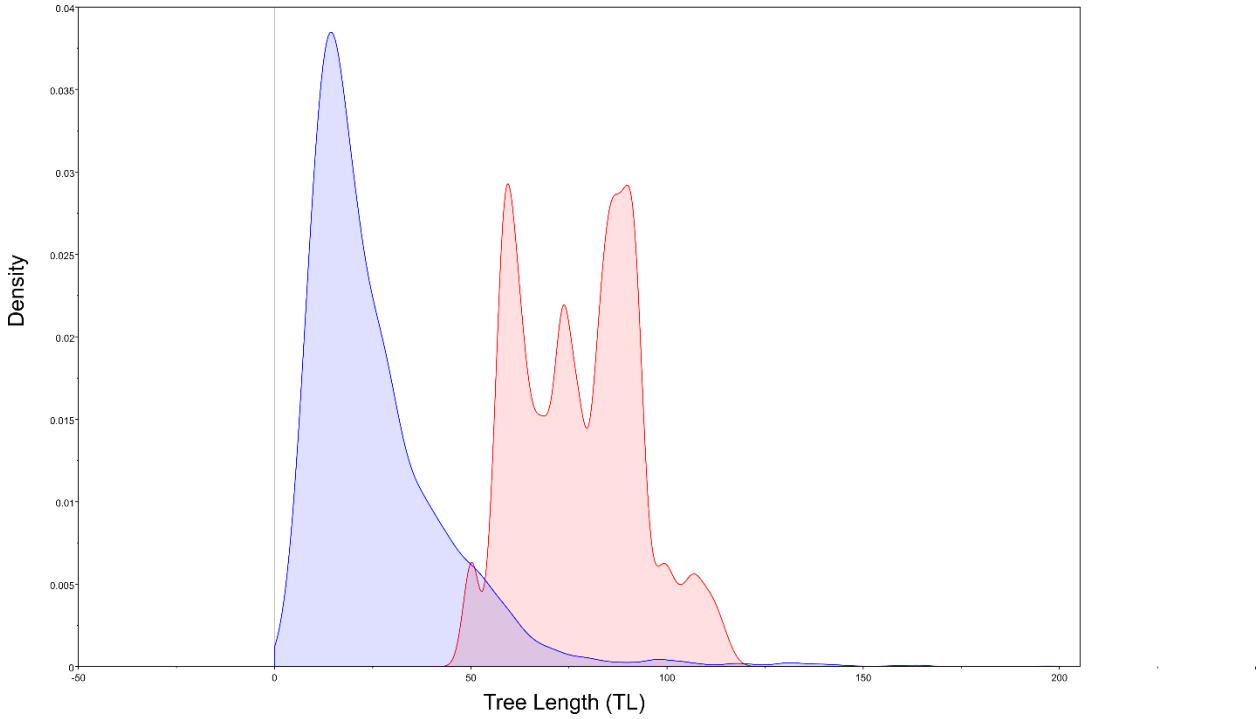


Fig S7. Kernel density plot for the tree length (TL) prior for a subset run with data (blue) and without data (red).

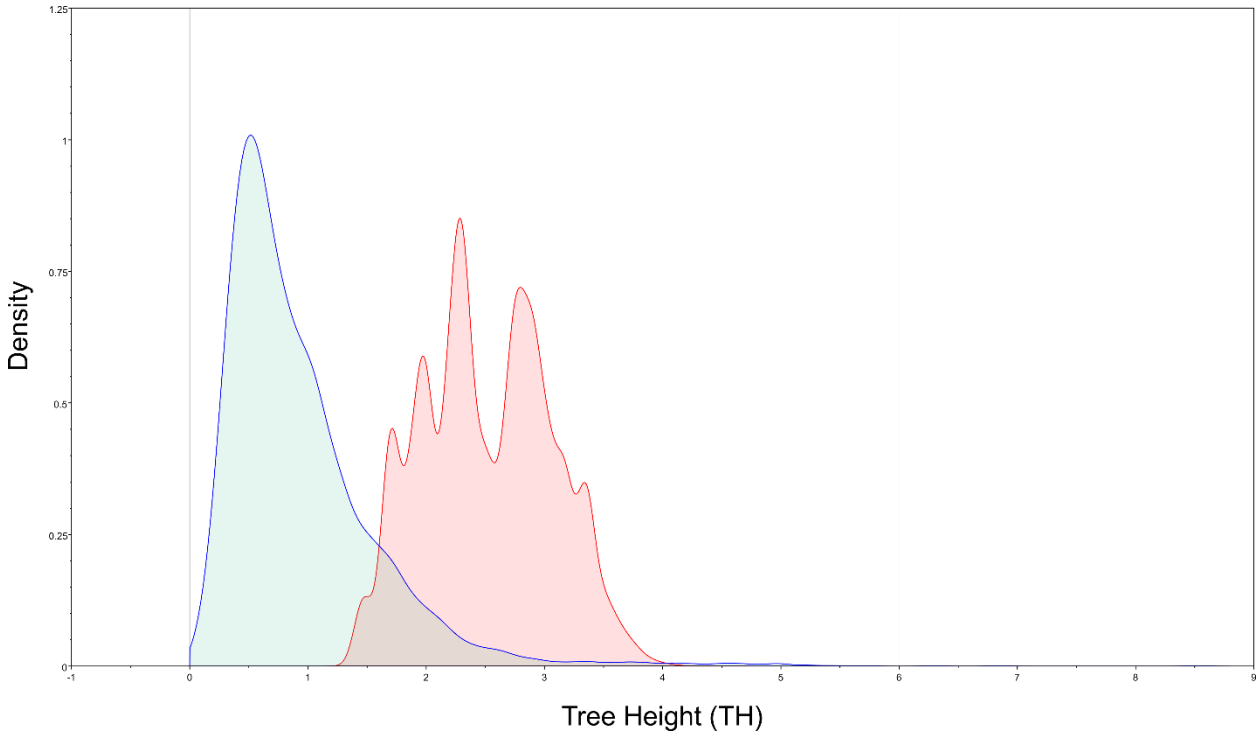


Fig. S8. Kernel density plot of the tree height (TH) prior for subset run with data (blue) and without data (red).

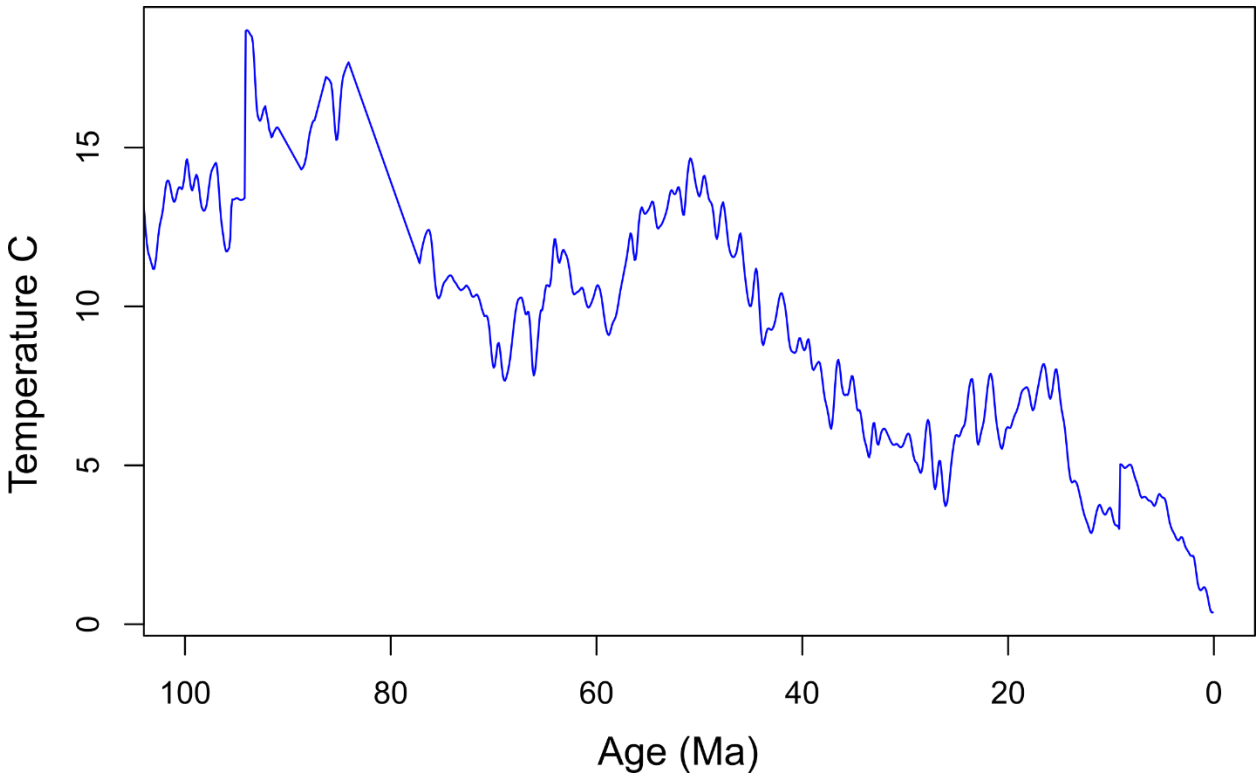


Fig. S9. Deep-sea temperature curve for the past 100 Ma based on data from Cramer et al. (33).

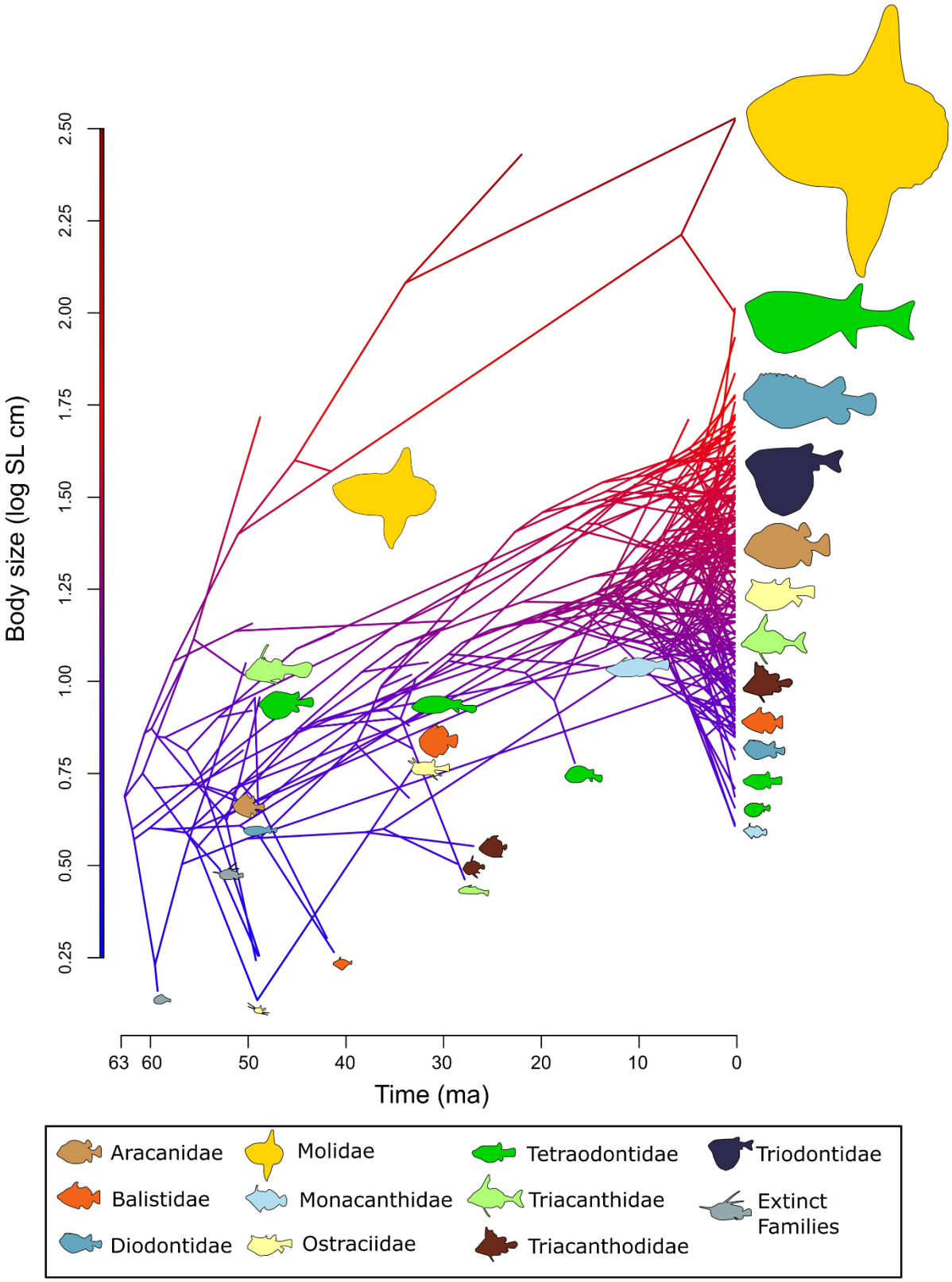

Body size (log SL cm)
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
63
60
50
40
30
20
10
0
Time (ma)
Aracanidae
Balistidae
Diodontidae
Molidae
Monacanthidae
Ostraciidae
Tetraodontidae
Triacanthidae
Triacanthodidae
Triodontidae
Extinct Families

Fig. S10. Evolution of tetraodontiform body size over time, excluding superfamily Plectocretacicoidea. Ancestral reconstruction of body size in tetraodontiforms, as estimated using the R package 'phytools' (28). The log-transformed mean maximum standard length for each species is plotted as a traitgram on the Y axis, with time on the X axis. Fish silhouettes are scaled to represent proportional log body size and colored by family, with extinct families in grey. The estimated ancestral body size of tetraodontiforms is 2-3 times smaller than that of present-day taxa.

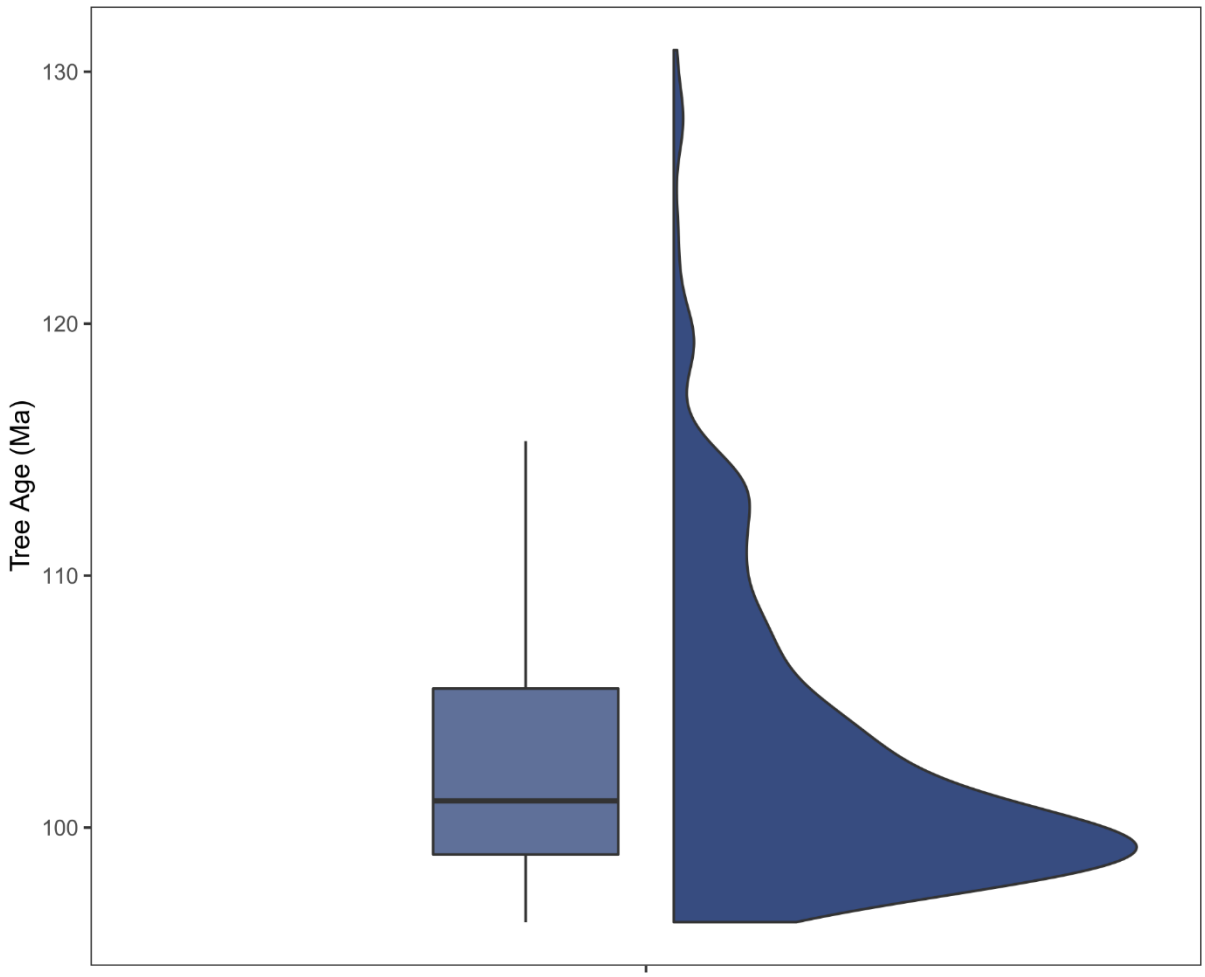


Fig. S11. Distribution plot showing the ages of 500 trees pulled from the posterior distribution. Tree ages range from 96.25 Ma to 130.86 Ma with a mean age of 103.18 Ma.

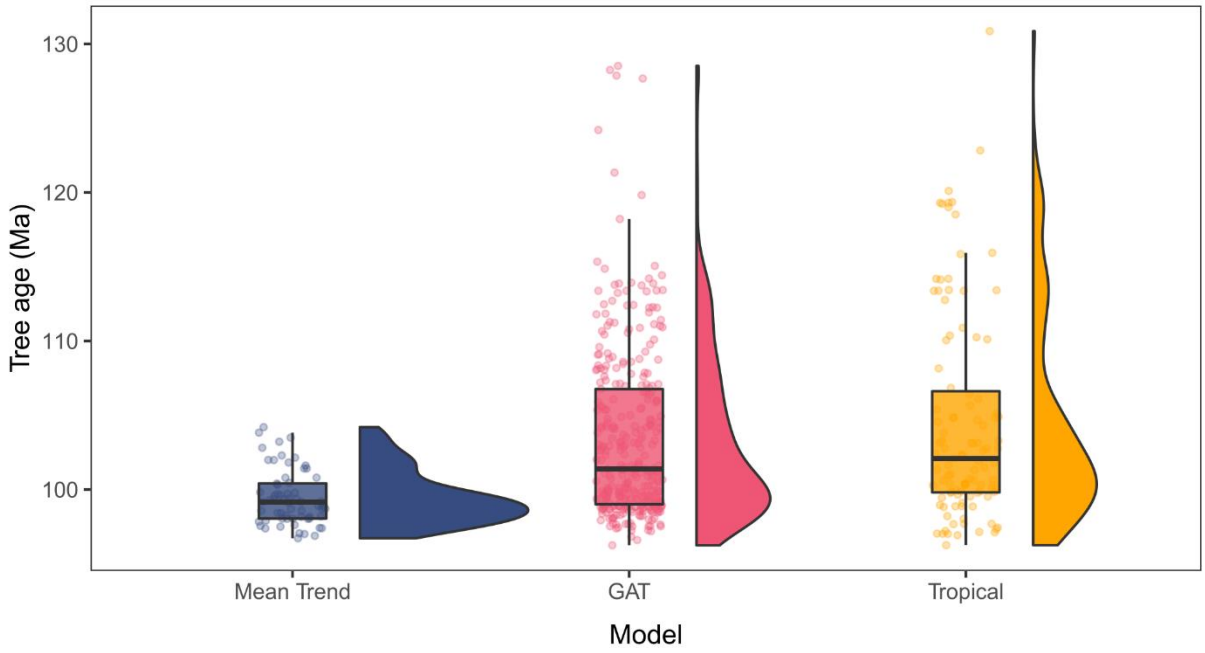


Fig. S12. Raincloud plots (half-violin plots and boxplots) showing the distribution of root ages of the 500 trees pulled from the posterior, separated by best-fit model.

