UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

DIVERSIFICATION DYNAMICS IN SPACE AND TIME OF TWO MARINE FISH GROUPS

A DISSERTATION SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By
AINTZANE SANTAQUITERIA GIL
Norman, Oklahoma
2024

DIVERSIFICATION DYNAMICS IN SPACE AND TIME OF TWO MARINE FISH GROUPS

A DISSERTATION APPROVED FOR THE

SCHOOL OF BIOLOGICAL SCIENCES

BY THE COMMITTEE CONSISTING OF

Chair: Dr. Richard Broughton

Dr. Ricardo Betancur

Dr. Katharine Marske

Dr. Gavin Woodruff

Dr. Lena Cole

Acknowledgments

I am deeply grateful for the unwavering support and guidance of my supervisor, Ricardo Betancur, whose mentorship has been influential in navigating the challenges of my Ph.D. journey. Thanks, Ricardo, for your passion, knowledge, hard work, and for the valuable insights you shared along the way. Thank you for the opportunity you gave me to work with you and for teaching me so much so I could become the scientist I am now. I'm excited to continue this productive collaboration in the future!

Many thanks to Dahiana Arcila and all my labmates from the Fish Evolution Lab family, who enriched my research experience and contributed to my personal and professional growth. All the adventures I lived with *La loca académica de Biología* will always be in my heart. Going to national and international conferences, spending hours in a car talking about research, anecdotes, and of course singing our favorite songs on road trips, diving in paradisiacal places, conducting fieldwork day and night, and our get-togethers. I hope we can continue sharing these crazy adventures together in the times ahead.

To my dissertation committee, Richard Broughton, Katie Marske, Gavin Woodruff, and Lena Cole as well as Daniel Moen and Cecil Lewis, I extend my sincere appreciation for their constructive feedback and encouragement throughout the research process. I am indebted to both the University of Puerto Rico-Río Piedras (UPR-RP) and the University of Oklahoma (OU) for providing the necessary resources and conducive environment for pursuing my doctoral studies. I am very thankful to all the researchers and institutions for providing tissue samples, conducted DNA extractions, library preparation, and sequencing. Also, to the bioinformaticians from the High-Performance Computing facility of UPR-RP and the OU Supercomputing Center for Education & Research (OSCER). This research would have not been possible without the funding support from the National Science Foundation grants awarded to my advisor as well as the Bullard Dissertation Completion Award I received from OU.

My heartfelt thanks go to my family and friends for their boundless love and encouragement, which sustained me through the highs and lows of this PhD journey. From Pamplona, to Norway, and now Oklahoma, I am truly grateful for your friendship and the way you have made each of these places more enjoyable to live in through your acts of kindness. Whether it's lending a helping hand, sharing laughter, playing different sports, or offering support during challenging times, your presence has added warmth and joy to my journey in each of these locations. Thank you for being such wonderful friends. Special thanks to two incredible women, Melissa Rincón and Jennifer Londoño, for their wise advice, positive energy, memorable trips, and shared laughter—I will cherish these memories forever. And finally, I am forever grateful for the support of my parents, Juan and Begoña, and my sister Enara, who have shown me that no matter where I am, they will support me and be proud of all my achievements.

Mila esker nigatik egin duzuen guztiagatik
&
Gracias por hacer realidad este sueño.

Abstract

Understanding the evolutionary processes shaping species distributions in both marine and terrestrial environments has been a central interest among evolutionary biologists, biogeographers, and ecologists. Species richness in a given region is directly influenced by three key processes: speciation, extinction, and dispersal. Variations in the rates and timing of these processes are responsible for shaping diversity gradients such as those observed along latitudinal, longitudinal, elevation, and depth gradients. These variations arise from a complex interplay of biotic and abiotic factors, encompassing climatic stability, geographical barriers, trophic specializations, productivity, competition, and predation. Macroevolutionary studies using phylogenies can illuminate these evolutionary patterns and processes over extensive timescales and diverse taxonomic groups. Specifically, integration of comprehensive phylogenetic trees, derived from extensive taxonomic sampling of both extinct and extant species, with thorough genetic analysis, and supplemented by ecological and morphological datasets, can facilitate identifying factors influencing diversification and biogeographic trends across taxa.