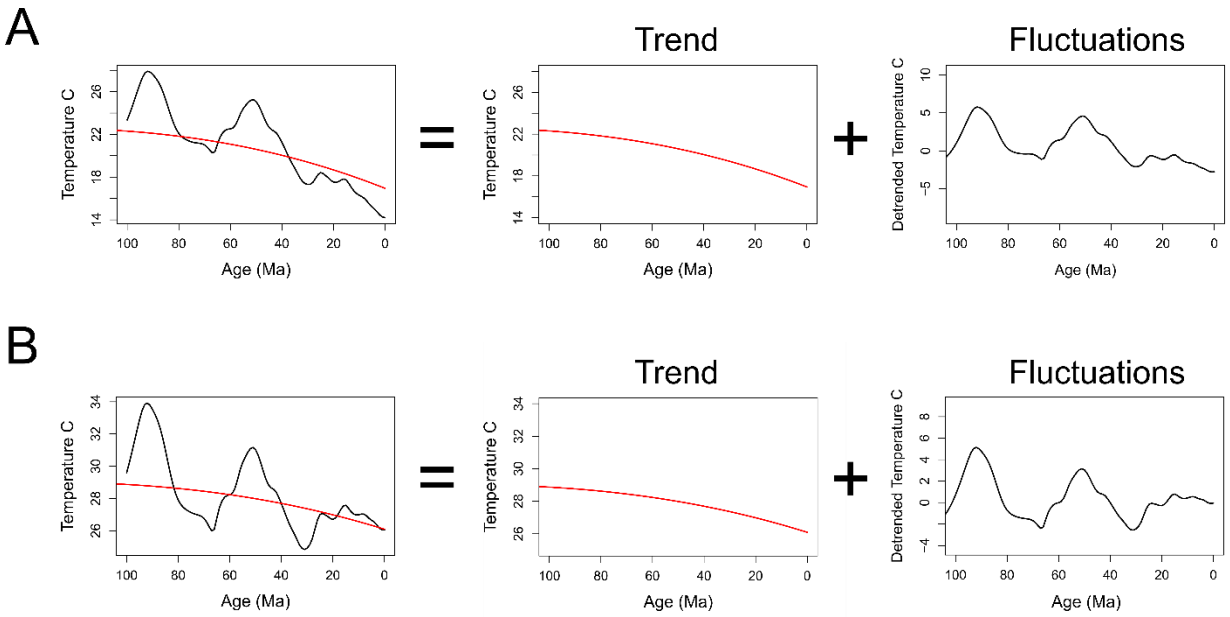
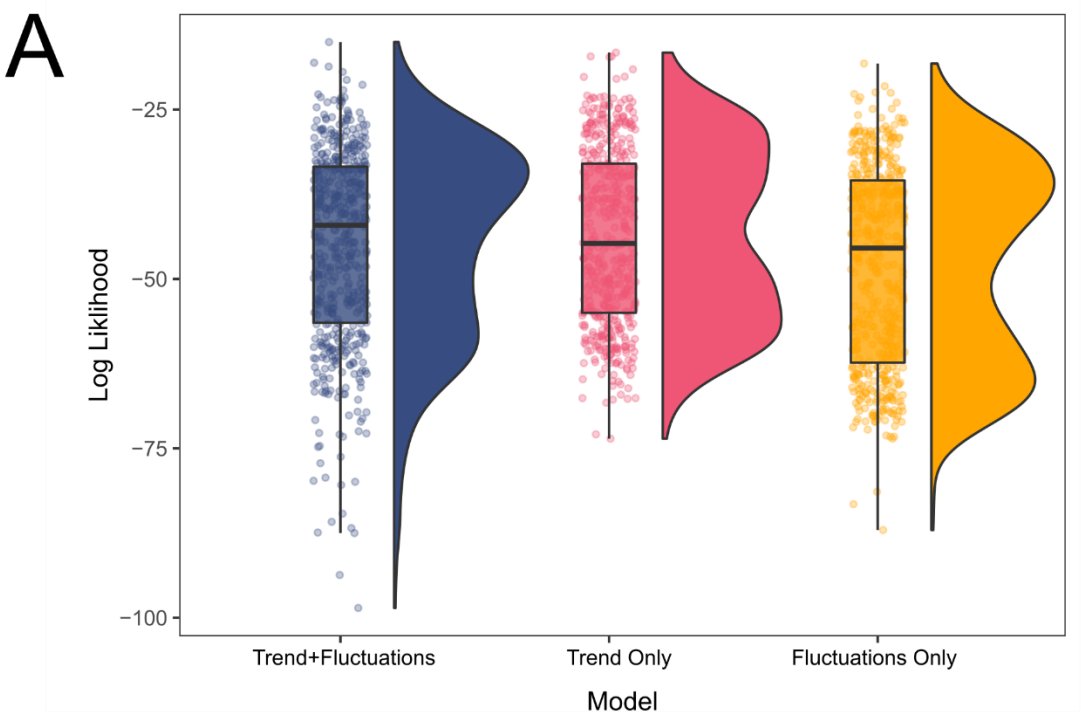


Fig. S13. The temperature curves for the **A)** global average temperature (GAT) and **B)** tropical latitude temperatures through the late Cretaceous to present day decomposed into an overall trend and the fluctuations around this trend. The temperature curve is thus the combination of the two curves.


A
−25
−50
−75
−100
Log Liklihood
Trend+Fluctuations
Trend Only
Fluctuations Only
Model

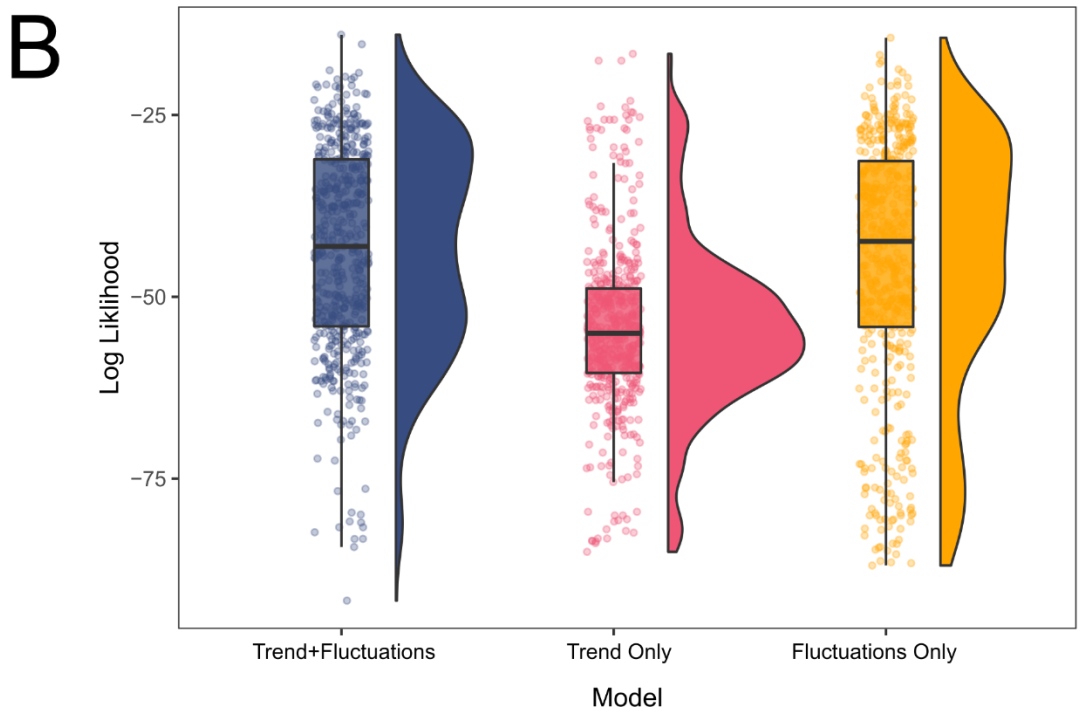

B
−25
−50
−75
Log Liklihood
Trend+Fluctuations
Trend Only
Fluctuations Only
Model

Fig. S14. Raincloud plots (half-violin plots and boxplots) of the decomposed OU climate models showing three tested models: both the overall global trend and the fluctuations around that trend, the trend only, and the fluctuations only. Figure displays the distribution of likelihood scores from 500 trees evenly selected from the posterior distribution of five independent gene subsets in the Bayesian analyses for **A)** the global average temperature (GAT) curve and **B)** the tropical latitude temperature curve.

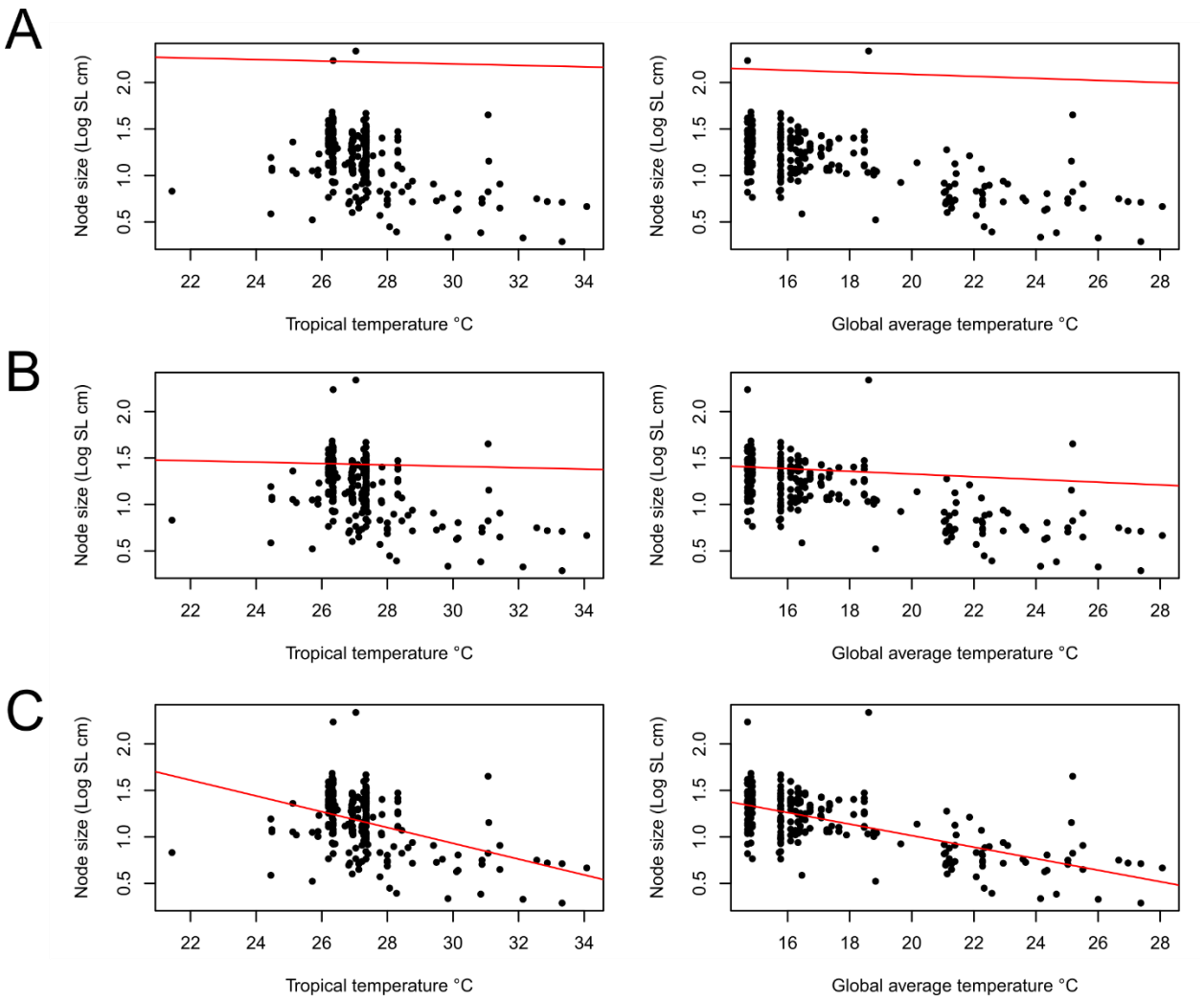


Fig. S15. Relationship between tetraodontiform body size and ocean temperature. Ancestral tetraodontiform body size, inclusive of superfamily Plectocretacicoidea, was reconstructed at each node in the maximum clade credibility tree and PGLS analyses under three different models: A) Ornstein-Uhlenbeck (OU) model, B) Brownian Motion (BM) model, and C) Ordinary Least Squares (OLS) model, were performed for two different curves: tropical (15°N-15°S) sea surface temperatures and global average sea temperatures. Under the best fit model (OU), Tetraodontiform body size is correlated with ocean temperatures using the global average curve (p = 5.571e-03) but not for the tropical curve (p =0. 0653).

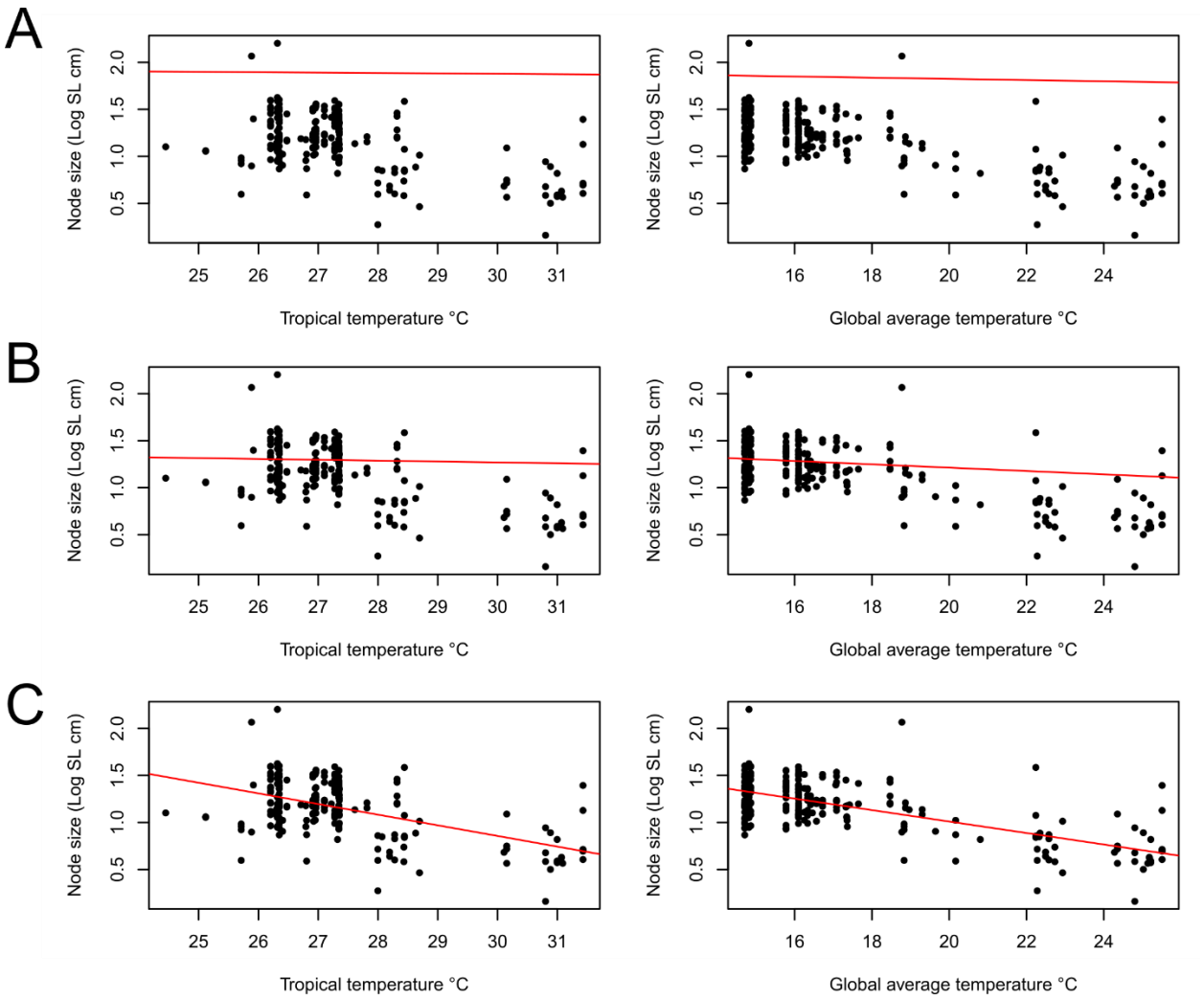


Fig. S16. Relationship between tetraodontiform body size and ocean temperature, without superfamily Plectocretacicoidea. Ancestral tetraodontiform body size, inclusive of superfamily Plectocretacicoidea, was reconstructed at each node in the maximum clade credibility tree and PGLS analyses under three different models: A) Ornstein-Uhlenbeck (OU) model, B) Brownian Motion (BM) model, and C) Ordinary Least Squares (OLS) model, were performed for two different curves: tropical (15°N-15°S) sea surface temperatures and global average sea temperatures (GAT). Under the best-fit model, Tetraodontiform body size is significantly correlated with GAT (BM model, p = 4.84e-05), but not tropical temperatures (OU model, p = 0.447).

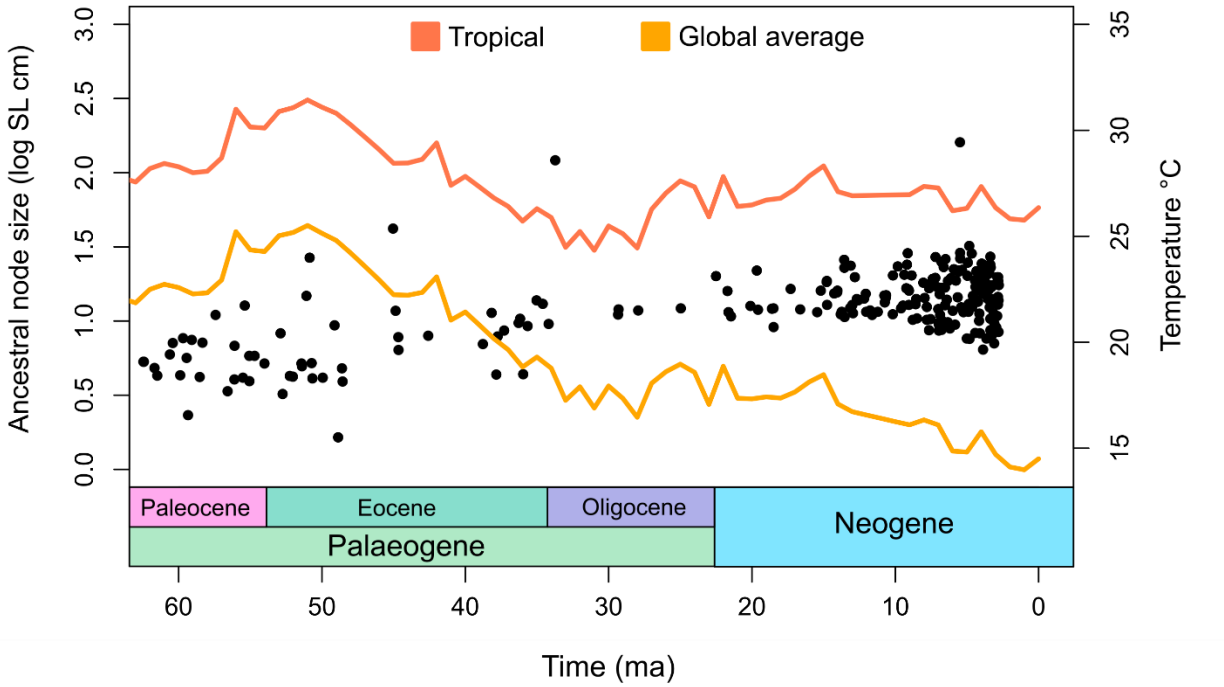


Fig. S17. Tetraodontiform body size and temperature over time, excluding Plectocretacicoidea. Sea surface temperature for tropical latitudes (15°N-15°S) (orange line) and a global average sea temperature (yellow line) are plotted for the past 63 Ma. The reconstructed ancestral node body size (log mean maximum standard length in cm) for tetraodontiforms is plotted in the same time scale. Sea temperatures have been slowly cooling over time, while tetraodontiform body size has been slowly increasing.

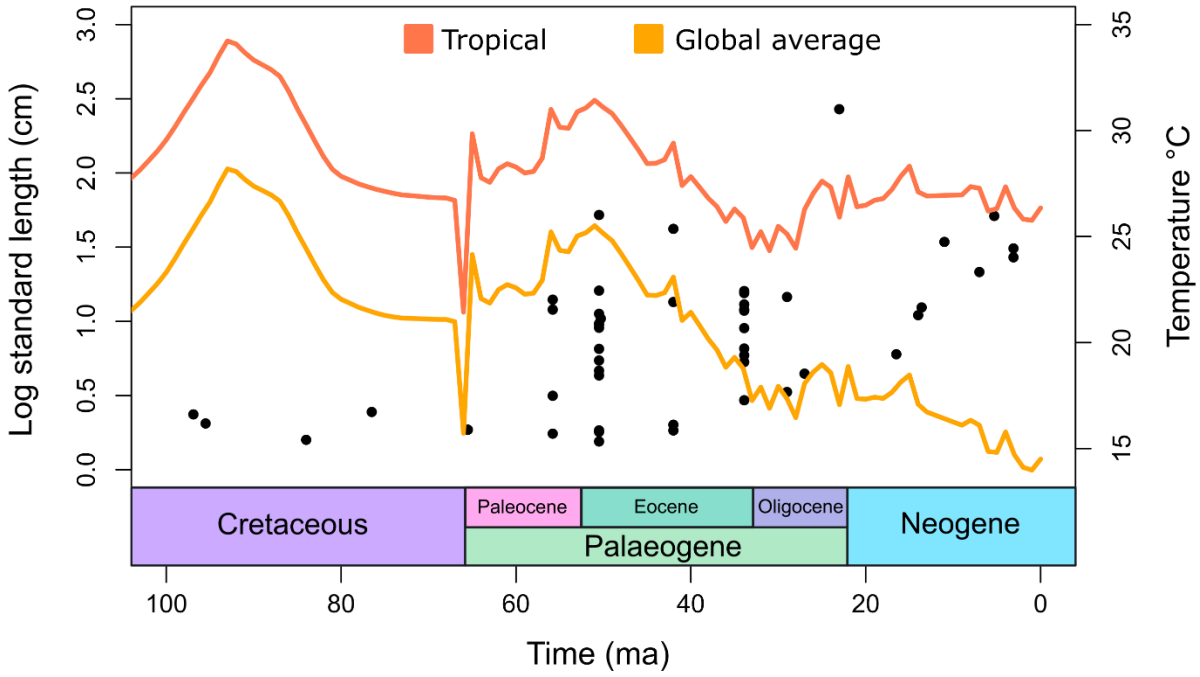


Fig. S18. Fossil tetraodontiform body size and temperature over time. Sea surface temperature for tropical latitudes (15°N-15°S) (orange line) and a global average sea temperature (yellow line) are plotted for the past 100 Ma. Body sizes for only fossil tetraodontiform species (log mean maximum standard length in cm) is plotted in the same time scale. Sea temperatures have been slowly cooling over time, while fossil tetraodontiform body size has been slowly increasing.