The overarching goal of my dissertation is to understand how extrinsic (e.g., formation of historical barriers, temperature) and intrinsic (e.g., life history processes such as feeding mode, dispersal ability) factors may have shaped the evolution of two charismatic groups of marine reef fishes. The first two chapters aim at examining Syngnatharia, an extraordinarily diverse clade (>660 species) that includes trumpetfishes, goatfishes, dragonets, seahorses, pipefishes, and allies. The third chapter focuses on fishes in the order Acanthuriformes, which comprises surgeonfishes, the louvar, and the moorish idol (87 species). Despite progress made in unravelling the relationships of these and other clades of disparate marine fish groups based on a handful of genetic markers sequenced from a few representative lineages, the vast majority of the species lack phylogenetic placement. Additionally, very few studies have looked at genes associated with phenotypic or ecological changes in reef fishes from a macroevolutionary perspective. To fill in these gaps, my research aims to examine the evolutionary history of these groups using state-of-the-art approaches, including phylogenomics, phylogenetic comparative methods, and phylogenetically-informed genotype-to-phenotype (PhyloG2P) comparative genomic approaches based on whole genomes.

In my first chapter, I applied an integrative phylogenomic approach to elucidate the evolutionary history and biogeography of Syngnatharia. I collected genome-wide DNA sequence and geographic distribution data for 169 species to cover ~25% of the species diversity and all 10 families in the group, and complemented these datasets with paleontological and geological information. With these datasets I inferred a set of time-calibrated trees and reconstructed the ancestral ranges of the group. I then examined the sensitivity of biogeographic analyses to phylogenetic uncertainty (estimated from multiple genomic subsets), area delimitation, and biogeographic models. After accounting for these uncertainties, my results reveal that syngnatharians originated in the ancient Tethys Sea at the Late Cretaceous, 87 million years ago (Ma) and subsequently occupied the Indo-Pacific Ocean. Throughout syngnatharian history, multiple independent lineages colonized the Eastern Pacific (6–8 times) and the Atlantic (6–14

times) from their center of origin, with most events taking place following an east-to-west route prior to the closure of the Tethys Seaway between 12–18 Ma. These colonizations were facilitated by the long-distance dispersal ability of syngnatharians during their pelagic larval stages or through rafting, such as with sargassum-associated species, aided by oceanic currents.

For my second chapter, I examined factors driving syngnatharians species richness along the longitudinal diversity gradient across oceans and assessed whether patterns of morphological diversity are congruent with this gradient. I increased the taxonomic sampling of syngnatharians from my first chapter to 323 species (50% of the species diversity) to test three non-mutually exclusive evolutionary hypotheses proposed to explain the longitudinal diversity gradient: time-for-speciation, center of accumulation, and *in situ* diversification rates. I estimated diversification rates and body shape disparity broadly across the group, considering biogeographic regions within all three major oceanic realms (Indo-Pacific, Atlantic, and eastern Pacific), as well as within the Indo-Pacific region. The analyses showed that the extensive diversity of syngnatharian species in the Indo-Pacific region primarily stems from ancient colonizations, leading to *in situ* speciation during the Palaeogene, shortly after the Paleocene-Eocene Thermal Maximum (PETM), and subsequent lineage accumulation during the Miocene coinciding with the initiation of the Indo-Australian Archipelago (IAA) rearrangement. Conversely, the eastern Pacific and Atlantic regions exhibit lower regional diversities, largely due to more recent colonization events and the onset of diversification, with most lineages in these areas emerging during the Miocene. These findings strongly support the time for speciation and center of accumulation hypotheses. My study also reveals that a significant portion of syngnatharian morphological diversity originated early in their evolutionary history within the Tethys Sea, followed by a gradual decline in subclade disparity marked by the emergence of multiple adaptive peaks, particularly in head morphology. This suggests that while high species richness exists, it does not necessarily correlate with high morphological disparity across various biogeographic contexts. All in all, colonization dynamics explain the longitudinal diversity patterns of syngnatharian fishes across marine realms while morphological similarities persist among them.

In my third chapter**,** I examined the ecological drivers of trophic transitions among fossil and extant acanthuriforms as well as the genomic basis of these transitions. By combining genomic data for 80 extant species (~93% of total diversity) with morphological characters for 32 fossil taxa, I inferred total evidence time-calibrated phylogenies. Using these phylogenies, I reconstructed the diet of acanthuriforms and investigated the number of times the planktivory lifestyle evolved, along with the geographic location and timing of these transitions. The analyses indicate an origin of acanthuriforms approximately 64 Ma following the K-Pg mass extinction event, with at least seven documented transitions to planktivory from non-planktivorous lineages, followed by at least four reversals to non-planktivorous diets. While the earliest transitions occurred in the ancient Tethys Sea, the most recent ones happened within the Indo-Pacific region. I then evaluated the effect of the convergently evolved diets on acanthuriforms' diversification, finding no significant effect as diversification rates remain constant across trophic guilds. However, transition rates are higher from planktivores to non-planktivores compared to the opposite direction. Diversification of planktivore species does appear to be influenced by cool past climatic temperatures, although there is also a confounding effect from phylogenetic signal.

Despite ecological and morphological factors commonly driving this trophic specialization, the extent to which this adaptive convergence is caused by convergent changes at the molecular level remains understudied in reef fishes. Therefore, in this study I performed PhyloG2P analyses, using newly-generated chromosome-level (*Acanthurus chirurgus*) and short-read (45 species) genomes to identify genes under positive selection across planktivore lineages and along branches where a transition to planktivory occurred. We identified a total of 91 genes that underwent convergent positive selection in planktivorous lineages, along with three genes unique to planktivores. These genes are implicated in metabolic processes and adaptations in body shape, consistent with the repeated instances of convergence towards a pelagic environment, which are associated with planktivory and specialized morphological traits.

In summary, my dissertation explores the evolutionary processes shaping the distributions of marine fish species, highlighting the pivotal roles of speciation, extinction, and dispersal in driving diversity across oceans. My research also underscores the importance of integrating data from both fossil and living species to obtain a more comprehensive picture of the evolutionary history of groups. Through comprehensive analyses based on genomic, ecological, and morphological data, I emphasize the need to address various factors generating uncertainty in macroevolutionary and biogeographic inferences. Furthermore, my findings contribute to our understanding of the evolutionary dynamics as well as genetic underpinnings of trophic transitions in marine fishes, shedding light on the adaptive mechanisms driving diversification. Overall, my thesis represents an important step towards understanding the evolutionary history of marine fishes by disentangling their diversification patterns in space and time.

Contents

List of figures

List of tables

Chapter 1

Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences

Published in *Systematic Biology* (https://doi.org/10.1093/sysbio/syab028)

Aintzane Santaquiteria, Alexandre C. Siqueira, Emanuell Duarte-Ribeiro, Giorgio Carnevale, William White, John Pogonoski, Carole C. Baldwin, Guillermo Ortí, Dahiana Arcila, and Ricardo Betancur-R.

1.1 Abstract

The charismatic trumpetfishes, goatfishes, dragonets, flying gurnards, seahorses, and pipefishes encompass a recently defined yet extraordinarily diverse clade of percomorph fishes—the series Syngnatharia. This group is widely distributed in tropical and warm-temperate regions, with a great proportion of its extant diversity occurring in the Indo-Pacific. Because most syngnatharians feature long-range dispersal capabilities, tracing their biogeographic origins is challenging. Here, we applied an integrative phylogenomic approach to elucidate the evolutionary biogeography of syngnatharians. We built upon a recently published phylogenomic study that examined ultraconserved elements by adding 62 species (total 169 species) and one family (Draconettidae), to cover *ca.* 25% of the species diversity and all 10 families in the group. We inferred a set of time-calibrated trees and conducted ancestral range estimations. We also examined the sensitivity of these analyses to phylogenetic uncertainty (estimated from multiple genomic subsets), area delimitation, and biogeographic models that include or exclude the jump-dispersal parameter (*j*). Of the three factors examined, we found that the *j* parameter has the strongest effect in ancestral

range estimates, followed by number of areas defined, and tree topology and divergence times. After accounting for these uncertainties, our results reveal that syngnatharians originated in the ancient Tethys Sea *ca.* 87 Ma (84–94 Ma; Late Cretaceous) and subsequently occupied the Indo-Pacific. Throughout syngnatharian history, multiple independent lineages colonized the eastern Pacific (6–8 times) and the Atlantic (6–14 times) from their center of origin, with most events taking place following an east-to-west route prior to the closure of the Tethys Seaway *ca.* 12–18 Ma. Ultimately, our study highlights the importance of accounting for different factors generating uncertainty in macroevolutionary and biogeographic inferences.