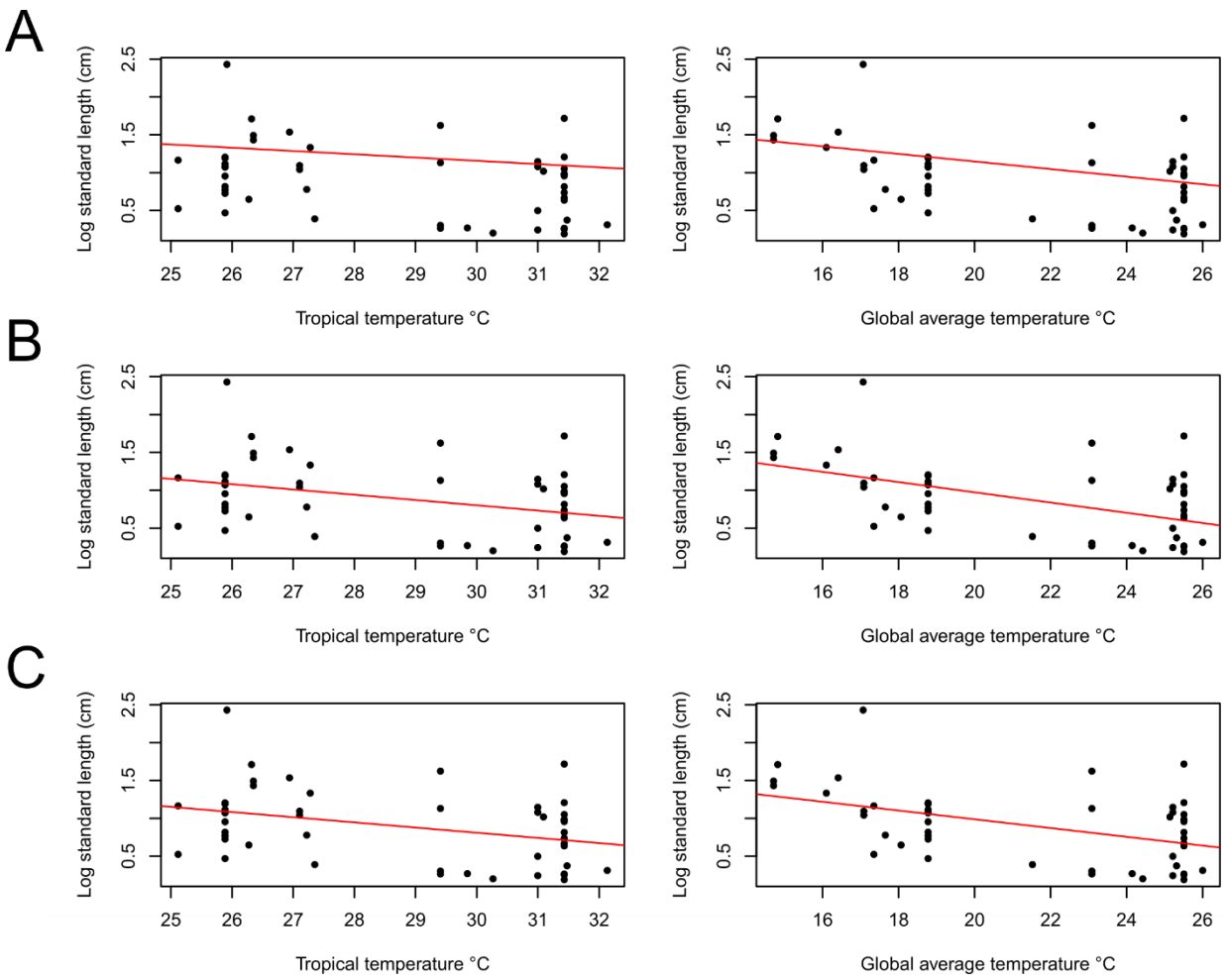


Fig. S19. Relationship between fossil tetraodontiform body size and ocean temperature. A PGLS analysis under three different models: A) Ornstein-Uhlenbeck (OU), B) Brownian Motion (BM), and C) Ordinary Least Squares (OLS) model, were performed between body size for fossil tetraodontiform species (log standard length (cm)) and ocean temperature for two different curves: tropical (15°N-15°S) sea surface temperatures and global average sea temperatures. Under the best fit model (OU), fossil body size is strongly correlated for global average temperature (p = 0.0311), but not tropical temperatures (p =0.1804).

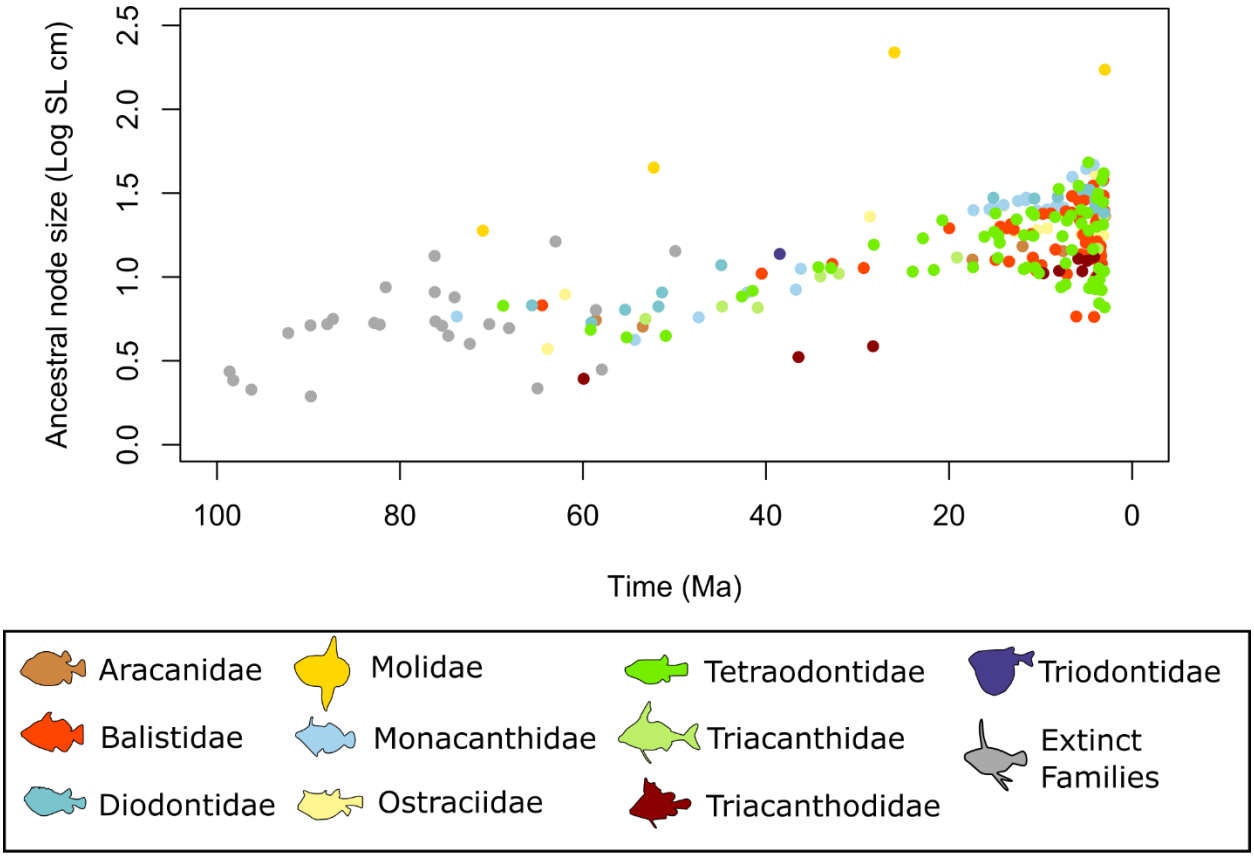


Fig. S20. Reconstructed ancestral node body size (log mean maximum standard length in cm) for tetraodontiforms is plotted against time. Nodal values are colored by family.

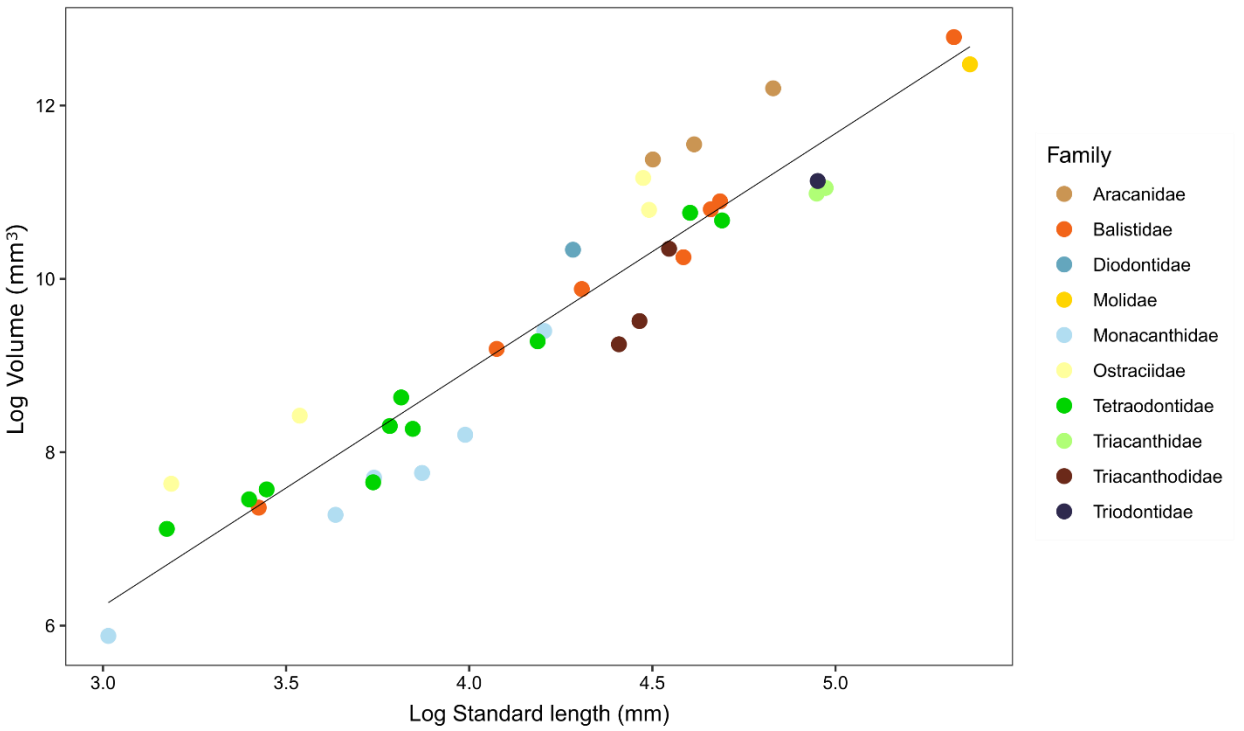


Fig. S21. Log standard length (mm) to log volume (mm^3) relationships between 41 species of tetraodontiform fishes across all 10 extant families. Data were collected from publicly available computed-tomography (CT) scan data accessed from MorphoSource.org as well as newly generated CT data. The image computing software Slicer was used to create three-dimensional models using CT data as input. Models were made solid using the WrapSolidify option to calculate volume. Standard length was measured using the ruler tool. A phylogenetic generalized least squares (PGLS) analysis shows the relationship between tetraodontiform standard length and volume is positively correlated ($p = 0.0005$).

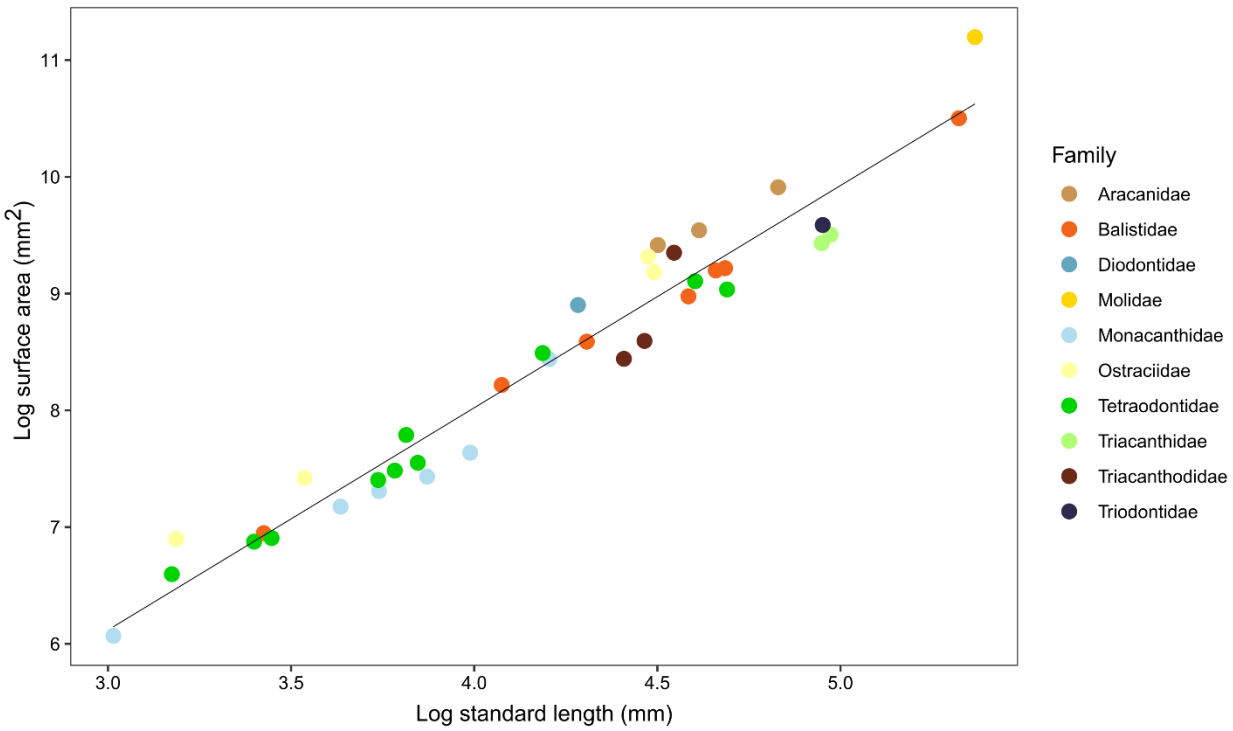


Fig. S22. Log standard length (mm) to log surface area (mm^2) relationships between 41 species of tetraodontiform fishes across all 10 extant families. Data were collected from publicly available computed-tomography (CT) scan data accessed from MorphoSource.org as well as newly generated CT data. The image computing software Slicer was used to create three-dimensional models using CT data as input. Models were made solid using the WrapSolidify option to calculate surface area. Standard length was measured using the ruler tool. A phylogenetic generalized least squares (PGLS) analysis shows the relationship between tetraodontiform standard length and volume is positively correlated ($p = 0.0248$).

Table S1. Summary of tree age prior effect on the root node (Tetraodontiformes+Outgroups). Table indicates the root prior distribution (offset exponential or uniform), whether or not the analysis ran with data, and a summary of the 95% highest posterior density (HPD) values for the mean root age, root age variance, upper and lower age bounds, and median root age. Bounds for the offset exponential root distribution were a minimum age of 95 Ma and a mean age of 107.9 Ma. Bounds for the uniform root distribution were a minimum age of 95.9 Ma and maximum age of 110 Ma.

		95% HPD Interval				
Root Distribution	**With data?**	**Mean**	**Variance**	**Lower**	**Upper**	**Median**
Offset Exponential	Yes	132.3739	204.7114	111.412	161.6181	129.3902
Offset Exponential	No	112.5834	55.90838	99.3592	126.3776	111.5328
Uniform	Yes	108.9578	0.964899	107.2127	110	109.267
Uniform	No	106.3641	6.494229	101.5038	109.9977	106.7817

Table S2. Summary of tree age prior effect on the root node (Tetraodontiformes+Outgroups) when Plectocretacicoidea is excluded. Table indicates the root prior distribution (offset exponential or uniform), whether or not the analysis ran with data, and a summary of the 95% highest posterior density (HPD) values for the mean root age, root age variance, upper and lower age bounds, and median root age. Bounds for the offset exponential root distribution were a minimum age of 59 Ma and a mean age of 61.8 Ma. Bounds for the uniform root distribution were a minimum age of 59 Ma and maximum age of 67.43 Ma

		95% HPD Interval				
Root Distribution	**With data?**	**Mean**	**Variance**	**Lower**	**Upper**	**Median**
Offset Exponential	Yes	93.30206	85.84346	73.55835	110.9355	92.23749
Offset Exponential	No	74.26792	19.14679	66.2045	82.83702	73.96716
Uniform	Yes	67.15891	0.07243	66.53637	67.42502	67.25915
Uniform	No	66.48487	0.714063	64.6329	67.42978	66.73349

Table S3. Summary of tree age prior effect on the crown node (crown Tetraodontiformes). Table indicates the root prior distribution (offset exponential or uniform), whether or not the analysis ran with data, and a summary of the 95% highest posterior density (HPD) values for the mean crown age, crown age variance, upper and lower age bounds, and median crown age. No age priors were placed onto the node for crown Tetraodontiformes.

				95% HPD Interval		
Root Distribution	**With data?**	**Mean**	**Variance**	**Lower**	**Upper**	**Median**
Offset Exponential	Yes	99.8399	49.56933	84.2824	112.6877	100.0196
Offset Exponential	No	94.55334	55.47974	79.4895	108.6024	94.97438
Uniform	Yes	94.72298	11.65628	87.54332	100.2006	95.3393
Uniform	No	92.28921	43.32038	79.27937	103.7278	92.94941

Table S4. Summary of tree age prior effect on the crown node (crown Tetraodontiformes) when Plectocretacicoidea is excluded. Table indicates the root prior distribution (offset exponential or uniform), whether or not the analysis ran with data, and a summary of the 95% highest posterior density (HPD) values for the mean crown age, crown age variance, upper and lower age bounds, and median crown age. No age priors were placed onto the node for crown Tetraodontiformes.

		95% HPD Interval				
Root Distribution	**With data?**	**Mean**	**Variance**	**Lower**	**Upper**	**Median**
Offset Exponential	Yes	80.66592	42.09142	67.29957	93.55297	79.65927
Offset Exponential	No	71.04824	13.99969	64.13806	78.6117	70.88791
Uniform	Yes	62.45651	1.159584	60.80456	64.29212	62.4217
Uniform	No	65.11428	1.201485	62.871	66.96123	65.25346

Table S5. Evolutionary model results, including the deep-sea temperature curve from Cramer et al. 2011 (33). Number of model parameters, mean values for the corrected Akaike information criterion (AICc), mean log likelihood (lnL), and weighted AIC (AICw) are reported.

Evolutionary Model	Parameters	AICc	lnL	AICw
OU climate deep-sea	5	89.87	-39.81	0.9849
OU climate GAT	5	98.25	-43.99	0.0149
OU climate tropical	5	112.42	-51.08	1.25e-05
Mean trend	3	126.91	-60.40	8.95e-09
OU	3	137.88	-65.89	3.71e-11
Rate trend	3	146.78	-70.34	4.33e-13
BM	2	148.01	-71.98	2.34e-13
EB	3	148.03	-70.96	2.31e-13

Table S6. **Evolutionary model results, excluding superfamily Plectocretacicoidea.** Model fitting results for the seven macroevolutionary models tested. Models were tested on 500 trees selected from the combined posterior distributions of five gene subsets. Number of model parameters and mean values for the corrected Akaike information criterion (AICc), log likelihood (lnL), and weighted AIC (AICw) are reported. The strongest support went to the climate-driven model using the global average temperature curve.

Evolutionary Model	Parameters	AICc	lnL	AICw
OU climate GAT	5	111.51	-50.62	0.996
OU climate tropical	5	122.52	-56.13	4.04e-03
OU	3	141.92	-67.91	2.47e-07
Mean trend	3	149.42	-71.66	5.82e-09
EB	3	165.91	-79.90	1.53e-12
Rate trend	3	166.55	-80.22	1.11e-12
BM	2	169.80	-82.87	2.18e-13

Table S7. Number of times each evolutionary model was the best fit. Summary of the best fitting evolutionary model for each of the 500 trees selected from the posterior distribution of the Bayesian analysis, for both fossil schemes (with/without superfamily Plectocretacicoidea). The best fit model for each of the 500 trees was determined from the highest weighted Akaike information criterion (AICw) score. The OU climate model using the global average temperature (GAT) curve was the best fitting model overall (329/500 trees; 65.8%) for the scheme inclusive of the superfamily Plectocretacicoidea and also for the scheme excluding them (395/500 trees; 79%).

	Number of trees (%)	
Evolutionary Model	**With Plectocretacicoidea**	**Without Plectocretacicoidea**
OU climate GAT	329 (65.8%)	395 (79%)
OU climate tropical	107 (21.4 %)	102 (20.4%)
Mean trend	64 (12.8%)	3 (0.006%)
OU	0 (0%)	0 (0%)
Rate trend	0 (0%)	0 (0%)
BM	0 (0%)	0 (0%)
EB	0 (0%)	0 (0%)

Table S8. Decomposed climate OU model fitting results for the global average temperature curve (GAT). Number of model parameters, mean values for corrected Akaike information criterion (AICc), mean log likelihood (lnL), and weighted AICc (AICw) are reported. Most model support went to a temperature curve composed of a trend parameter and a fluctuations parameter.

Model	Parameters	AICc	lnL	AICw
Trend+Fluctuations	6	96.53	-42.08	0.7697
Trend	5	99.77	-44.75	0.1524
Fluctuations	5	101.12	-45.43	0.0777

Table S9. Decomposed climate OU model fitting results for the tropical latitude temperature curve. Number of model parameters, mean values for corrected Akaike information criterion (AICc), mean log likelihood (lnL), and weighted AICc (AICw) are reported. Most model support went to a temperature curve composed of only the fluctuations parameter.

Model	Parameters	AICc	lnL	AICw
Fluctuations	5	95.01	-42.37	0.8522
Trend+Fluctuations	6	98.52	-43.07	0.1477
Trend	5	120.29	-55.01	2.77e-06

Table S10. PGLS model fitting results, including superfamily Plectocretacicoidea. Model fitting results for two temperature curves (global average temperature (GAT) and tropical latitudes) for phylogenetic generalized least squares (PGLS) analyses under an Ornstein-Uhlenbeck (OU) model, a Brownian Motion (BM) model, and a model where phylogeny is not considered (i.e. Ordinary Least Squares (OLS)). The Akaike information criterion (AIC), log likelihood (lnL) and p values are reported for each model. Highest support went to an OU model for both temperature curves.

Temperature curve	Model	AIC	lnL	p value
GAT	OU	-480.584	244.292	5.571E-03
GAT	BM	-452.838	229.418	4.51E-04
GAT	OLS	-2.9266	4.463	2.56E-29
Tropical	OU	-476.312	242.156	6.54E-02
Tropical	BM	-442.797	224.3984	1.23E-01
Tropical	OLS	81.05182	-37.5259	4.76E-11

Table S11. PGLS model fitting results, excluding superfamily Plectocretacicoidea. Model fitting results for two temperature curves (global average temperature (GAT) and tropical latitudes) for phylogenetic generalized least squares (PGLS) analyses under an Ornstein-Uhlenbeck (OU) model, a Brownian Motion (BM) model, and a model where phylogeny is not considered (i.e. Ordinary Least Squares (OLS)). The Akaike information criterion (AIC), log likelihood (lnL) and p values are reported for each model. Highest support went to a BM model for the GAT curve and an OU model for the tropical curve.