1.2 Introduction

Molecular phylogenetic studies are steadily resolving long-lasting uncertainties in the Fish Tree of Life, most notably within percomorphs, a clade of spiny-finned fishes often referred to as the "bush at the top" (Nelson 1989) that is characterized by spectacular morphological and taxonomic diversity (*ca.* 18,000 species). Interrelationships and composition of major percomorph lineages remained controversial after decades of anatomical studies (Johnson 1993; Johnson and Patterson 1993), but recent phylogenetic analyses based on molecular evidence have unambiguously supported the resolution of the percomorph bush into nine supraordinal clades (Near et al. 2012; Betancur-R. et al. 2013, 2017; Sanciangco et al. 2016; Alfaro et al. 2018; Hughes et al. 2018). Most of these clades have never been inferred on the basis of morphological evidence, and therefore the origin and mechanisms shaping the evolutionary radiation of percomorphs remain poorly explored (Alfaro et al. 2018). Recent studies have suggested an association between the Cretaceous-Palaeogene (K-Pg) mass extinction and the origin of five of these nine percomorph crown groups, implying an important effect of extinctions on the evolutionary dynamics that resulted in the clades' astonishing diversity (Alfaro et al. 2018; Ribeiro et al. 2018).

One such percomorph clade has been classified as the series Syngnatharia, which includes *ca*. 670 described species arranged in 10 families, four suborders (Callionymoidei, Dactylopteroidei, Mulloidei, and Syngnathoidei) and a single order (Syngnathiformes; Betancur-R. et al. 2017). Best-known among syngnatharians are seahorses and pipefishes (family Syngnathidae), forming a clade with over 320 species (Fricke et al. 2020). The fossil record of Syngnatharia dates back to the Late Cretaceous and includes representatives from all suborders (Carnevale et al. 2006; Bannikov 2014; Cantalice and Alvarado-Ortega 2016; Carnevale and Bannikov 2019). Most of these fossils have been found in the Eocene strata of Monte Bolca (Italy), at the ancient western Tethys Sea—one of the most important localities bearing exquisitely preserved fossil teleosts known from the paleontological record (Bannikov 2014; Carnevale et al. 2014; Friedman and Carnevale 2018). Many extant syngnatharian families are circumglobally distributed, occurring in tropical and temperate marine waters, although some species inhabit brackish and freshwater environments (Whitfield 1999; York et al. 2006; IUCN 2019; OBIS 2021). Syngnatharians live in diverse habitats from soft substrates (e.g., goatfishes, seamoths, flying gurnards) to more complex environments such as coral and rocky reefs (seahorses, trumpetfishes) and mangrove forests (seahorses); other groups also inhabit sea-grass beds (pipefishes) and drifting macroalgae (*Sargassum*-associated pipefishes; Froese and Pauly 2019). Reflecting this diversity in habitats, and the associated challenges in locomotion, reproduction and feeding ecology, fishes in this group display a variety of body plans. Some of the most conspicuous morphological traits and behaviors observed in syngnatharians include snout and body elongation (e.g., pipefishes, pipehorses, trumpetfishes and cornetfishes; Bergert and Wainwright 1997), vertical swimming (e.g., shrimpfishes and seahorses; Atz 1962; Lin et al. 2016; Fish and Holzman 2019), prehensile tails (seahorses and pipehorses; Neutens et al. 2014; Hamilton et al. 2017), hyoid barbels (goatfishes; Gosline 1984), and, most remarkably, male pregnancy (pipefishes, pipehorses and seahorses; Berglund et al. 1986). Although there is no morphological evidence that unifies the disparate array of clades included in Syngnatharia (e.g., goatfishes and seahorses), the monophyly of the group is consistently resolved by different molecular studies, whether based on mitochondrial markers alone (Kawahara et al. 2008; Song et al. 2014), a handful of nuclear and mitochondrial markers (Chen et al. 2003; Near et al. 2012; Betancur-R. et al. 2013) or genome-scale datasets (Longo et al. 2017; Alfaro et al. 2018; Hughes et al. 2018).