Temperature curve	Model	AIC	lnL	p value
GAT	OU	-476.94	242.470	0.208
GAT	BM	-488.651	247.325	4.84E-05
GAT	OLS	-38.771	22.385	1.82E-31
Tropical	OU	-475.923	241.961	0.447
Tropical	BM	-474.703	240.352	0.0995
Tropical	OLS	25.211	-9.605	1.37E-17

Table S12. PGLS model fitting results for only fossil Tetraodontiformes. Model fitting results for two temperature curves (global average temperature (GAT) and tropical latitudes) for phylogenetic generalized least squares (PGLS) analyses under an Ornstein-Uhlenbeck (OU) model, a Brownian Motion (BM) model, and a model where phylogeny is not considered (i.e. Ordinary Least Squares (OLS)). The Akaike information criterion (AIC), log likelihood (lnL) and p values are reported for each model. Highest support went to an OU model for both temperature curves.

Temperature curve	Model	AIC	lnL	p value
GAT	OU	48.627	-20.313	0.0311
GAT	BM	51.578	-22.789	1.53E-03
GAT	OLS	60.689	-27.344	7.53E-04
Tropical	OU	51.543	-21.771	0.1804
Tropical	BM	56.779	-25.389	0.0239
Tropical	OLS	66.134	-30.067	0.0128

Table S13. List of specimens newly sequenced (141 Tetraodontiformes representing 131 unique species) and four outgroup Lophiiformes). The museum catalog tissue and voucher numbers are provided. One specimen (*Rhinecanthus verrucosus*) was excluded from the analyses due to low capture efficiency and only being represented in a single gene.

Affiliation	**Family**	**Species Name**	**Voucher Number**	**Tissue Collection Number**
Outgroup	Lophiidae	*Lophiodes caulinaris*	USNM 421345	USNM AG7PC44
Outgroup	Lophiidae	*Lophiodes spilurus*	USNM 421229	USNM AG7PF91
Outgroup	Lophiidae	*Lophiomus setigerus*	CSIRO H 6570-04	CSIRO GT 1833
Outgroup	Lophiidae	*Lophius litulon*	CSIRO H 7394-66	CSIRO GT 6889
Eupercaria	Aracanidae	*Anoplocapros amygdaloides*	CSIRO H 6351-13	CSIRO GT 5967
Eupercaria	Aracanidae	*Anoplocapros inermis*	CSIRO H 6836-11	CSIRO GT 3695
Eupercaria	Aracanidae	*Anoplocapros lenticularis*	CSIRO H 6341-11	CSIRO GT 214
Eupercaria	Aracanidae	*Aracana aurita*	CSIRO H 6812-01	CSIRO GT 2350
Eupercaria	Aracanidae	*Caprichthys gymnura*	CSIRO H 6341-02	CSIRO GT 203
Eupercaria	Aracanidae	*Capropygia unistriata*	CSIRO H 6350-23	CSIRO GT 5956
Eupercaria	Balistidae	*Abalistes filamentosus*	CSIRO unreg LM964	CSIRO IN02982
Eupercaria	Balistidae	*Balistes capriscus*	USNM 405175	USNM AD9NF67
Eupercaria	Balistidae	*Balistes polylepis*	USNM 421241	USNM AG7PC77
Eupercaria	Balistidae	*Balistes vetula*	No voucher	UPRFL0353
Eupercaria	Balistidae	*Balistoides viridescens*	USNM 403379	USNM AG9RF59
Eupercaria	Balistidae	*Canthidermis sufflamen*	USNM 423028	USNM AC7BZ85
Eupercaria	Balistidae	*Melichthys niger*	No voucher	UPRFL0061
Eupercaria	Balistidae	*Melichthys vidua*	STRI-X-60	STRI BFT11786
Eupercaria	Balistidae	*Odonus niger*	USNM 409200	USNM AG7PW08
Eupercaria	Balistidae	*Pseudobalistes fuscus*	USNM 400523	USNM AG9RP93
Eupercaria	Balistidae	*Rhinecanthus aculeatus*	USNM 400511	USNM AG9RU13

Eupercaria	Balistidae	*Rhinecanthus lunula*	USNM 392212	USNM AG5NP92
Eupercaria	Balistidae	*Rhinecanthus rectangulus*	USNM 400515	USNM AG9RQ72
Eupercaria	Balistidae	*Rhinecanthus verrucosus*	USNM 435559	USNM AB6QS48
Eupercaria	Balistidae	*Sufflamen bursa*	USNM 392517	USNM AG5NQ79
Eupercaria	Balistidae	*Sufflamen chrysopterum*	AMS I.44740-006	CSIRO UG0735
Eupercaria	Balistidae	*Sufflamen chrysopterum*	USNM 435999	USNM AC1VD99
Eupercaria	Balistidae	*Xanthichthys auromarginatus*	USNM 409459	USNM AG7PY24
Eupercaria	Balistidae	*Xanthichthys caeruleolineatus*	USNM 409149	USNM AG7PV58
Eupercaria	Balistidae	*Xanthichthys caeruleolineatus*	CSIRO H 6318-01	CSIRO GT 168
Eupercaria	Balistidae	*Xanthichthys lineopunctatus*	CSIRO unreg KD795	CSIRO IN02440
Eupercaria	Diodontidae	*Chilomycterus antennatus*	USNM 414367	USNM AC2WT95
Eupercaria	Diodontidae	*Chilomycterus reticulatus*	CSIRO H 7305-08	CSIRO IN02878
Eupercaria	Diodontidae	*Chilomycterus schoepfii*	USNM 415605	USNM AC2WQ63
Eupercaria	Diodontidae	*Chilomycterus spinosus*	USNM 405184	USNM AD9NF85
Eupercaria	Diodontidae	*Diodon holocanthus*	No voucher	UPRFL0286
Eupercaria	Diodontidae	*Diodon hystrix*	STRI-X-325	STRI BFT11552
Eupercaria	Diodontidae	*Diodon nicthemerus*	CSIRO H 6347-31	CSIRO GT 5963
Eupercaria	Diodontidae	*Tragulichthys jaculiferus*	CSIRO unreg	CSIRO GT 4734
Eupercaria	Monacanthidae	*Acanthaluteres spilomelanurus*	CSIRO H 7101-12	CSIRO GT 5571
Eupercaria	Monacanthidae	*Acanthaluteres vittiger*	CSIRO H 6945-04	CSIRO GT 4819
Eupercaria	Monacanthidae	*Aluterus heudelotii*	USNM 405092	USNM AD9NE01
Eupercaria	Monacanthidae	*Aluterus monoceros*	No voucher	UPRFL1497
Eupercaria	Monacanthidae	*Aluterus scriptus*	No voucher	UPRFL0003

Eupercaria	Monacanthidae	*Amanses scopas*	USNM 392430	USNM AG5NQ54
Eupercaria	Monacanthidae	*Anacanthus barbatus*	CSIRO H 6146-06	CSIRO GT 4577
Eupercaria	Monacanthidae	*Brachaluteres jacksonianus*	CSIRO H 7103-08	CSIRO GT 5588
Eupercaria	Monacanthidae	*Brachaluteres taylori*	CSIRO H 6914-01	CSIRO GT 4430
Eupercaria	Monacanthidae	*Cantherhines dumerilii*	USNM 399512	USNM AG9RT41
Eupercaria	Monacanthidae	*Cantherhines fronticinctus*	USNM 436443	USNM AC1VI44
Eupercaria	Monacanthidae	*Cantherhines longicaudus*	USNM 400537	USNM AG9RP49
Eupercaria	Monacanthidae	*Cantherhines nukuhiva*	USNM 409267	USNM AG7PW33
Eupercaria	Monacanthidae	*Cantherhines pardalis*	USNM 435717	USNM AB6QU07
Eupercaria	Monacanthidae	*Cantherhines pullus*	No voucher	UPRFL0403
Eupercaria	Monacanthidae	*Cantherhines sandwichiensis*	USNM 392012	USNM AG5NP68
Eupercaria	Monacanthidae	*Cantheschenia grandisquamis*	CSIRO unreg	CSIRO GT 4763
Eupercaria	Monacanthidae	*Chaetodermis penicilligerus*	CSIRO H 6911-05	CSIRO GT 4402
Eupercaria	Monacanthidae	*Chaetodermis penicilligerus*	CSIRO H 8251-01	CSIRO GT 10694
Eupercaria	Monacanthidae	*Eubalichthys bucephalus*	Not retained	CSIRO GT 216
Eupercaria	Monacanthidae	*Eubalichthys caeruleoguttatus*	CSIRO H 8282-02	CSIRO GT 10791
Eupercaria	Monacanthidae	*Eubalichthys gunnii*	CSIRO H 6944-01	CSIRO GT 4806
Eupercaria	Monacanthidae	*Eubalichthys mosaicus*	CSIRO H 6350-07	CSIRO GT 5914
Eupercaria	Monacanthidae	*Eubalichthys quadrispinis*	CSIRO H 6348-07	CSIRO GT 311
Eupercaria	Monacanthidae	*Meuschenia trachylepis*	CSIRO H 6838-09	CSIRO GT 3654
Eupercaria	Monacanthidae	*Oxymonacanthus longirostris*	AMS I.44739-002	CSIRO UG0755

Eupercaria	Monacanthidae	*Paramonacanthus choirocephalus*	CSIRO H 6145-06	CSIRO GT 4410
Eupercaria	Monacanthidae	*Paramonacanthus filicauda*	CSIRO H 6319-07	CSIRO GT 4617
Eupercaria	Monacanthidae	*Paramonacanthus japonicus*	CSIRO unreg BY011	CSIRO IN00550
Eupercaria	Monacanthidae	*Paramonacanthus lowei*	CSIRO unreg	CSIRO CMR005207a
Eupercaria	Monacanthidae	*Paramonacanthus lowei*	CSIRO unreg	CSIRO CMR005207b
Eupercaria	Monacanthidae	*Paramonacanthus oblongus*	CSIRO unreg	CSIRO GT 4244
Eupercaria	Monacanthidae	*Pervagor aspricaudus*	USNM 390982	USNM AG5NP14
Eupercaria	Monacanthidae	*Pervagor janthinosoma*	AMS I.44714-026	CSIRO UG0193
Eupercaria	Monacanthidae	*Pervagor marginalis*	USNM 409029	USNM AG7PU39
Eupercaria	Monacanthidae	*Pervagor melanocephalus*	USNM 435755	USNM AB6QU45
Eupercaria	Monacanthidae	*Pseudalutarius nasicornis*	CSIRO H 6937-08	CSIRO GT 4711
Eupercaria	Monacanthidae	*Pseudomonacanthus elongatus*	CSIRO H 6904-02	CSIRO GT 4289
Eupercaria	Monacanthidae	*Pseudomonacanthus elongatus*	CSIRO H 8249-02	CSIRO GT 10812
Eupercaria	Monacanthidae	*Pseudomonacanthus macrurus*	USNM 435396	USNM AB6QQ85
Eupercaria	Monacanthidae	*Pseudomonacanthus peroni*	CSIRO H 4643-03	CSIRO 10V
Eupercaria	Monacanthidae	*Rudarius minutus*	WAM P.33523-001	CSIRO UG0547
Eupercaria	Monacanthidae	*Stephanolepis hispidus*	USNM 405062	USNM AD9ND41
Eupercaria	Monacanthidae	*Stephanolepis setifer*	USNM 419328	USNM AC4YM29
Eupercaria	Monacanthidae	*Thamnaconus degeni*	CSIRO H 6942-04	CSIRO GT 4786
Eupercaria	Monacanthidae	*Thamnaconus striatus*	CSIRO H 7220-09	CSIRO IN02545
Eupercaria	Monacanthidae	*Thamnaconus tessellatus*	CSIRO H 6422-04	CSIRO GT 715

Eupercaria	Ostraciidae	*Acanthostracion polygonius*	USNM 421721	USNM AC2WV43
Eupercaria	Ostraciidae	*Acanthostracion quadricornis*	No voucher	UPRFL0075
Eupercaria	Ostraciidae	*Lactophrys trigonus*	USNM 415636	USNM AC2WQ88
Eupercaria	Ostraciidae	*Lactophrys triqueter*	No voucher	UPRFL0225
Eupercaria	Ostraciidae	*Lactophrys triqueter*	No voucher	UPRFL0072
Eupercaria	Ostraciidae	*Lactoria cornuta*	USNM 403207	USNM AG9RD87
Eupercaria	Ostraciidae	*Ostracion cubicus*	USNM 391997	USNM AG5NP65
Eupercaria	Ostraciidae	*Ostracion meleagris*	USNM 400508	USNM AG9RT93
Eupercaria	Ostraciidae	*Ostracion nasus*	CSIRO unreg	CSIRO GT 4682
Eupercaria	Ostraciidae	*Ostracion whitleyi*	USNM 409290	USNM AG7PW56
Eupercaria	Tetraodontidae	*Arothron hispidus*	USNM 400510	USNM AG9RU11
Eupercaria	Tetraodontidae	*Arothron hispidus*	CSIRO H 8228-02	CSIRO GT 10604
Eupercaria	Tetraodontidae	*Arothron meleagris*	USNM 391239	USNM AG5NP43
Eupercaria	Tetraodontidae	*Arothron nigropunctatus*	NMV A.29880-008	CSIRO UG0991
Eupercaria	Tetraodontidae	*Canthigaster axiologus*	USNM 400503	USNM AG9RT89
Eupercaria	Tetraodontidae	*Canthigaster bennetti*	USNM 439632	USNM AG5NT52
Eupercaria	Tetraodontidae	*Canthigaster criobe*	USNM 400521	USNM AG9RL45
Eupercaria	Tetraodontidae	*Canthigaster cyanospilota*	CSIRO H 8206-01	CSIRO GT 10464
Eupercaria	Tetraodontidae	*Canthigaster janthinoptera*	USNM 392325	USNM AG5NQ15
Eupercaria	Tetraodontidae	*Canthigaster marquesensis*	USNM 409458	USNM AG7PY23
Eupercaria	Tetraodontidae	*Canthigaster punctatissima*	No voucher	UPRFL0354
Eupercaria	Tetraodontidae	*Canthigaster rapaensis*	USNM 400531	USNM AG9RO95
Eupercaria	Tetraodontidae	*Canthigaster rostrata*	No voucher	UPRFL0412
Eupercaria	Tetraodontidae	*Canthigaster solandri*	USNM 400518	USNM AG9RM90

Eupercaria	Tetraodontidae	*Canthigaster solandri*	SAIAB 78221	KUIT 7224
Eupercaria	Tetraodontidae	*Canthigaster valentini*	USNM 390990	USNM AG5NP15
Eupercaria	Tetraodontidae	*Chelonodontops patoca*	CSIRO unreg LM921	CSIRO IN02921
Eupercaria	Tetraodontidae	*Feroxodon multistriatus*	CSIRO unreg	CSIRO GT 4433
Eupercaria	Tetraodontidae	*Feroxodon multistriatus*	CSIRO H 8257-01	CSIRO GT 10714
Eupercaria	Tetraodontidae	*Lagocephalus cheesemanii*	CSIRO H 7285-04	CSIRO GT 6570
Eupercaria	Tetraodontidae	*Lagocephalus guentheri*	USNM 437742	USNM AC1VJ91
Eupercaria	Tetraodontidae	*Lagocephalus laevigatus*	No voucher	UPRFL1635
Eupercaria	Tetraodontidae	*Lagocephalus* cf. *suezensis*	CSIRO unreg	CSIRO GT4214
Eupercaria	Tetraodontidae	*Lagocephalus spadiceus*	USNM 403468	USNM AG9RG48
Eupercaria	Tetraodontidae	*Sphoeroides annulatus*	HB-1190	STRI BFT05089
Eupercaria	Tetraodontidae	*Sphoeroides dorsalis*	USNM 433079	USNM AB4OP01
Eupercaria	Tetraodontidae	*Sphoeroides greeleyi*	No voucher	UPRFL1637
Eupercaria	Tetraodontidae	*Sphoeroides lobatus*	No voucher	UPRFL1301
Eupercaria	Tetraodontidae	*Sphoeroides maculatus*	USNM 423825	USNM AE2QZ45
Eupercaria	Tetraodontidae	*Sphoeroides marmoratus*	USNM 405093	USNM AD9NE03
Eupercaria	Tetraodontidae	*Sphoeroides nephelus*	USNM 415007	USNM AC4YB58
Eupercaria	Tetraodontidae	*Sphoeroides pachygaster*	USNM 405072	USNM AD9ND61
Eupercaria	Tetraodontidae	*Sphoeroides spengleri*	No voucher	UPRFL0410
Eupercaria	Tetraodontidae	*Sphoeroides testudineus*	USNM 414276	USNM AC8CC45
Eupercaria	Tetraodontidae	*Takifugu niphobles*	No voucher	ANSP 206060 large
Eupercaria	Tetraodontidae	*Torquigener hicksii*	CSIRO H 6451-01	CSIRO GT 741
Eupercaria	Tetraodontidae	*Torquigener parcuspinus*	CSIRO H 8267-06	CSIRO GT 10748
Eupercaria	Triacanthidae	*Pseudotriacanthus strigilifer*	CSIRO H 7217-12	CSIRO IN02762

Eupercaria	Triacanthidae	*Triacanthus nieuhofi*	CSIRO H 8157-03	CSIRO IN02487
Eupercaria	Triacanthidae	*Tripodichthys blochii*	USNM 424823	USNM AH0SW64
Eupercaria	Triacanthodidae	*Atrophacanthus* sp.	USNM 440414	USNM AB4OL14
Eupercaria	Triacanthodidae	*Halimochirurgus centriscoides*	No voucher	ASIZP 0913987
Eupercaria	Triacanthodidae	*Halimochirurgus centriscoides*	CSIRO H 6574-18	CSIRO GT 1416
Eupercaria	Triacanthodidae	*Halimochirurgus* sp.	CSIRO H 7135-19	CSIRO GT 5806
Eupercaria	Triacanthodidae	*Hollardia hollardia*	VIMS 40109	VIMS 40109
Eupercaria	Triacanthodidae	*Hollardia meadi*	USNM 431710	USNM AG9RL06
Eupercaria	Triacanthodidae	*Parahollardia lineata*	VIMS 40107	VIMS 40107
Eupercaria	Triacanthodidae	*Paratriacanthodes retrospinis*	NMV A 29672-007	CSIRO GT 1472
Eupercaria	Triacanthodidae	*Triacanthodes* sp. 1	CSIRO H 6570-33	CSIRO GT 1862

Table S14. List of prior distributions used for node dating in previous studies that included Tetraodontiformes. Lists the hard minimum ages and 95% soft maximum ages, if provided. Indicates whether or not the superfamily Plectocretacicoidea was included in the divergence time estimations.

Clade	Paper	Hard minimum age	95% Soft maximum age	Includes Plectocretacicoidea?
Tetraodontiformes + Lophiiformes	Hughes et al. 2018 (3)	85 Ma	122 Ma	Yes- *Cretatriacanthus guidottii*
Tetraodontiformes	Near et al. 2013 (5)	---	---	No
Tetraodontiformes	Dornburg et al. 2014 (29)	70.08 Ma	109.845 Ma	Yes- *Plectocretacicus clarae*, *Cretatriacanthus guidottii*
Tetraodontiformes	Betancur-R et al. 2013 (1)	85 Ma	122 Ma	Yes- *Cretatriacanthus guidottii*
Tetraodontiformes + Lophiiformes	Chen et al. 2014 (30)	83 Ma	124 Ma	Yes- *Cretatriacanthus guidottii*
Tetraodontiformes	Santini et al. 2009 (31)	59 Ma	98 Ma	Yes- *Plectocretacicus clarae*
Tetraodontiformes	Near et al. 2012 (32)	---	71.4 Ma	No
Diodontidae + Tetraodontidae	Hughes et al. 2018 (3)	50 Ma	85 Ma	N/A
Tetraodontidae	Hughes et al. 2018 (3)	32 Ma	50 Ma	N/A
Tetraodontidae	Betancur-R et al. 2013 (1)	32 Ma	50 Ma	N/A
Diodontidae + Tetraodontidae	Betancur-R et al. 2013 (1)	50 Ma	85 Ma	N/A
Molidae	Betancur-R et al. 2013 (1)	41 Ma	85 Ma	N/A
Aracanidae + Ostraciidae	Betancur-R et al. 2013 (1)	50 Ma	85 Ma	N/A
Balistidae	Betancur-R et al. 2013 (1)	35 Ma	85 Ma	N/A

Supplementary References

1. R. Betancur-R., *et al.*, The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr* **5** (2013).
2. L. C. Hughes, *et al.*, Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Molecular Ecology Resources* **21**, 816–833 (2021).
3. L. C. Hughes, *et al.*, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci USA* **115**, 6249–6254 (2018).
4. C. Li, G. Ortí, G. Zhang, G. Lu, A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* **7**, 44 (2007).
5. T. J. Near, *et al.*, Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences* **110**, 12738–12743 (2013).
6. V. Ranwez, E. J. P. Douzery, C. Cambon, N. Chantret, F. Delsuc, MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution* **35**, 2582–2584 (2018).
7. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
8. M. Kearse, *et al.*, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
9. M. L. Borowiec, AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660 (2016).
10. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).
11. D. Arcila, *et al.*, Testing the Utility of Alternative Metrics of Branch Support to Address the Ancient Evolutionary Radiation of Tunas, Stromateoids, and Allies (Teleostei: Pelagiaria). *Systematic Biology* (2021) https:/doi.org/10.1093/sysbio/syab018 (September 30, 2021).
12. S. RATNASINGHAM, P. D. N. HEBERT, bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Mol Ecol Notes* **7**, 355–364 (2007).
13. C. Yang, *et al.*, Efficient COI barcoding using high throughput single-end 400 bp sequencing. *BMC Genomics* **21**, 862 (2020).
14. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
15. R. Lanfear, B. Calcott, D. Kainer, C. Mayer, A. Stamatakis, Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol* **14**, 82 (2014).
16. B. Q. Minh, M. A. T. Nguyen, A. von Haeseler, Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution* **30**, 1188–1195 (2013).
17. D. Arcila, J. C. Tyler, Mass extinction in tetraodontiform fishes linked to the Palaeocene–Eocene thermal maximum. *Proc. R. Soc. B.* **284**, 20171771 (2017).
18. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
19. M. Rincon-Sandoval, *et al.*, Evolutionary determinism and convergence associated with water-column transitions in marine fishes. *Proc Natl Acad Sci USA* **117**, 33396–33403 (2020).

20. A. Santaquiteria, *et al.*, Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences. *Systematic Biology*, syab028 (2021).
21. M. E. Alfaro, *et al.*, Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. *Nat Ecol Evol* **2**, 688–696 (2018).
22. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214 (2007).
23. R. Froese, D. Pauly, FishBase. *www.fishbase.org*.
24. R. Kikinis, S. D. Pieper, K. G. Vosburgh, "3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support" in *Intraoperative Imaging and Image-Guided Therapy*, F. A. Jolesz, Ed. (Springer, 2014), pp. 277–289.
25. C. R. Scotese, H. Song, B. J. W. Mills, D. G. van der Meer, Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. *Earth-Science Reviews* **215**, 103503 (2021).
26. R Core Team, *R: A language and environment for statistical computing* (2020).
27. J. Clavel, A. Brinkworth, OUenv: Climatic/Environment dependent Ornstein-Uhlenbeck model of trait evolution. GitHub Repository. https://github.com/JClavel/OUenv
28. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).
29. A. Dornburg, J. P. Townsend, M. Friedman, T. J. Near, Phylogenetic informativeness reconciles ray-finned fish molecular divergence times. *BMC Evol Biol* **14**, 169 (2014).
30. W.-J. Chen, *et al.*, New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). *Front. Mar. Sci.* **1** (2014).
31. F. Santini, L. J. Harmon, G. Carnevale, M. E. Alfaro, Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol Biol* **9**, 194 (2009).
32. T. J. Near, *et al.*, Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences* **109**, 13698–13703 (2012).
33. B. S. Cramer, K. G. Miller, P. J. Barrett, J. D. Wright, Late Cretaceous–Neogene trends in deep ocean temperature and continental ice volume: Reconciling records of benthic foraminiferal geochemistry (δ^{18}O and Mg/Ca) with sea level history. *J. Geophys. Res.* **116**, C12023 (2011).

Appendix 2

Supplemental Materials for Chapter 2:

Evolutionary innovation accelerates morphological diversification in pufferfishes and their relatives

Emily M. Troyer[1], Kory M. Evans[2], Christopher Goatley[3-5], Matt Friedman[6], Giorgio Carnevale[7], Benjamin Nicholas[8], Matthew Kolmann[9], Katherine E. Bemis[10,11], Dahiana Arcila[12*]

[1]*Department of Biology and Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, 730 Van Vleet Oval, Richards Hall, Norman, OK 73019, USA*
[2]*Rice University, Biosciences Department, 6100 Main Street, Houston, TX 77005, USA*
[3]*School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton, Hampshire SO14 3ZH, UK*
[4]*Australian Museum Research Institute, Australian Museum, 1 William Street, Sydney, NSW 2010, Australia*
[5]*Function, Evolution and Anatomy Research (FEAR) Lab, School of Environmental and Rural Science, University of New England, Armidale, NSW 235, Australia*
[6]*Museum of Paleontology and Department of Earth and Environmental Sciences, University of Michigan, 1105 North University Avenue, Ann Arbor, Michigan 48109, USA*
[7]*Dipartimento di Scienze della Terra, Università degli Studi di Torino, Via Valperga Caluso 35, 10125 Turin, Italy*
[8]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA*
[9]*Department of Biology, University of Louisville, Louisville, KY 40292, USA*
[10]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, USA*
[11]*National Systematics Laboratory, Office of Science and Technology, NOAA Fisheries, Washington, D.C. 20560, USA*
[12]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA*

Appendix 2 Includes:

Figures S1 to S8
Tables S1 to S3
Supplemental References

Other supplementary materials for this manuscript include the following:

Datasets are available on Dryad (doi:10.5061/dryad.4f4qrfjjx). These include:

- Appendix 1: List of species CT scanned, including museum collection code, specimen number, where specimen was scanned, whether scan is available on MorphoSource, and trait categorizations (habitat, mouth type).
- Appendix 2: All R scripts and code used in this study, along with all files needed to run code (i.e. phylogenetic trees, landmark files).

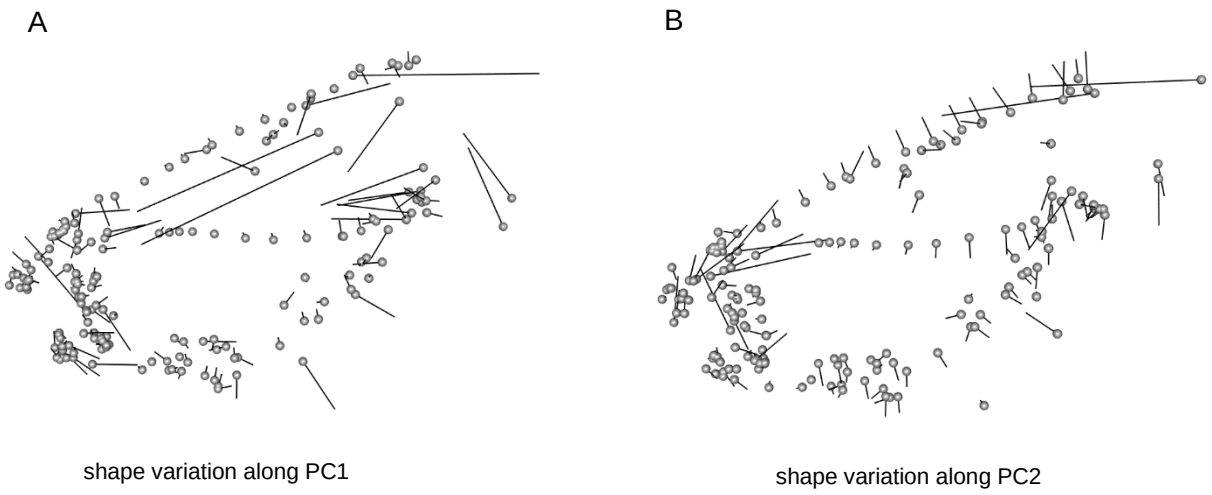
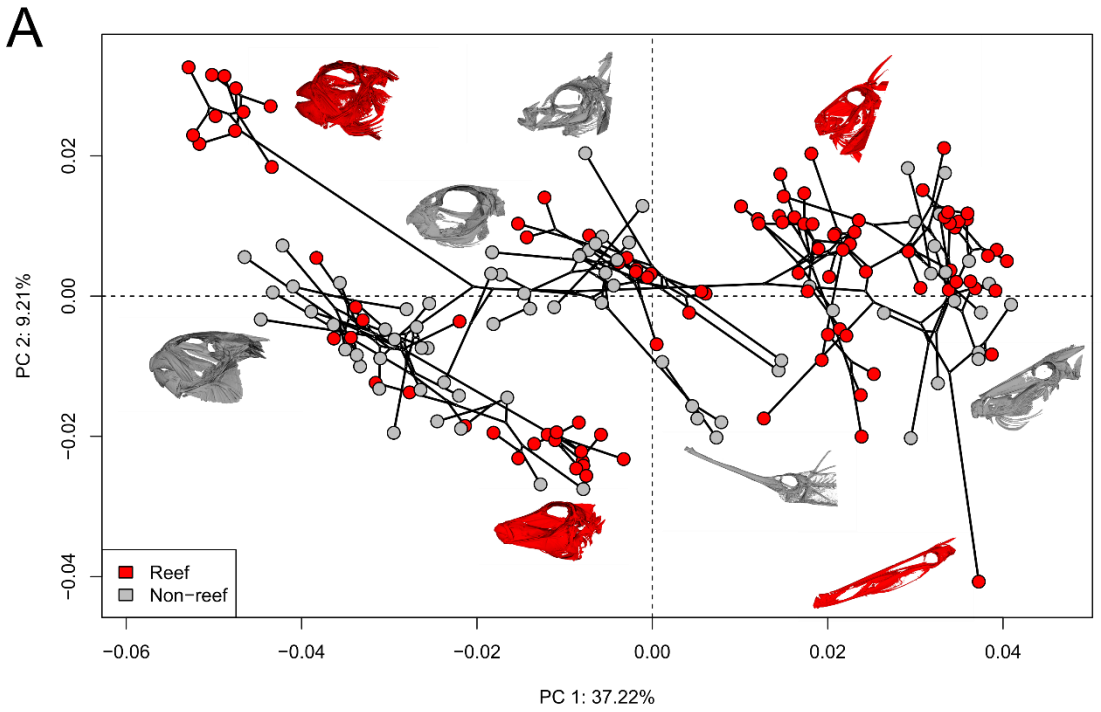
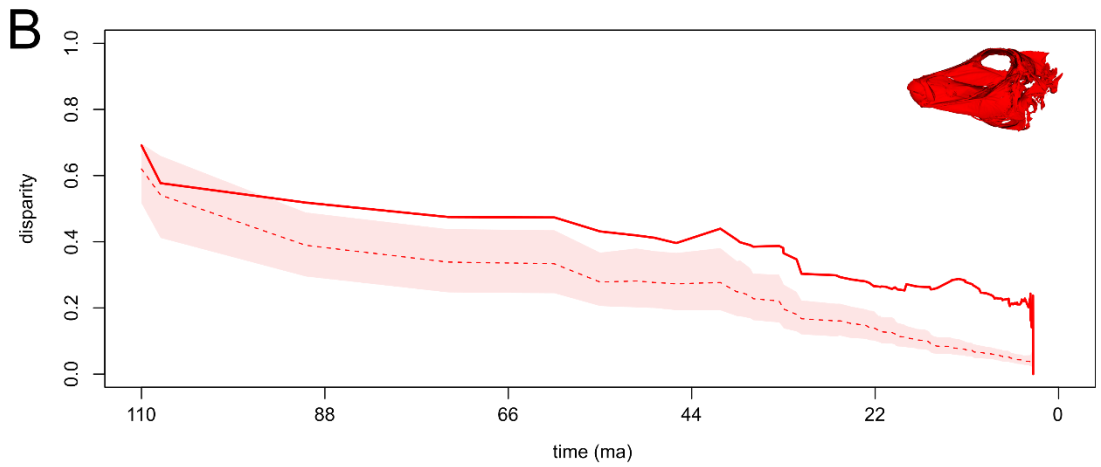


Figure S1. Major axes of shape variation for PCs 1 and 2. Ball-and-stick models show changes in landmark positions between species with the most positive (grey balls) and the most negative PC scores (vector tips). **A)** PC1 explains 37.22% of the shape variance. The positive values are characterized by a more posterior positioned orbit, shorter supraoccipital crest, thinner preopercle, and thinner margin of the first tooth. The negative values are characterized by a more anterior positioned orbit, longer supraoccipital crest, wider preopercle, and a more elongated distal margin of the first tooth, corresponding to a beak. **B)** PC2 explains 9.21% of the shape variance. The positive values are characterized by a wider ethmoid and frontal, a more curved parasphenoid, and a more distal orbit. The negative values are characterized by a thinner ethmoid and frontal, a straighter parasphenoid, and proximal orbit.


A
PC 2: 9.21%
0.02
0.00
−0.02
−0.04
Reef
Non−reef
−0.06
−0.04
−0.02
0.00
0.02
0.04
PC 1: 37.22%


B
disparity
0.0
0.2
0.4
0.6
0.8
1.0
110
88
66
44
22
0
time (ma)

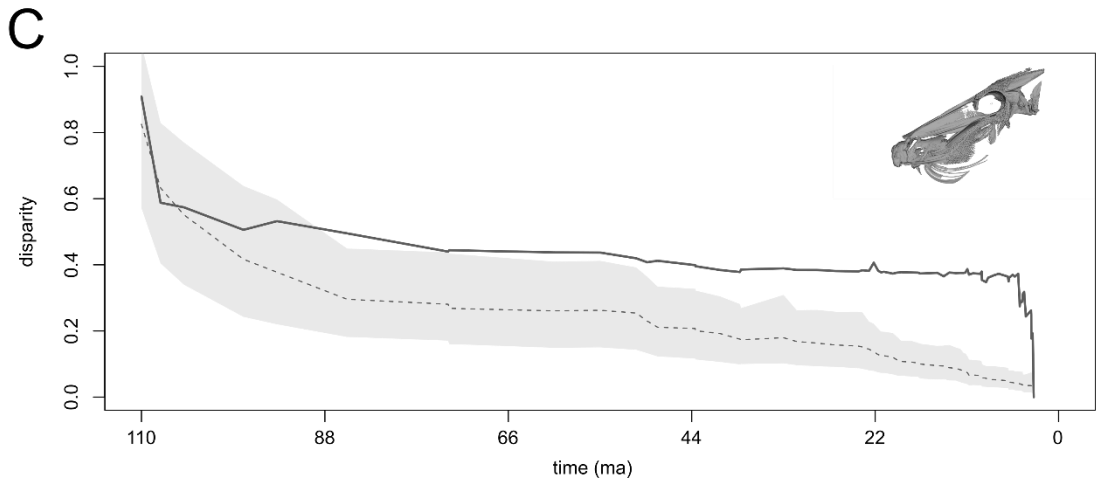

C
disparity
0.0
0.2
0.4
0.6
0.8
1.0
110
88
66
44
22
0
time (ma)

Figure S2. Skull shape disparity and evolution across 176 tetraodontiform species. A) Phylomorphospace analysis of skull shape in tetraodontiform fishes. Reef-associated species are indicated in red while non-reef-associated species are in grey. B) Disparity through time plots of reef-associated and C) non-reef-associated tetraodontiform species showing disparity rate over scaled time. Dashed lines indicate the Brownian motion expectation and shaded regions represent the 95% confidence interval. Solid lines indicate the actual measured disparity.

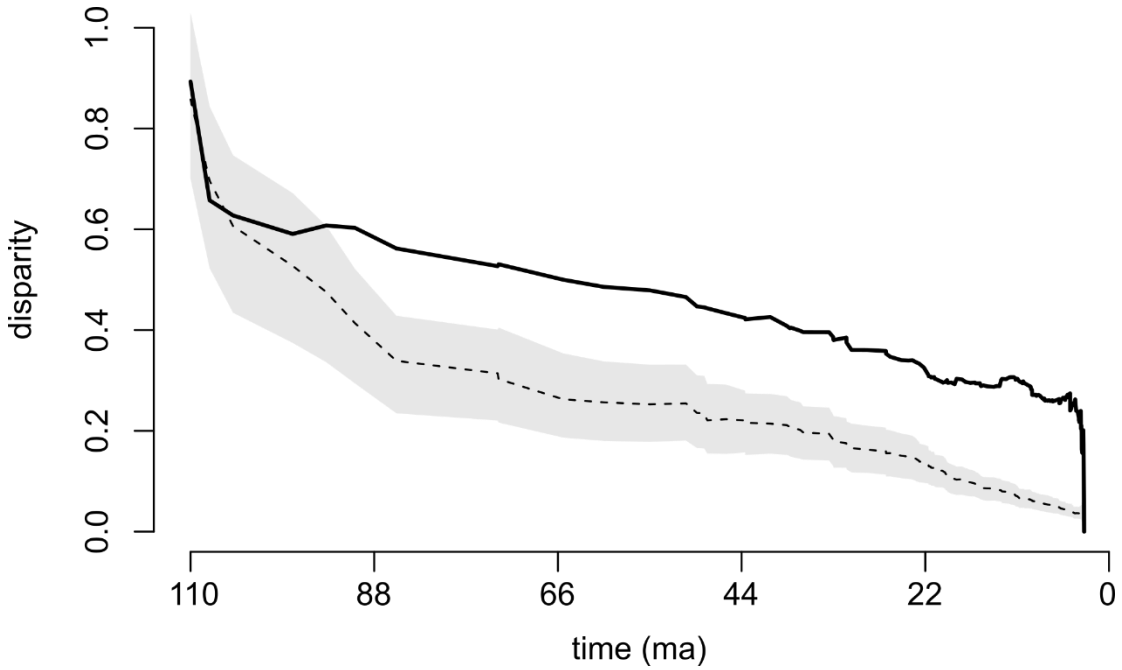


Figure S3. **Disparity through time plot for all Tetraodontiformes.** Dashed line indicates the Brownian motion expectation, and the shaded region represents the 95% confidence interval. Solid line indicates the actual measured disparity.

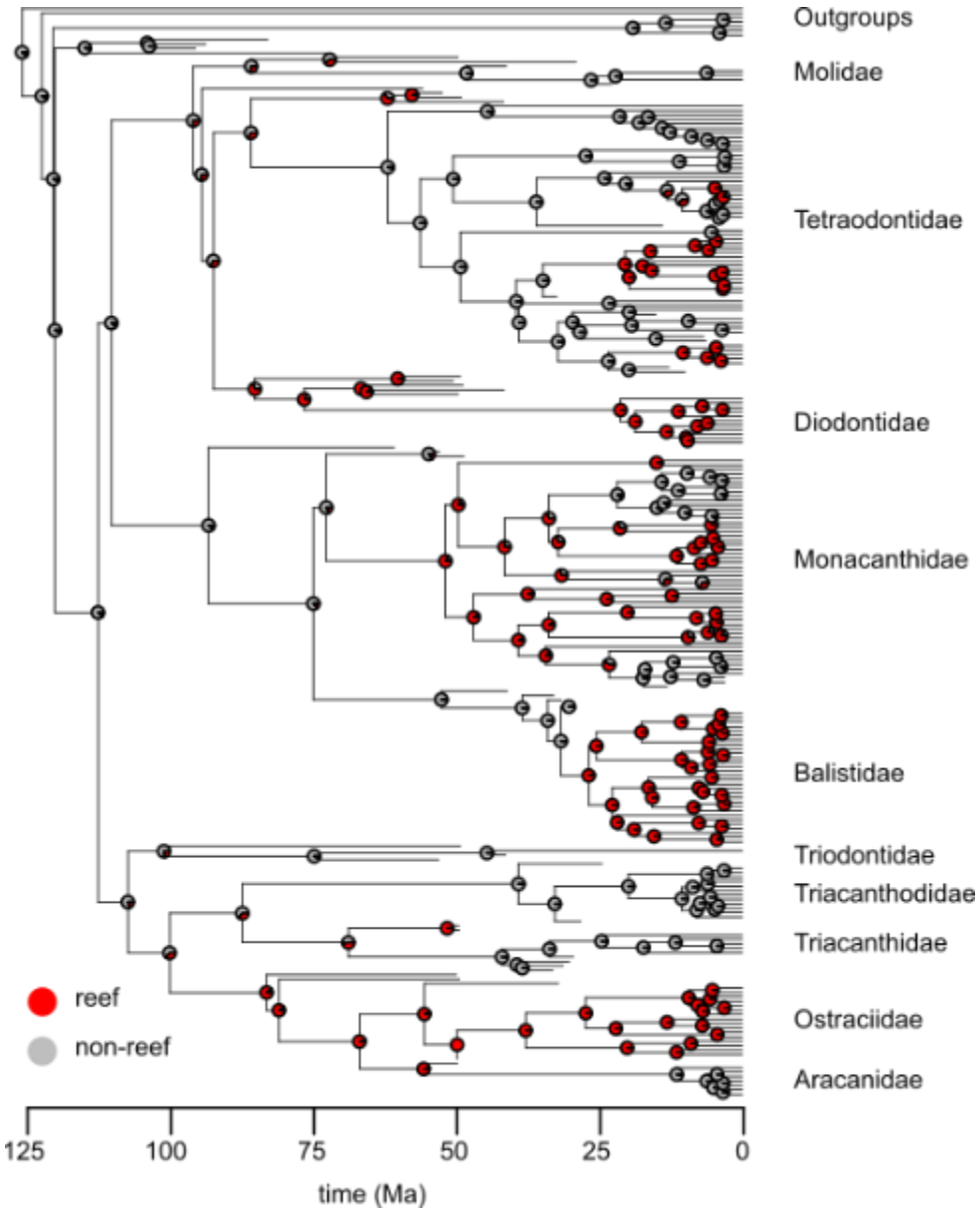


Figure S4. **Ancestral state reconstruction of habitat type across Tetraodontiformes.** Pie charts at nodes represent posterior probability of habitat state (red = reef, grey = non-reef).