Despite robust support for the monophyly of Syngnatharia, phylogenetic relationships among early branching (suborder-level) clades have been controversial due to a seemingly explosive radiation during the early stages of their evolution (Near et al. 2013; Betancur-R. et al. 2017; Longo et al. 2017; Alfaro et al. 2018). Short internodes along the backbone of the syngnatharian phylogeny make it challenging to resolve these relationships due to high levels of incomplete lineage shorting (Maddison 1997) and/or low signal-to-noise ratios (Rokas and Carroll 2006; Townsend et al. 2012). Whereas previous studies using sparse taxonomic sampling or few nuclear markers failed to provide robust resolution at the suborder level (Near et al. 2012; Betancur-R. et al. 2013; Song et al. 2014; Hughes et al. 2018), more recent phylogenetic analyses based on genome-scale data coupled with dense-taxonomic sampling (Longo et al. 2017) delineated the four major, morphologically-supported suborders (Kim 2002; Wiley and Johnson 2010; Nelson et al. 2016; see also Betancur-R. et al. 2017), making progress towards the resolution of this spectacular radiation.

Resolving syngnatharian relationships will open new avenues to address a variety of macroevolutionary inferences, such as historical biogeography. Previous attempts to investigate the biogeographic history of this group have been restricted to particular families or genera (e.g., Syngnathidae, *Hippocampus, Mulloidichthys*; Bowen et al. 2001; Teske et al. 2004, 2007; Boehm et al. 2013; Lessios and Robertson 2013; Hamilton et al. 2017; Li et al. 2021). These studies have consistently identified the Indo-Pacific (IP) as the center of origin for these geographically widespread clades, but their results differed regarding the inferred routes and timing of the colonization of the Atlantic and the eastern Pacific (EP).

These previous studies, and others that have examined the historical biogeography of other clades, do not typically account for factors generating uncertainty in comparative analyses, such as variations in tree topology and divergence times (e.g., Batista et al. 2020) or alternative delineation of biogeographic areas and models (e.g., Richardson et al. 2004). Recent advances in statistical approaches for ancestral range estimations now allow the implementation of alternative biogeographic models—e.g., dispersal-extinction-cladogenesis (DEC; Ree and Smith 2008a), dispersal-vicariance–analyses (DIVA; Ronquist 1997), and Bayesian inference of historical biogeography for discrete areas (BayAREA; Landis et al. 2013)—and parameters (e.g., the jump-dispersal or founder-event speciation [*j*]) with different assumptions regarding the dispersal abilities of groups between areas at different time slices (e.g., LaGrange: Ree and Smith 2008b; BioGeoBEARS: Matzke 2013). Remarkably, among the alternative biogeographic models available for analyses, variants that include the *j* parameter have been recently criticized because their likelihood can be artificially inflated leading to an overestimation of jump-dispersal events (Ree and Sanmartín 2018; but see Klaus and Matzke 2020).

Here, we assess the evolutionary and biogeographic history of Syngnatharia using a variety of sensitivity analyses in a phylogenomic comparative framework. We first expanded a recently constructed molecular phylogeny for Syngnatharia that examined ultraconserved element (UCE) loci (Longo et al. 2017) to include 62 additional species (169 species total), and inferred robust time-calibrated trees by integrating paleontological and geological information. To account for topological and divergence-time uncertainty in downstream comparative analyses, we also partitioned the complete matrix into genomic subsets. With the resulting, comprehensive phylogenetic framework, we explored the global biogeography of the group by incorporating geographic distribution data from extant and fossil species in three different time slices. Additionally, we examined uncertainties in ancestral range estimations using different sets of predefined area schemes, biogeographic models (including and excluding the *j* parameter), and topologies. Finally, in light of these results, we inferred possible routes through which different lineages colonized the EP and the Atlantic.

1.3 Materials and methods

See Appendix A, Supplementary Materials and Methods section for additional methodological details.

Taxonomic sampling and genomic data

Our genomic dataset was built upon a recently published phylogenomic analysis of Syngnatharia based on UCE data for 113 species (Longo et al. 2017), including a scombroid species as the outgroup (*Taractichthys longipinnis*). We added 78 newly sequenced specimens (62 species), including one additional family (Draconettidae) thereby covering all 10 representative families in the group. We initially assembled a dataset of 190 individuals. We updated the total number species after implementing steps for sample quality-control (see Appendix A, Supplementary Materials and Methods), based on CO1 sequence comparisons to the Barcode of Life Database (BOLD) and National Center for Biotechnology Information (NCBI) databases. These steps resulted in the reidentification of six taxa, including four sequenced by Longo et al. (2017) that turned out to be duplicates (see Appendix A, Supplementary Materials and Methods). Our final dataset consisted of 184 individuals comprising 169 syngnatharian species (107 previously published and 62 newly sequenced) or 25.3% of the clade’s diversity (Supplementary Appendix 1 available on Dryad).