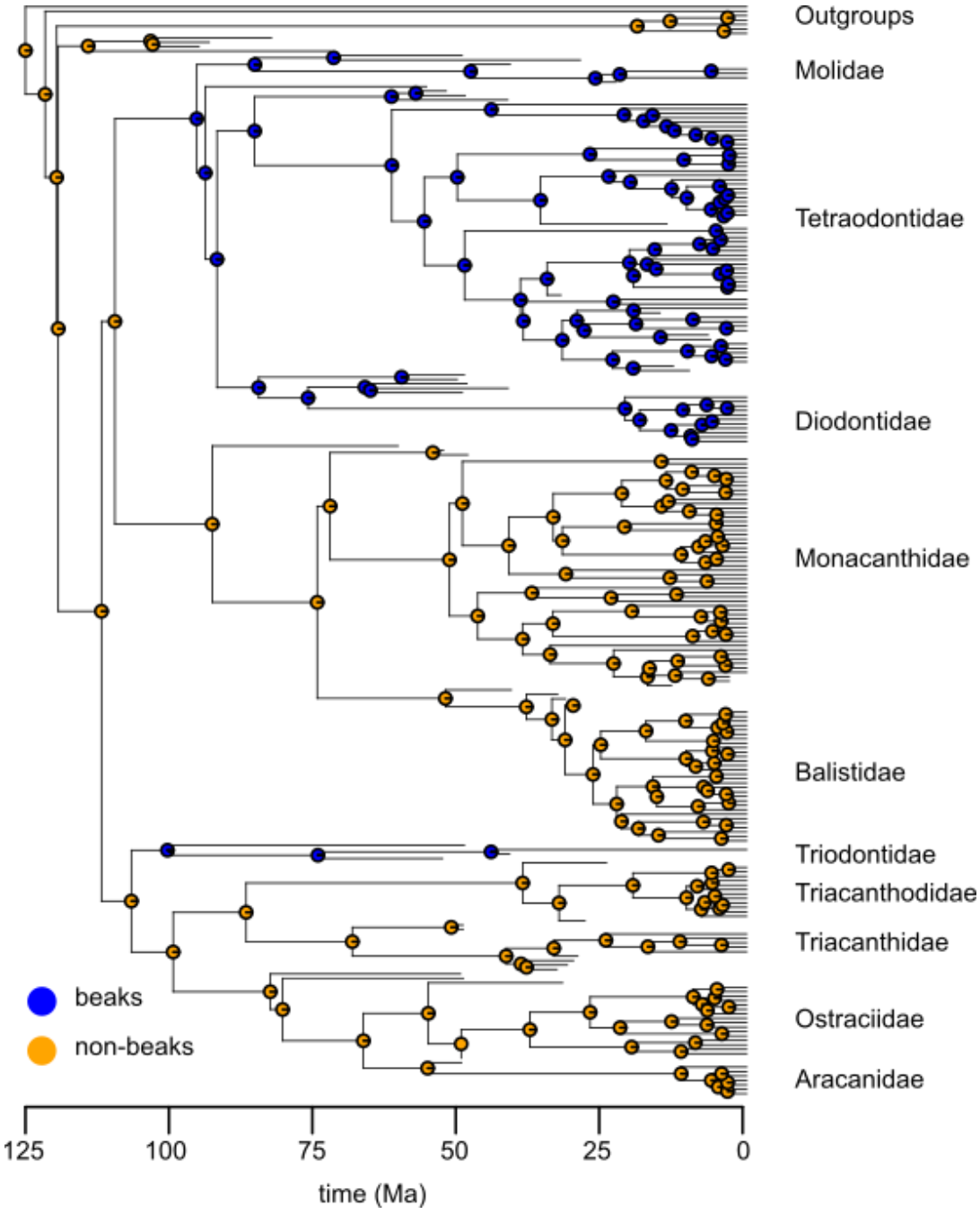
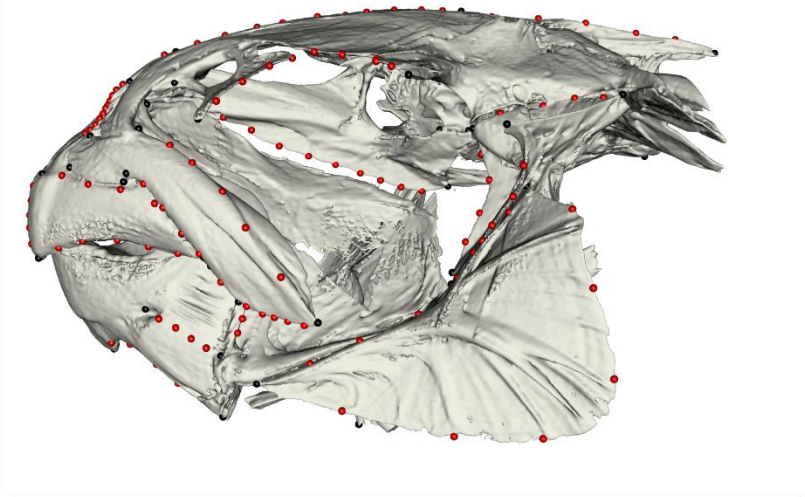
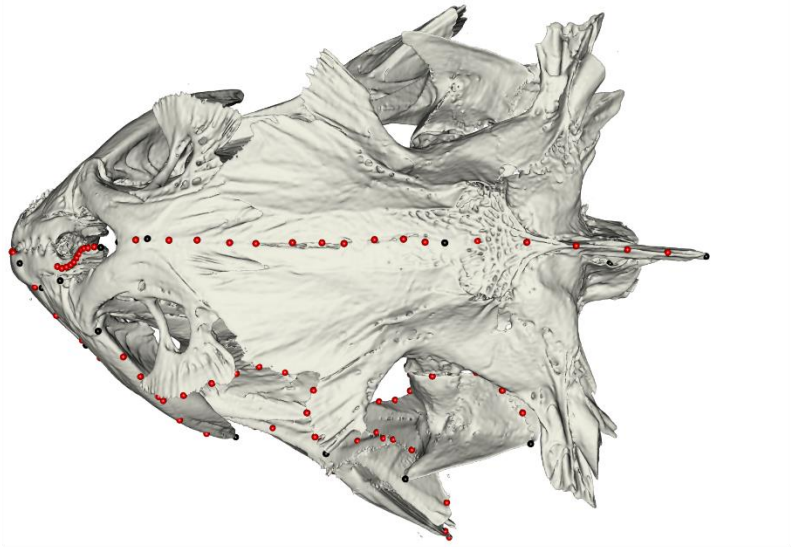


Figure S5. Ancestral state reconstruction of mouth type across Tetraodontiformes. Pie charts at nodes represent posterior probability of mouth state (blue = beaked, orange = non-beaked).

A



B



C

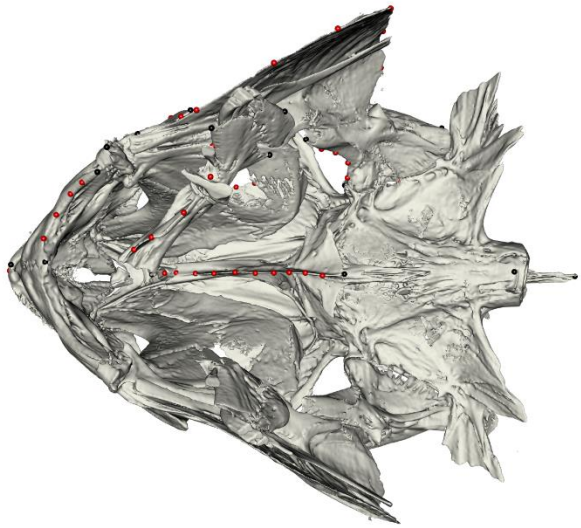


Figure S6. Landmark scheme used in this study showing 48 fixed (black) and 122 semi-sliding landmarks (red) on representative skull of *Marilyna darwinii* for **A)** left lateral view, **B)** dorsal view, and **C)** ventral view.

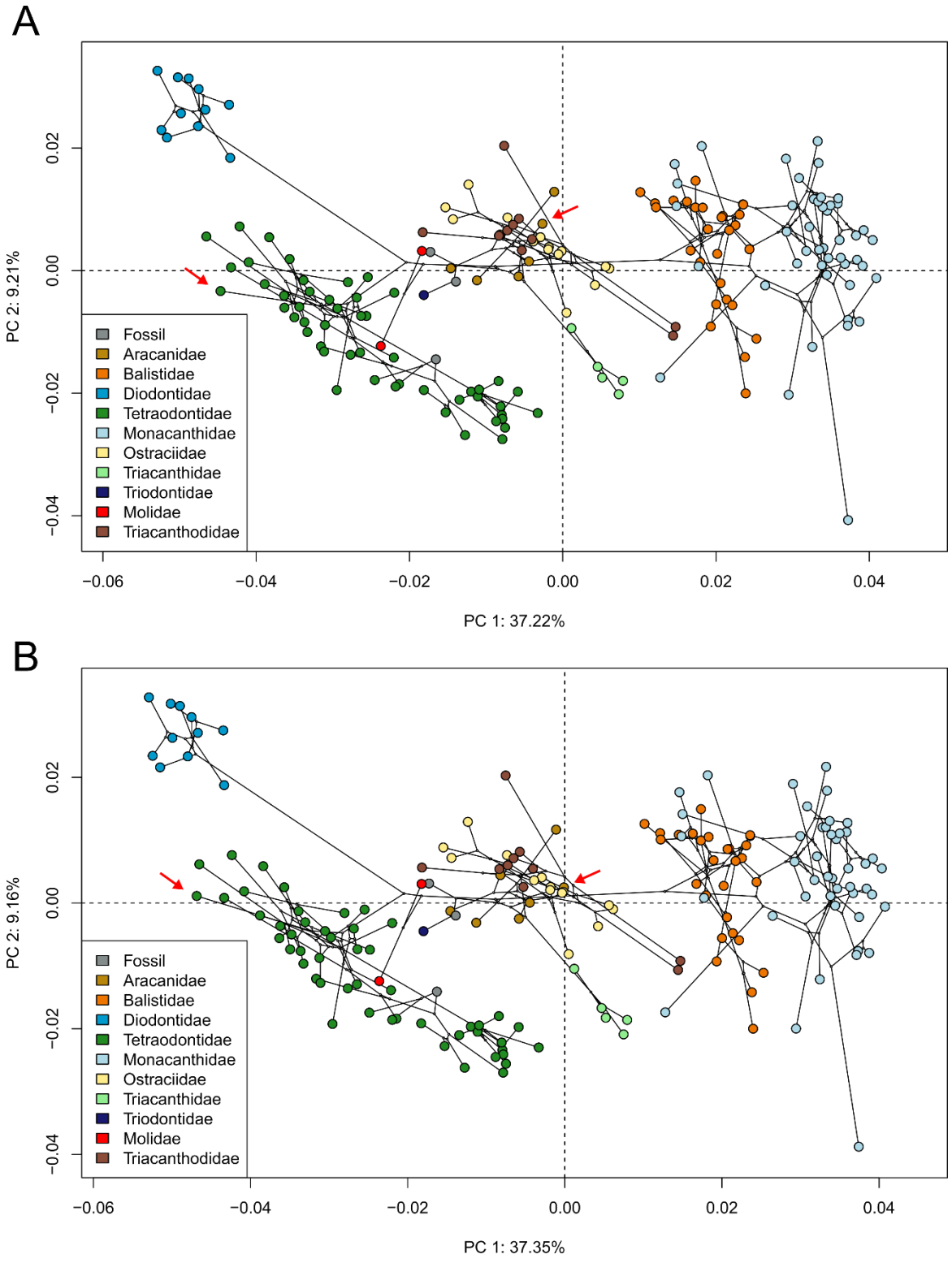
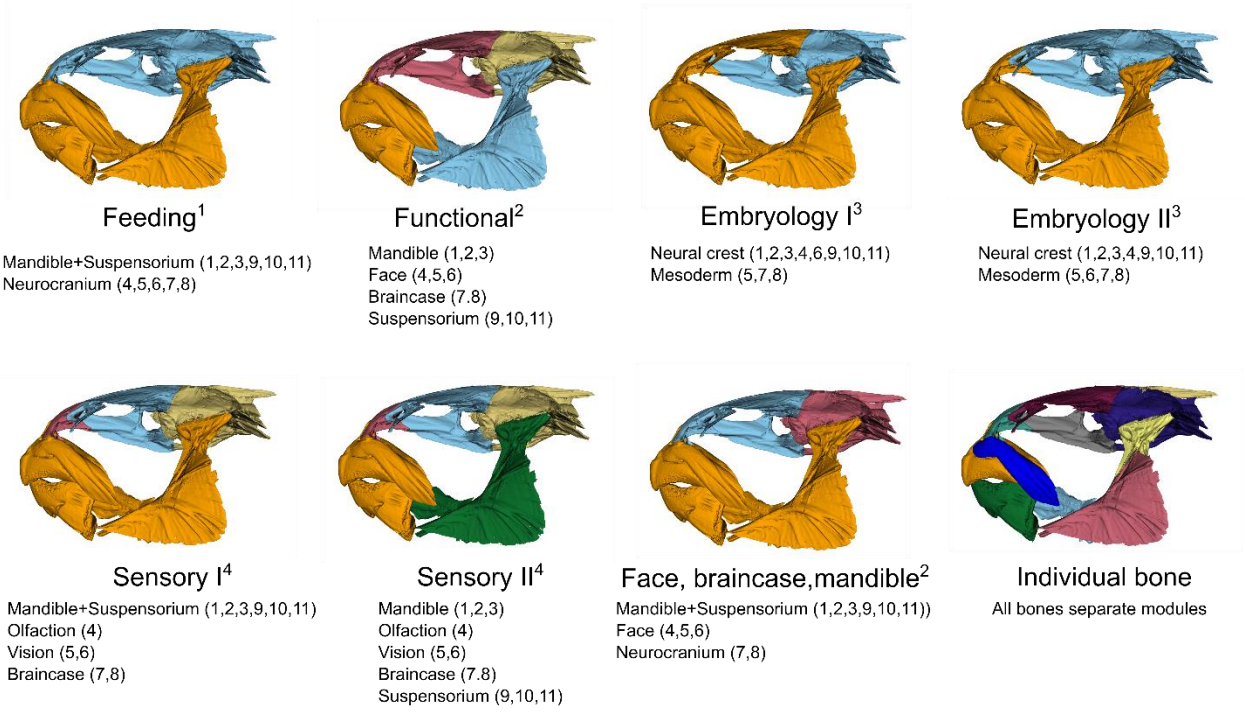

A
PC 2: 9.21%
PC 1: 37.22%
Fossil
Aracanidae
Balistidae
Diodontidae
Tetraodontidae
Monacanthidae
Ostraciidae
Triacanthidae
Triodontidae
Molidae
Triacanthodidae
B
PC 2: 9.16%
PC 1: 37.35%

Figure S7. Morphospace reproducibility check. Red arrows point to species selected for reproducibility checks (one tetraodontid, dark green; one aracanid, light brown). **A)** Original morphospace. **B)** Morphospace after relandmarking two randomly selected species.



Skull Regions

1. premaxilla	4. vomer	7. supraoccipital	10. preopercle
2. maxilla	5. parasphenoid	8. basioccipital	11. ceratohyal
3. dentary	6. frontal	9. hyomandibula	

Figure S8. Eight hypotheses of modularity used in this study. For each hypothesis, the specific modules, and the skull regions they encompass, are listed. Each hypothesis, with the exception of the individual bone module, is referenced from a past study, indicated in superscript.

Table S1. Log marginal likelihoods and Bayes factors for the single rate Brownian motion model and the Variable rates model used to analyze trait evolution of the tetraodontiform skull shape dataset in BayesTraitsv4. A Bayes factor greater than 10 indicates strong support for that model. Best supported model is shown in bold.

Model	Log marginal likelihood	Log Bayes factor
Brownian motion	-13542.09	-2061.9
Variable rates	**-12511.14**	**2061.9**

Table S2. Result of covariance ratio (CR) analyses for all modularity hypotheses. Shown are the effect sizes, CR scores, and p values for each of the eight modularity hypotheses tested. More negative effect sizes indicate a better fit and stronger modular signal. A CR score closer to 1 represents a weaker modular signal. Best supported modular hypothesis is shown in bold.

Model	Effect size	CR score	P value
Feeding	-2.33	0.915	0.01
FBM	-3.95	0.826	0.001
Embryology I	-4.70	0.808	0.001
Embryology II	-2.21	0.914	0.017
Sensory I	-2.11	0.811	0.024
Sensory II	-2.40	0.824	0.014
Functional	-3.64	0.870	0.001
Individual bone	**-4.91**	**0.745**	**0.001**

Table S3. Landmark scheme with description of landmarks and cranial regions. Sliding landmarks indicated as curves.

Landmark	Region	Description
1	Premaxilla	Anterior of skull, tip of first tooth
2	Premaxilla	Base of first tooth
3	Premaxilla	Top of premaxilla
4	Premaxilla	In the crook of the L of the premaxilla
5	Premaxilla	Far posterior point of premaxilla
6	Dentary	Lower jaw tip of front tooth
7	Dentary	Posterior dorsal tip of dentary
8	Dentary	Ventral medial anterior of dentary
9	Dentary	Articulation of dentary and articular
10	Dentary	Posteriorly-oriented projection of peak of articular
11	Dentary	Posterior lower tip of dentary
12	Dentary	Threeway point between articular, retroarticular, and preopercle (jaw joint)
13	Dentary	Bottom posterior of retroarticular
14	Vomer	Anterior center of toothy vomer plate, on anterior end of parasphenoid
15	Vomer	Distal anterior tip of vomer plate
16	Ethmoid	Posterior proximal tip of lateral ethmoid right next to parasphenoid
17	Ethmoid	Articulation of distal lateral ethmoid tip and suborbit
18	Orbit/Frontal	Articulation of lateral ethmoid and frontal bone/on anterior tip of orbit
19	Ethmoid	Proximal most point of lateral ethmoid/frontal margin
20	Supraoccipital	Anterior medial fusion of parietals before ridge/origin of supraoccipital
21	Orbit/Frontal	Posterior tip of orbit, just over the hyomandibula
22	Parasphenoid	Crook of parasphenoid where extends dorsally to contact frontal
23	Parasphenoid	Medial margin between parasphenoid and basioccipital
24	Basioccipital/Prootic	Prootic foramen
25	Supraoccipital	Posterior end of supraoccipital crest
26	Basioccipital/Prootic	Dorsal point of articulation between basioccipital and first vertebrae
27	Basioccipital/Prootic	Ventral point of articulation between basioccipital and first vertebrae
Curve 1: 28-38	Premaxilla	Ascending process of premaxilla- landmarks 2 to 3 (11 points)
Curve 2: 39-42	Dentary	Ventral margin of dentary- landmarks 8 to 11 (4 points)
Curve 3: 43-52	Parasphenoid	Ventral margin of parasphenoid- landmarks 14 to 23 (10 points)
Curve 4: 53-61	Orbit/Frontal	Distal rim of upper orbit- landmarks 18 to 21 (9 points)
62	Maxilla	Distal, dorsal anterior margin of maxilla
63	Maxilla	Distal, dorsal posterior margin of maxilla
64	Maxilla	Distal, ventral anterior margin of maxilla
65	Maxilla	Distal, ventral posterior margin of maxilla
66	Maxilla	Proximal, dorsal anterior margin of maxilla

67	Maxilla	Proximal, dorsal posterior margin of maxilla
68	Maxilla	Anterior most point of the proximal descending process of the maxilla
69	Hyomandibula	Anterior point of contact of distal face of hyomandibula with sphenotic
70	Hyomandibula	Posterior point of contact of distal face of hyomandibula with pterotic bone
71	Hyomandibula	Distal most point of lateral projection/ridge of hyomandibula
72	Hyomandibula	Anterior-ventral most point of hyomandibula below landmark 71
73	Hyomandibula	Ventral-most point of hyomandibula
Curve 5: 74-77	Angular	Angular ascending process- landmarks 10 to 12 (4 points)
Curve 6: 78-81	Angular	Angular lateral process- landmarks 9 to 12 (4 points)
Curve 7: 82-86	Premaxilla	Lateral arm premaxilla- landmarks 2 to 5 (5 points)
Curve 8: 87-90	Hyomandibula	Anterior face hyomandibula- landmarks 69 to 72 (4 points)
Curve 9: 91-94	Hyomandibula	Dorsal surface hyomandibula- landmarks 69 to 70 (4 points)
Curve 10: 95-98	Hyomandibula	Posterior face hyomandibula- landmarks 71 to 73 (4 points)
99	Ceratohyal	Most anterior point of ceratohyal
100	Ceratohyal	Interior ridge of ceratohyal
101	Ceratohyal	Posterior dorsal tip of "triangle" shape at back of ceratohyal
102	Ceratohyal	Most posterior point of epihyal
103	Ceratohyal	Posterior ventral tip of "triangle" shape at back of ceratohyal
Curve 11: 104-108	Ceratohyal	Ceratohyal dorsal ridge- landmarks 99 to 101 (5 points)
Curve 12: 109-113	Ceratohyal	Ceratohyal ventral ridge- landmarks 99 to 103 (5 points)
Curve 13: 114-118	Supraoccipital	Supraoccipital crest- landmarks 20 to 25 (5 points)
119	Ethmoid	Center-anterior most point of ethmoid
Curve 14: 120-130	Orbit/Frontal	Central axis of cranium- landmarks 119 to 20 (11 points)
131	Preopercle	Anterior most point of preopercle
132	Preopercle	Posterior most point of preopercle
Curve 15: 133-138	Preopercle	Dorsal curve of preopercle- landmarks 131 to 132 (6 points)
Curve 16: 139-144	Preopercle	Ventral curve of preopercle- landmarks 131 to 132 (6 points)
Curve 17: 145-149	Maxilla	Anterior curve of maxilla- landmarks 62 to 64 (5 points)
Curve 18: 150-154	Maxilla	Posterior curve of maxilla- landmarks 63 to 65 (5 points)
Curve 19: 155-159	Maxilla	Ventral curve of maxilla- landmarks 64 to 65 (5 points)
Curve 20: 160-164	Premaxilla	Proximal margin of first tooth- landmarks 2 to 1 (5 points)
165	Premaxilla	Dorsal distal edge of first tooth
Curve 21: 166-170	Premaxilla	Distal margin of first tooth- landmarks 1 to 165 (5 points)

Supplemental References

1. Westneat, M. W. Skull Biomechanics and Suction Feeding in Fishes. in *Fish Physiology* vol. 23 29–75 (Elsevier, 2005).
2. Evans, K. M., Vidal-García, M., Tagliacollo, V. A., Taylor, S. J. & Fenolio, D. B. Bony Patchwork: Mosaic Patterns of Evolution in the Skull of Electric Fishes (Apteronotidae: Gymnotiformes). *Integr. Comp. Biol.* **59**, 420–431 (2019).
3. Kague, E., Gallagher, M., Burke, S., Parsons, M., Franz-Odendall, T., Fisher, S. Skeletogenic Fate of Zebrafish Cranial and Trunk Neural Crest. *PLoS ONE* **7**, e47394 (2012).
4. Helfman, G., Collette, B., Facey, D. E., Bowen, B. W. *The diversity of fishes: biology, evolution, and ecology*. (Blackwell, 2009).

Appendix 3

Supplemental Materials for Chapter 3:

They might not be giants: Genetic convergence on miniaturization in gobiid fishes

Emily M. Troyer[1], Dahiana Arcila[2]

[1]*Department of Biology and Sam Noble Oklahoma Museum of Natural History, University of Oklahoma, 730 Van Vleet Oval, Richards Hall, Norman, OK 73019, USA*

[2]*Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA*

Appendix 3 Includes:

Figures S1 to S4
Tables S1 to S8

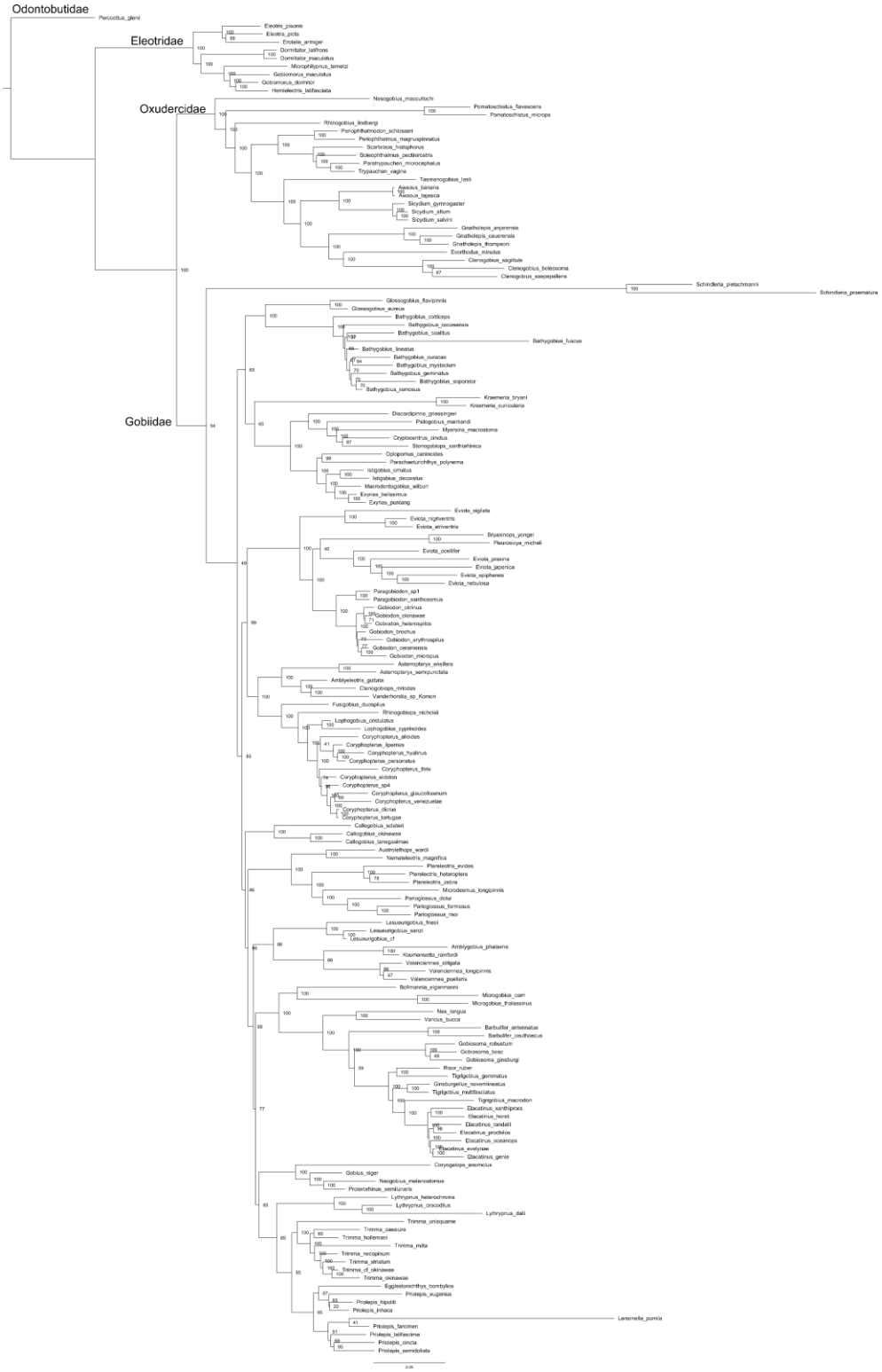


Figure S1. Maximum likelihood phylogeny of 165 species of Gobiodei based on concatenation analysis of 980 exons and inferred using RAxML-ng and partitioned by codon. Nodal values indicate bootstrap support values based on 1,000 replicates. Branch lengths represent the number of nucleotide substitutions per site, indicated by the scale bar.

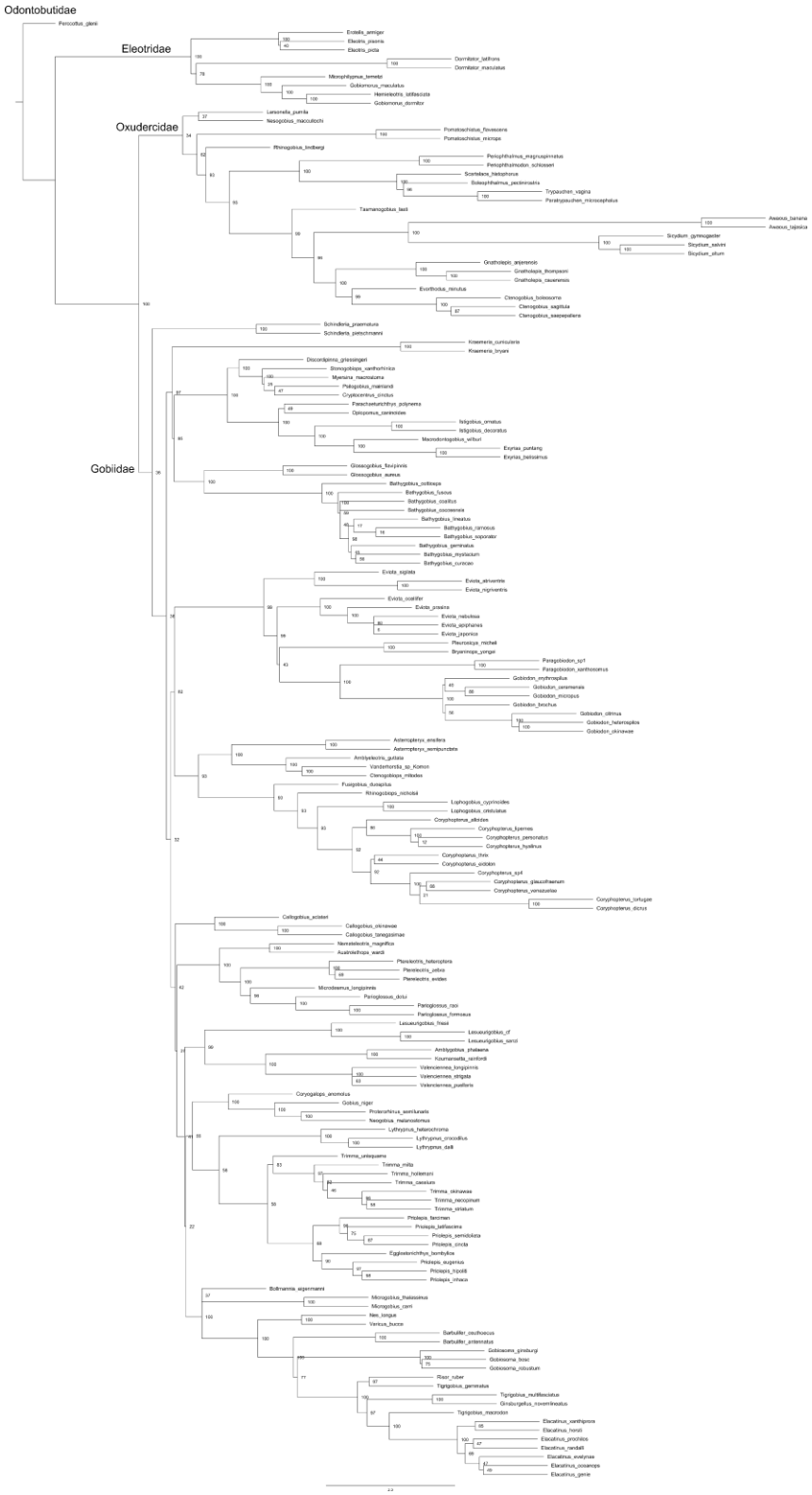


Figure S2. Cladogram phylogeny of 165 Gobioidei species based on multi-species coalescent analysis of 980 exons inferred in ASTRAL. Nodal values indicate bootstrap support estimated using 1,000 replicates.

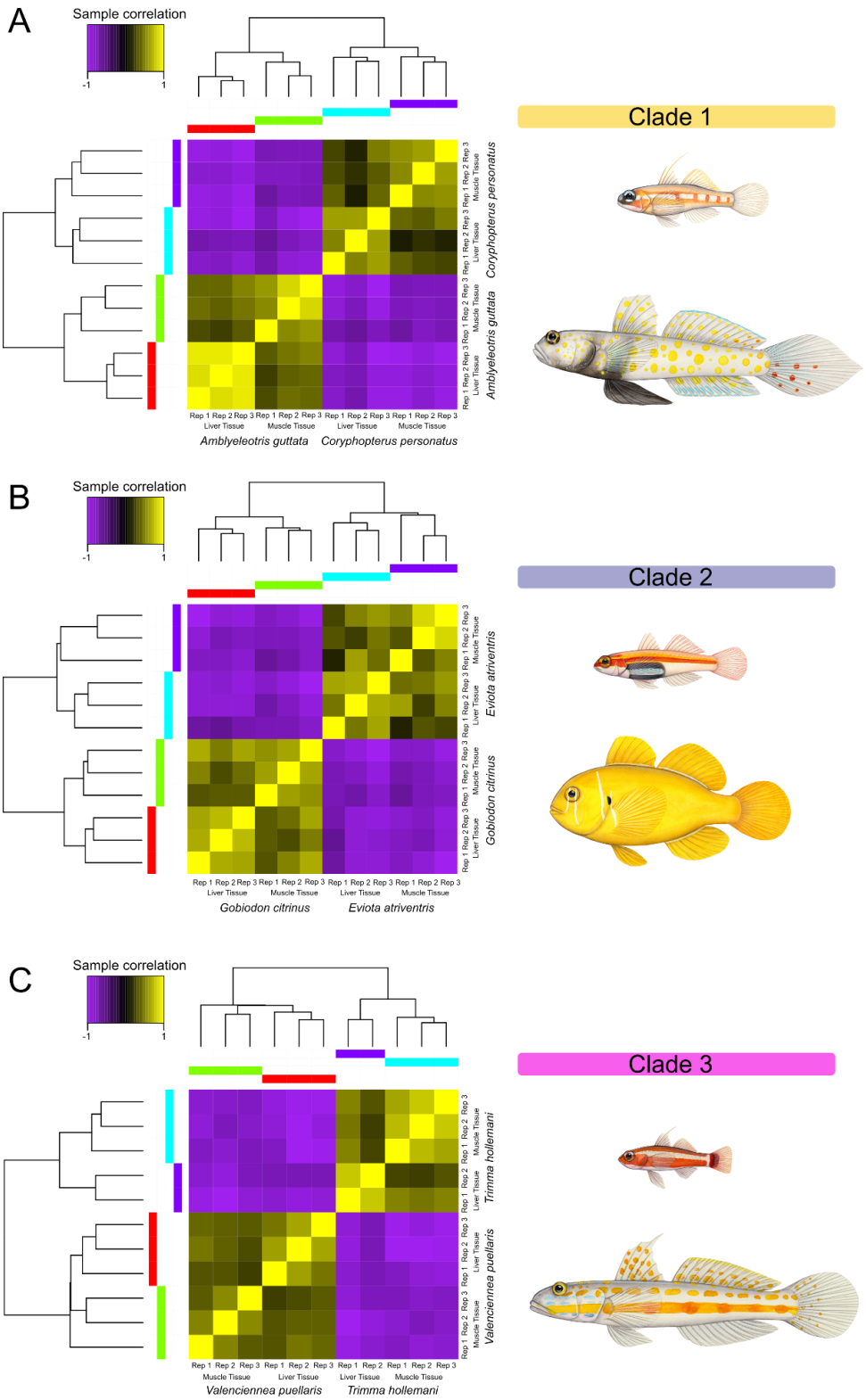


Figure S3. Heatmaps based on hierarchical cluster analyses depicting sample correlations for each tissue sample (muscle and liver) for three clades of gobiid species containing a closely related large-bodied and small-bodied species. A) Clade 1: *Amblyeleotris guttata* and *Coryphopterus personatus*, B) Clade 2: *Eviota atriventris* and *Gobiodon citrinus*, and C) Clade 3: *Trimma hollemani* and *Valenciennea puellaris.* Illustrations by Julie Johnson.

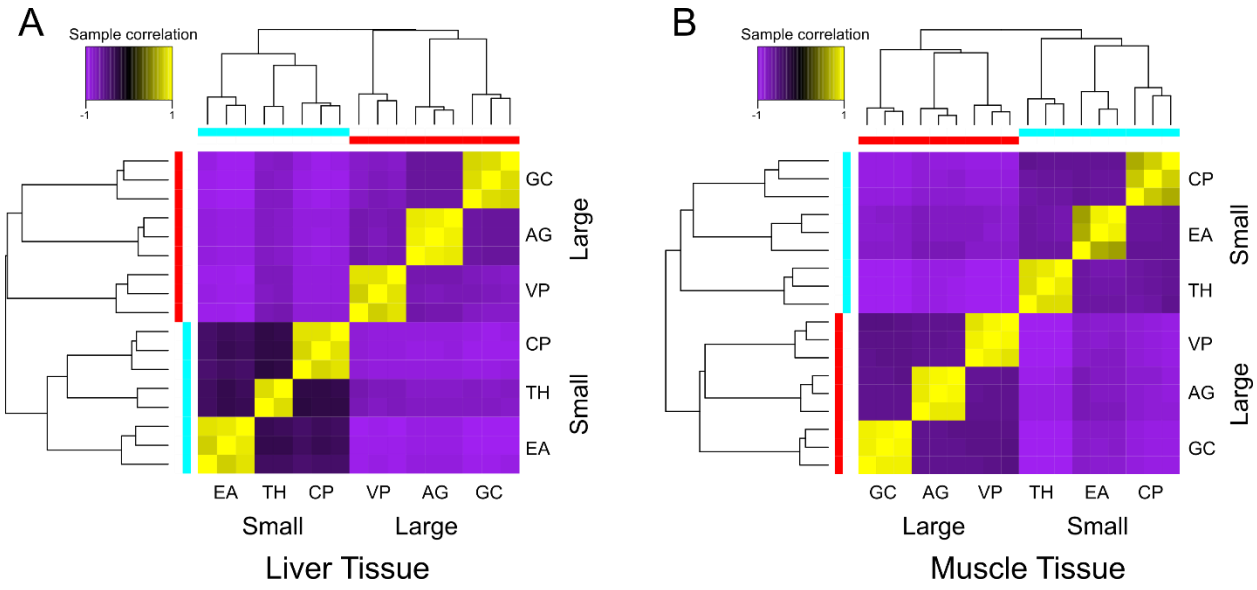


Figure S4. Heatmaps based on hierarchical cluster analyses depicting sample correlation matrices for six species of gobiids, including three large-bodied and three small-bodied species for A) liver tissue and B) skeletal muscle tissue. Species abbreviations are AG: *Amblyeleotris guttata*; CP: *Coryphopterus personatus*; EA: *Eviota atriventris*; GC: *Gobiodon citrinus*; TH: *Trimma hollemani*; and VP: *Valenciennea puellaris*.

Table S1. Mapping results from the reference-based transcriptome assembly using the closest available reference genome, *Boleophthalmus pectinirostris*, the great blue-spotted mudskipper.

Species	Replicate Number	Tissue Type	Sample Name	Sample Batch	Total Paired Reads	Total Mapped Reads
Amblyeleotris guttata	Rep 1	Liver	AG1LIV	2	54535844	609145 (1.12%)
Amblyeleotris guttata	Rep 1	Muscle	AG1MUS	2	42664314	929187 (2.18%)
Amblyeleotris guttata	Rep 2	Liver	AG2LIV	2	53641650	867374 (1.62%)
Amblyeleotris guttata	Rep 2	Muscle	AG2MUS	2	52167206	3680388 (7.05%)
Amblyeleotris guttata	Rep 3	Liver	AG3LIV	2	40599514	672284 (1.66%)
Amblyeleotris guttata	Rep 3	Muscle	AG3MUS	2	46049726	3554770 (7.72%)
Coryphopterus personatus	Rep 1	Liver	CP1LIV	4	46258910	807007 (1.74%)
Coryphopterus personatus	Rep 1	Muscle	CP1MUS	4	42020522	3317367 (7.89%)
Coryphopterus personatus	Rep 2	Liver	CP2LIV	4	46176486	842179 (1.82%)
Coryphopterus personatus	Rep 2	Muscle	CP2MUS	4	58405918	6476804 (11.09%)
Coryphopterus personatus	Rep 3	Liver	CP3LIV	4	41527944	4739604 (11.41%)
Coryphopterus personatus	Rep 3	Muscle	CP3MUS2	4	44678018	2164166 (4.84%)
Eviota atriventris	Rep 1	Liver	EA1LIV	5	45551830	371082 (0.81%)
Eviota atriventris	Rep 1	Muscle	EA1MUS	5	48698616	750234 (1.54%)
Eviota atriventris	Rep 2	Muscle	EA2MUS	5	43254500	1822490 (4.21%)
Eviota atriventris	Rep 2	Liver	EA4LIV	7	41624070	317043 (0.76%)
Eviota atriventris	Rep 3	Muscle	EA4MUS	7	51440866	3385570 (6.58%)
Eviota atriventris	Rep 3	Liver	EA5LIV	7	46264854	466576 (1.01%)
Gobiodon citrinus	Rep 1	Liver	GC1LIV	3	48441004	668385 (1.38%)
Gobiodon citrinus	Rep 1	Muscle	GC1MUS	3	47333204	2027812 (4.28%)
Gobiodon citrinus	Rep 2	Liver	GC2LIV	4	41068662	1136580 (2.77%)
Gobiodon citrinus	Rep 2	Muscle	GC2MUS	4	57515632	1202882 (2.09%)
Gobiodon citrinus	Rep 3	Liver	GC3LIV	4	43056264	1025055 (2.38%)
Gobiodon citrinus	Rep 3	Muscle	GC3MUS2	4	43804724	1736944 (3.97%)
Trimma hollemani	Rep 1	Muscle	TH1MUS	5	52844800	3597917 (6.81%)
Trimma hollemani	Rep 2	Liver	TH2LIV	5	46833976	1052805 (2.25%)
Trimma hollemani	Rep 2	Muscle	TH2MUS	5	54741102	5374927 (9.82%)
Trimma hollemani	Rep 3	Liver	TH3LIV	5	41757644	2060978 (4.94%)
Trimma hollemani	Rep 3	Muscle	TH3MUS	5	41307108	2003921 (4.85%)
Valenciennea puellaris	Rep 1	Liver	VP1LIVER	1	49488552	1341417 (2.71%)
Valenciennea puellaris	Rep 1	Muscle	VP1MUSC	1	48030550	3996925 (8.32%)
Valenciennea puellaris	Rep 2	Liver	VP2LIV	2	43692562	496434 (1.14%)
Valenciennea puellaris	Rep 2	Muscle	VP2MUS	2	46757462	1567974 (3.35%)
Valenciennea puellaris	Rep 3	Muscle	VP3MUS	6	44351946	2002948 (4.52%)
Valenciennea puellaris	Rep 3	Liver	VP4LIV	8	44327074	428767 (0.97%)

Table S2. Total body lengths recorded from FishBase for all species in the phylogenetic analysis. Species are categorized according to their affiliation (either outgroup or focal clade) and family.

Affiliation	Family	Species	Total Length (cm)
Outgroup	Eleotridae	*Dormitator latifrons*	41
Outgroup	Eleotridae	*Dormitator maculatus*	70
Outgroup	Eleotridae	*Eleotris picta*	55.2262
Outgroup	Eleotridae	*Eleotris pisonis*	25
Outgroup	Eleotridae	*Erotelis armiger*	9.0943
Outgroup	Eleotridae	*Gobiomorus dormitor*	90
Outgroup	Eleotridae	*Gobiomorus maculatus*	33.8572
Outgroup	Eleotridae	*Hemieleotris latifasciata*	12
Outgroup	Eleotridae	*Microphilypnus ternetzi*	2.4322
Gobiidae	Gobiidae	*Amblyeleotris guttata*	13.7452
Gobiidae	Gobiidae	*Amblygobius phalaena*	15
Gobiidae	Gobiidae	*Asterropteryx ensifera*	3.6892
Gobiidae	Gobiidae	*Asterropteryx semipunctata*	6.5
Gobiidae	Gobiidae	*Austrolethops wardi*	6
Gobiidae	Gobiidae	*Barbulifer antennatus*	2.4322
Gobiidae	Gobiidae	*Barbulifer ceuthoecus*	3
Gobiidae	Gobiidae	*Bathygobius coalitus*	12.4882
Gobiidae	Gobiidae	*Bathygobius cocosensis*	12
Gobiidae	Gobiidae	*Bathygobius cotticeps*	11
Gobiidae	Gobiidae	*Bathygobius curacao*	7.5
Gobiidae	Gobiidae	*Bathygobius fuscus*	12
Gobiidae	Gobiidae	*Bathygobius geminatus*	4.6948
Gobiidae	Gobiidae	*Bathygobius lineatus*	12
Gobiidae	Gobiidae	*Bathygobius mystacium*	15
Gobiidae	Gobiidae	*Bathygobius ramosus*	20
Gobiidae	Gobiidae	*Bathygobius soporator*	15
Gobiidae	Gobiidae	*Bollmannia eigenmanni*	17.5162
Gobiidae	Gobiidae	*Bryaninops yongei*	4.3177
Gobiidae	Gobiidae	*Callogobius okinawae*	6.2032
Gobiidae	Gobiidae	*Callogobius sclateri*	7
Gobiidae	Gobiidae	*Callogobius tanegasimae*	6.8317
Gobiidae	Gobiidae	*Coryogalops anomolus*	5.8
Gobiidae	Gobiidae	*Coryphopterus alloides*	4
Gobiidae	Gobiidae	*Coryphopterus dicrus*	5
Gobiidae	Gobiidae	*Coryphopterus eidolon*	6
Gobiidae	Gobiidae	*Coryphopterus glaucofraenum*	8
Gobiidae	Gobiidae	*Coryphopterus hyalinus*	2.5
Gobiidae	Gobiidae	*Coryphopterus lipernes*	3
Gobiidae	Gobiidae	*Coryphopterus personatus*	4
Gobiidae	Gobiidae	*Coryphopterus* sp4	NA
Gobiidae	Gobiidae	*Coryphopterus thrix*	5
Gobiidae	Gobiidae	*Coryphopterus tortugae*	5.0719

Gobiidae	Gobiidae	*Coryphopterus venezuelae*	7.5
Gobiidae	Gobiidae	*Cryptocentrus cinctus*	12.4882
Gobiidae	Gobiidae	*Ctenogobiops mitodes*	6.5803
Gobiidae	Gobiidae	*Discordipinna griessingeri*	3.6892
Gobiidae	Gobiidae	*Egglestonichthys bombylios*	4.762678
Gobiidae	Gobiidae	*Elacatinus evelynae*	4
Gobiidae	Gobiidae	*Elacatinus genie*	4.5
Gobiidae	Gobiidae	*Elacatinus horsti*	5
Gobiidae	Gobiidae	*Elacatinus oceanops*	5
Gobiidae	Gobiidae	*Elacatinus prochilos*	4
Gobiidae	Gobiidae	*Elacatinus randalli*	4.6
Gobiidae	Gobiidae	*Elacatinus xanthiprora*	4
Gobiidae	Gobiidae	*Eviota atriventris*	2.0551
Gobiidae	Gobiidae	*Eviota epiphanes*	1.678
Gobiidae	Gobiidae	*Eviota japonica*	2.935
Gobiidae	Gobiidae	*Eviota nebulosa*	2.3065
Gobiidae	Gobiidae	*Eviota nigriventris*	3.6892
Gobiidae	Gobiidae	*Eviota ocellifer*	2.1808
Gobiidae	Gobiidae	*Eviota prasina*	3.8149
Gobiidae	Gobiidae	*Eviota sigilata*	3.6892
Gobiidae	Gobiidae	*Exyrias belissimus*	18.7732
Gobiidae	Gobiidae	*Exyrias puntang*	16.2
Gobiidae	Gobiidae	*Fusigobius duospilus*	6
Gobiidae	Gobiidae	*Ginsburgellus novemlineatus*	2.5
Gobiidae	Gobiidae	*Glossogobius aureus*	31.3432
Gobiidae	Gobiidae	*Glossogobius flavipinnis*	8
Gobiidae	Gobiidae	*Gobiodon brochus*	3.1864
Gobiidae	Gobiidae	*Gobiodon ceramensis*	3.5
Gobiidae	Gobiidae	*Gobiodon citrinus*	6.6
Gobiidae	Gobiidae	*Gobiodon erythrospilus*	5.4
Gobiidae	Gobiidae	*Gobiodon heterospilos*	6.6
Gobiidae	Gobiidae	*Gobiodon micropus*	3.5
Gobiidae	Gobiidae	*Gobiodon okinawae*	3.5
Gobiidae	Gobiidae	*Gobiosoma bosc*	6
Gobiidae	Gobiidae	*Gobiosoma ginsburgi*	6
Gobiidae	Gobiidae	*Gobiosoma robustum*	5
Gobiidae	Gobiidae	*Gobius niger*	25.3
Gobiidae	Gobiidae	*Istigobius decoratus*	13
Gobiidae	Gobiidae	*Istigobius ornatus*	11
Gobiidae	Gobiidae	*Koumansetta rainfordi*	10.6027
Gobiidae	Gobiidae	*Kraemeria bryani*	2.46991
Gobiidae	Gobiidae	*Kraemeria cunicularia*	4.5
Gobiidae	Gobiidae	*Larsonella pumila*	2.3065
Gobiidae	Gobiidae	*Lesueurigobius* cf.	NA
Gobiidae	Gobiidae	*Lesueurigobius friesii*	13

Gobiidae	Gobiidae	*Lesueurigobius sanzi*	12.3
Gobiidae	Gobiidae	*Lophogobius cristulatus*	7.4
Gobiidae	Gobiidae	*Lophogobius cyprinoides*	10
Gobiidae	Gobiidae	*Lythrypnus crocodilus*	2.57047
Gobiidae	Gobiidae	*Lythrypnus dalli*	6.4
Gobiidae	Gobiidae	*Lythrypnus heterochroma*	2.5
Gobiidae	Gobiidae	*Macrodontogobius wilburi*	8.0887
Gobiidae	Gobiidae	*Microdesmus longipinnis*	27
Gobiidae	Gobiidae	*Microgobius carri*	12.4882
Gobiidae	Gobiidae	*Microgobius thalassinus*	6.4
Gobiidae	Gobiidae	*Myersina macrostoma*	6
Gobiidae	Gobiidae	*Nemateleotris magnifica*	9
Gobiidae	Gobiidae	*Neogobius melanostomus*	35
Gobiidae	Gobiidae	*Nes longus*	10
Gobiidae	Gobiidae	*Nesogobius maccullochi*	8.5915
Gobiidae	Gobiidae	*Oplopomus caninoides*	9.3457
Gobiidae	Gobiidae	*Parachaeturichthys polynema*	15
Gobiidae	Gobiidae	*Paragobiodon* sp1	NA
Gobiidae	Gobiidae	*Paragobiodon xanthosomus*	4
Gobiidae	Gobiidae	*Parioglossus dotui*	4.9462
Gobiidae	Gobiidae	*Parioglossus formosus*	5.5747
Gobiidae	Gobiidae	*Parioglossus raoi*	4.3177
Gobiidae	Gobiidae	*Pleurosicya micheli*	2.5
Gobiidae	Gobiidae	*Priolepis cincta*	7
Gobiidae	Gobiidae	*Priolepis eugenius*	6.9574
Gobiidae	Gobiidae	*Priolepis farcimen*	2.3065
Gobiidae	Gobiidae	*Priolepis hipoliti*	4.6
Gobiidae	Gobiidae	*Priolepis inhaca*	4
Gobiidae	Gobiidae	*Priolepis latifascima*	2.5579
Gobiidae	Gobiidae	*Priolepis semidoliata*	2.935
Gobiidae	Gobiidae	*Proterorhinus semilunaris*	11.2312
Gobiidae	Gobiidae	*Psilogobius mainlandi*	4.5691
Gobiidae	Gobiidae	*Ptereleotris evides*	14
Gobiidae	Gobiidae	*Ptereleotris heteroptera*	14
Gobiidae	Gobiidae	*Ptereleotris zebra*	15.0022
Gobiidae	Gobiidae	*Rhinogobiops nicholsii*	15
Gobiidae	Gobiidae	*Risor ruber*	2.5
Gobiidae	Gobiidae	*Schindleria pietschmanni*	2.0551
Gobiidae	Gobiidae	*Schindleria praematura*	2.5
Gobiidae	Gobiidae	*Stonogobiops xanthorhinica*	7.4602
Gobiidae	Gobiidae	*Tasmanogobius lasti*	5.5
Gobiidae	Gobiidae	*Tigrigobius gemmatus*	2.5
Gobiidae	Gobiidae	*Tigrigobius macrodon*	5
Gobiidae	Gobiidae	*Tigrigobius multifasciatus*	5
Gobiidae	Gobiidae	*Trimma caesiura*	4.3177

Gobiidae	Gobiidae	*Trimma hollemani*	3.4378
Gobiidae	Gobiidae	*Trimma milta*	3
Gobiidae	Gobiidae	*Trimma necopinum*	4
Gobiidae	Gobiidae	*Trimma okinawae*	4.3177
Gobiidae	Gobiidae	*Trimma striatum*	3
Gobiidae	Gobiidae	*Trimma unisquame*	3.1864
Gobiidae	Gobiidae	*Valenciennea longipinnis*	22.5442
Gobiidae	Gobiidae	*Valenciennea puellaris*	25.0582
Gobiidae	Gobiidae	*Valenciennea strigata*	18
Gobiidae	Gobiidae	*Vanderhorstia sp. 'Komon-yatsushi-haze'*	NA
Gobiidae	Gobiidae	*Varicus bucca*	5.7004
Outgroup	Odontobutidae	*Perccottus glenii*	25
Outgroup	Oxudercidae	*Awaous banana*	37.6282
Outgroup	Oxudercidae	*Awaous tajasica*	20.4073
Outgroup	Oxudercidae	*Boleophthalmus pectinirostris*	17.5
Outgroup	Oxudercidae	*Ctenogobius boleosoma*	7.5
Outgroup	Oxudercidae	*Ctenogobius saepepallens*	5
Outgroup	Oxudercidae	*Ctenogobius sagittula*	20
Outgroup	Oxudercidae	*Evorthodus minutus*	2.688
Outgroup	Oxudercidae	*Gnatholepis anjerensis*	10.477
Outgroup	Oxudercidae	*Gnatholepis cauerensis*	7.4602
Outgroup	Oxudercidae	*Gnatholepis thompsoni*	8.2
Outgroup	Oxudercidae	*Paratrypauchen microcephalus*	18
Outgroup	Oxudercidae	*Periophthalmodon schlosseri*	27
Outgroup	Oxudercidae	*Periophthalmus magnuspinnatus*	12.8653
Outgroup	Oxudercidae	*Pomatoschistus flavescens*	6
Outgroup	Oxudercidae	*Pomatoschistus microps*	9
Outgroup	Oxudercidae	*Rhinogobius lindbergi*	4.90849
Outgroup	Gobiidae	*Scartelaos histophorus*	17.5162
Outgroup	Oxudercidae	*Sicydium altum*	17.5162
Outgroup	Oxudercidae	*Sicydium gymnogaster*	12.065
Outgroup	Oxudercidae	*Sicydium salvini*	17.5162
Outgroup	Oxudercidae	*Trypauchen vagina*	22

Table S3. Sample statistics for 35 tissue samples of six gobiid species from RNA-seq analysis. Sample name, RNA Integrity Number (RIN) score, number of raw read pairs (before trimming of adaptors and low-quality reads), number of trimmed read pairs, percent of read pairs removed after trimming, and the percentage of reads mapping back to the sample in the Salmon analyses are provided. The total number of raw read pairs is provided at the end of the table.

Sample	RIN Score	Raw Read Pairs	Trimmed Read Pairs	Percent Removed	Salmon Percent Mapped
AG1LIV	6.7	27766238	27008708	2.73	97.98
AG1MUS	7.2	21747602	21172358	2.65	98.16
AG2LIV	5.3	27605368	26593983	3.66	97.75
AG2MUS	7	26510511	25868107	2.42	98.45
AG3LIV	5.7	20696320	20132024	2.73	97.72
AG3MUS	7.6	23341984	22865819	2.04	98.11
CP1LIV	5.9	23647953	22915744	3.1	97.01
CP1MUS	5	21400532	20837981	2.63	80.83
CP2LIV	6.1	23638815	22949560	2.92	94.44
CP2MUS	7.2	29873758	28978046	3	94.97
CP3LIV	4.1	21260057	20610333	3.06	94.17
CP3MUS2	5.6	22783985	22166546	2.71	89.55
EA1LIV	4.1	24781557	22574394	8.91	95.1
EA1MUS	5.2	25017113	24191521	3.3	96.4
EA2MUS	7.4	22085471	21493251	2.68	92.34
EA4LIV	7.2	21411020	20512686	4.2	96.19
EA4MUS	8.4	26362449	25333215	3.9	96.88
EA5LIV	6.4	23768933	22813583	4.02	96.1
GC1LIV	5.8	24965070	24039862	3.71	95.9
GC1MUS	7	24101614	23499340	2.5	95.47
GC2LIV	5.7	21392035	20372023	4.77	94.77
GC2MUS	6.6	29451825	28584672	2.94	94.54
GC3LIV	6	22097696	21369389	3.3	96.15
GC3MUS2	6.7	22427257	21728436	3.12	96.13
TH1MUS	7	26850265	26230073	2.31	97.3
TH2LIV	6.9	24084538	23280886	3.34	94.69
TH2MUS	7.2	27926319	27168439	2.71	96.97
TH3LIV	6.5	21540371	20772126	3.57	92.08
TH3MUS	6.2	21089625	20531014	2.65	96.84
VP1LIVER	4.3	25305413	24584148	2.85	97.69
VP1MUSC	7.4	24673068	23818817	3.46	98.43
VP2LIV	6.5	22392468	21680382	3.18	97.56
VP2MUS	6.9	24534859	23196453	5.46	97.88
VP3MUS	6.9	22576957	22018843	2.47	98.2
VP4LIV	5.2	22830829	21880229	4.16	97.99

Total Reads
841939875

Table S4. Species-specific transcriptome assembly statistics. Total number of reads per assembly, number of normalized reads, percent of reads selected after read normalization, total number of Trinity genes, total number of Trinity transcripts, percent GC content, median contig length, average contig length, the contig N50 score, total number of assembled bases, BUSCO complete scores from the Actinopterygii and Vertebrata databases, and overall alignment percentages from BowTie are listed.

Species	***A. guttata***	***C. personatus***	***E. atriventris***	***G. citrinus***	***T. hollemani***	***V. puellaris***	**Average**
Total Reads	287281998	276916420	273837300	279187444	235965076	274357744	271257664
Normalized Reads	41965198	49513268	50623049	46099265	32434222	33843388	42413065
Percent Selected	14.61	17.88	18.49	16.51	13.75	12.34	15.6
Total Trinity 'genes'	580881	328663	454525	340052	239601	224886	361434.67
Total Trinity transcripts	725858	430973	666674	455077	323694	314526	486133.67
Percent GC	44.74	42.01	43.84	44.02	44.06	44.42	43.85
Median Contig Length	245	286	313	313	324	401	313.67
Average Contig Length	554.23	519.59	605.18	637.94	596.02	907.13	636.68
Contig N50	1351	714	964	1110	922	1837	1149.67
Total Assembled Bases	402292600	223928465	403458039	290313788	192927485	285315354	299705955
Complete BUSCO actinopterygii	86.1	47.7	83.6	69.2	56.3	84.3	71.2
Complete BUSCO vertebrata	89.5	49.3	87	72.8	60.6	88.8	74.67
BowTie Overall Alignment	98.87	93.02	96.9	96.64	96.47	98.4	96.72

Table S5. OrthoFinder summary statistics for each of the six gobiid species.

Species	**A. guttata**	**C. personatus**	**E. atriventris**	**G. citrinus**	**T. hollemani**	**V. puellaris**
Number of genes	249005	136772	259589	194335	138510	204143
Number of genes in orthogroups	234486	126735	244198	181231	129772	193043
Number of unassigned genes	14519	10037	15391	13104	8738	11100
Percentage of genes in orthogroups	94.2	92.7	94.1	93.3	93.7	94.6
Percentage of unassigned genes	5.8	7.3	5.9	6.7	6.3	5.4
Number of orthogroups containing species	73437	49790	73662	63665	50784	63537
Percentage of orthogroups containing species	49	33.2	49.1	42.4	33.9	42.4
Number of species-specific orthogroups	12935	6282	15363	9252	6691	10169
Number of genes in species-specific orthogroups	45722	20634	60766	31679	23135	36217
Percentage of genes in species-specific orthogroups	18.4	15.1	23.4	16.3	16.7	17.7

Table S6. Summary of the differential expression (DE) analyses using the one-to-one orthologs for the cladewise comparisons: Clade 1 (*Amblyeleotris guttata* vs. *Coryphopterus personatus*), Clade 2 (*Eviota atriventris* vs. *Gobiodon citrinus*), and Clade 3 (*Trimma hollemani* vs. *Valenciennea puellaris*) and for the DE analyses using all orthogroups (regardless of being one-to-one. Analyses are further broken down according to tissue type. For each comparison, the number of total one-to-one orthologs/orthogroups, number of DE orthologs/orthogroups, percent DE, and number of upregulated and downregulated ortholog/orthogroups is listed.

Comparison	**1-1 Orthologs**	**DE orthologs**	**Percent DE**	**Upregulated**	**Downregulated**
All tissues, Clade 1	3236	196	6.06	106	90
Liver, Clade 1	2467	221	8.96	129	92
Muscle, Clade 1	2490	182	7.31	62	120
All tissues, Clade 2	4162	129	3.1	75	54
Liver, Clade 2	3130	458	14.63	404	54
Muscle, Clade 2	3397	220	6.48	57	163
All tissues, Clade 3	2991	125	4.18	43	82
Liver, Clade 3	1737	212	12.2	38	174
Muscle, Clade 3	2699	136	5.04	81	55

Comparison	**Orthogroups**	**DE orthogroups**	**Percent DE**	**Upregulated**	**Downregulated**
All tissues, All Species	7,871	3315	42.12	1824	2048
Liver, All Species	6595	1033	15.66	300	733
Muscle, All Species	7292	1505	20.64	1050	455

Table S7. BLAST results for the 19 one-to-one orthologs present across all six species in the RNA-seq analyses. The ortholog name, gene name, protein name, biological process, and molecular function, if known, is shown.

Ortholog	Gene Name	Protein Name	Biological process	Molecular Function
OG0054682	*ALDH9A1*	Aldehyde dehydrogenase family 9 member A1	----	oxidoreductase activity
OG0054798	*snw1*	SNW domain-containing protein 1	mRNA processing	----
OG0054841	*LOC115412578*	Bromodomain adjacent to zinc finger domain protein 2B-like	----	DNA binding
OG0054913	*OLFML2B*	Olfactomedin-like domain-containing protein	Signal transduction	----
OG0054940	*prim1*	DNA primase	DNA replication, transcription	DNA primase activity
OG0055100	*sec63*	SEC63 homolog, protein translocation regulator	Protein transport	RNA binding
OG0055440	*ivns1abp*	Influenza virus NS1A-binding protein	----	----
OG0055715	*unc13d*	Unc-13 homolog D (C. elegans)	Exocytosis, secretion	
OG0056127	----	Uncharacterized protein	----	----
OG0056552	*LOC114435552*	Poly(rC) binding protein 2	----	RNA binding
OG0056712	*SRSF11*	Serine and arginine rich splicing factor 11	RNA splicing	RNA binding
OG0056740	*LOC114432992*	Y+L amino acid transporter 2	----	----
OG0057028	*rtf1*	RNA polymerase-associated protein RTF1 homolog	Transcription, transcription regulation	RNA polymerase II complex binding
OG0057041	*vps26c*	VPS26 endosomal protein sorting factor C	Intracellular protein transport	----
OG0057662	*rplp2*	Large ribosomal subunit protein P2	Cytoplasmic translational elongation	Structural constituent of ribosome
OG0057751	*csnk1a1*	Non-specific serine/threonine protein kinase	Phosphorylation	ATP binding
OG0057764	*wsb1*	WD repeat and SOCS box-containing protein 1	Intracellular signal transduction	----
OG0057905	*hoxb6a*	Homeobox B6a	Transcription, transcription regulation	DNA binding
OG0057973	*agmat*	Agmatine ureohydrolase (agmatinase)	Alpha-amino acid metabolic process, polyamine biosynthetic process	Hydrolase activity

Table S8. Potential candidate genes involved in body size evolution and miniaturization. Orthogroup ID and gene name, whether the gene is upregulated or downregulated in which tissue for which size category, biological process and molecular function associated with the gene, and a brief summary of the gene from NCBI is listed.

Orthogroup	Gene Name	Up	Down	Biological process	Molecular Function	Summary
OG0010936	*aspdh*	Large liver	Small liver	involved in NAD biosynthetic process	enables NADP binding	Predicted to enable NADP binding activity; aspartate dehydrogenase NAD activity; and aspartate dehydrogenase NADP activity. Predicted to be involved in NAD biosynthetic process.
OG0030437	*bzw2*	Small muscle	Large muscle	involved in regulation of translational initiation	enables protein binding	Enables cadherin binding activity. Predicted to be involved in cell differentiation and nervous system development.
OG0033330	*chrnd*	Small muscle	Large muscle	Ion transport, muscle tissue growth, muscle contraction	transmembrane signaling receptor activity, acetylcholine binding	After binding acetylcholine, the AChR responds by an extensive change in conformation that affects all subunits and leads to opening of an ion-conducting channel across the plasma membrane.
OG0031054	*GGCX*	Large liver	Small liver	vitamin K metabolic process	gamma-glutamyl carboxylase activity, vitamin binding	Predicted to enable gamma-glutamyl carboxylase activity and vitamin binding activity.
OG0029191	*ifrd1*	Small muscle, Large liver	Large muscle, Small liver	involved in muscle cell differentiation skeletal muscle tissue regeneration, striated muscle tissue development	---	Immediate early gene which encodes for a protein which may function as a transcriptional co-activator/repressor which controls growth and differentiation of specific cell types during embryonic development and tissue regeneration.
OG0032206	*mrps9*	Small muscle, Large liver	Large muscle, Small liver	mitochondrial translation	structural constituent of ribosome, enables RNA binding	Helps in protein synthesis within the mitochondrion.
OG0036737	*nfu1*	Small liver, Large muscle	Large liver, small muscle	iron-sulfur cluster assembly	enables iron ion binding	Plays a critical role in iron-sulfur cluster biogenesis
OG0020938	*NIPSNAP2*	Small muscle	Large muscle			May be involved in vesicular transport.
OG0031036	*pcnp*	Small muscle, Large liver	Large muscle, Small liver	cell cycle, protein ubiquitination	enables protein binding	Involved in proteasome-mediated ubiquitin-dependent protein catabolic process and protein ubiquitination
OG0034868	*pdcd5*	Small muscle, Large liver	Large muscle, Small liver	apoptotic processes	DNA binding	May be an important regulator of K(lysine) acetyltransferase 5 (a protein involved in transcription, DNA damage response and cell cycle control) by inhibiting its proteasome-dependent degradation.
OG0037782	*rab6a*	Small muscle	Large muscle	protein transport and localization	GTP binding, GTPase activity	This gene encodes a member of the RAB family, which belongs to the small GTPase superfamily. GTPases of the RAB family bind to various effectors to regulate the targeting and fusion of transport carriers to acceptor compartments.

OG0038350	*sat1*	Small muscle, Large liver	Large muscle, Small liver	angiogenesi, spermidine acetylation, putrescine catabolic process	diamine N-acetyltransferase activity, protein binding, spermidine binding	Rate-limiting enzyme in the catabolic pathway of polyamine metabolism. It catalyzes the acetylation of spermidine and spermine, and is involved in the regulation of the intracellular concentration of polyamines and their transport out of cells.
OG0023773	*SDHAF2*	Small liver, Large muscle	Large liver, small muscle	mitochondrial electron transport	enables protein binding	Plays an essential role in the assembly of succinate dehydrogenase (SDH), an enzyme complex that is a component of both the tricarboxylic acid (TCA) cycle and the mitochondrial electron transport chain, and which couples the oxidation of succinate to fumarate with the reduction of ubiquinone (coenzyme Q) to ubiquinol.
OG0036887	*ybx1*	Small liver	Large liver	transport, embryonic morphogenesis and developmen, cell division	enables DNA and protein binding	The encoded protein functions as both a DNA and RNA binding protein and has been implicated in numerous cellular processes including regulation of transcription and translation, pre-mRNA splicing, DNA reparation and mRNA packaging.
OG0020678	*zgc:136908*	Small muscle	Large muscle	autophagy, cellular response to stress	ATP binding, ATP hydrolysis activity, lipid binding	'Predicted to have ATPase activity and polyubiquitin modification-dependent protein binding activity. Predicted to be involved in ERAD pathway; autophagosome maturation; and mitotic spindle disassembly.