DNA extractions, UCE library preparation and sequencing

We extracted DNA from tissue samples in a 96-well plate format on a GenePrep, following manufacturer's instructions at the Laboratory of Analytical Biology at the Smithsonian National Museum of Natural History. We checked the quality of DNA extractions by visually inspecting high molecular weight DNA stained with GelRed (Biotium) on a 1% agarose gel. High quality DNA extractions were sent to Arbor Biosciences for target enrichment using the MYbaits UCE Capture Kits, a custom bait set of approximately 1340 UCE loci identified in acanthomorph fishes (McGee et al. 2016; available from http://ultraconserved.org). Enriched libraries were quantified with qPCR (Kapa Biosystems) and sequenced using a single lane of Illumina HiSeq 2500 at the U. Chicago Genomics Facility.

Data assembly and alignment

We used the PHYLUCE computational pipeline (Faircloth 2016) to process the raw sequence reads (http://phyluce.readthedocs.io/en/latest/tutorial-one.html). This pipeline generated a 'monolithic' FASTA file with all UCEs for all newly sequenced individuals. After this step, we added to the 'monolithic' file UCE data for the 113 species (112 syngnatharians and the scombroid outgroup) previously sequenced (Longo et al. 2017). We 'exploded' this file to obtain individual UCE loci and aligned them using MAFFT (Katoh and Standley 2013) based on a maximum divergence of 0.2. We trimmed the resulting alignments using Gblocks v0.91b (Castresana 2000) to remove ambiguously aligned flanking regions. Edited alignments consisted of 1,186 UCE loci that we used to generate two subsets, each including all taxa examined: a 75% completeness matrix with 932 UCEs (142 taxa contain data in all gene alignments) and a 90% completeness matrix with 346 UCEs (171 taxa in all gene alignment). The resulting alignments included nine of ten syngnatharian families, with only Draconettidae missing. To cover all representative families, we probed the raw data from a specimen of *Draconetta xenica* that was sequenced using exon capture approaches (Hughes et al. 2020) to identify shared UCE loci using the map to reference function implemented in Geneious v.11.1.2 program (Biomatters 2019). We recovered a total of 17 and 50 loci present in the 90% and 75% completeness matrices, respectively. See Appendix A, Supplementary Materials and Methods for additional details.

Phylogenomic analyses

For both matrices, we determined the best-fit partitioning scheme as well as the best-fit model for each partition using the Bayesian Information Criterion (BIC) in Partition Finder v2.1.1 (Lanfear et al. 2017). We used the sliding-window approach and entropy site characteristic (SWSC-EN), a partition method proposed for UCE data (Tagliacollo and Lanfear 2018). The SWSC-EN produces partitions for each locus based on a core and two flanking regions. We estimated concatenation-based maximum-likelihood (ML) trees in RAxML v8.2.4 (Stamatakis 2014) using the best-fit partitioning schemes and the GTRGAMMA substitution model. We ran 30 independent ML searches and used nonparametric bootstrapping via autoMRE (number of bootstrap replicates automatically determined) to assess edge support. We also conducted coalescent-based species-tree analyses in ASTRAL-II (Mirarab and Warnow 2015) using RAxML gene trees as input. Gene trees were inferred using the UCE core-flank partitions and the same parameters applied for concatenation analyses. Before phylogenetic dating analyses, we pruned duplicate individuals per species from the corresponding trees.

To account for variation in topology and divergence times in biogeographic analyses (see below), we also assembled 12 largely independent subsets (subsampled from the 75% matrix), each with a sufficient number of genes to overcome sampling error. Preliminary tests including a higher number of subsets, each with fewer genes (18 subsets), resulted in high levels of topological discordance, particularly for trees estimated with ASTRAL-II. We thus reduced the number of subsets to 12 (two with 89 loci, six with 90 loci and four with 91 loci), all of which produced trees with lower levels of topological incongruence compared to those obtained using fewer genes. To maintain *Draconetta xenica* across all trees, all subsets overlapped in 17 anchor UCE markers that include this taxon. We conducted phylogenetic analyses using the 12 subsets in RAxML and ASTRAL-II, as explained above, producing a total of 24 trees (two per subset). Finally, we used the 28 trees inferred (two each with the 75% and 90% completeness matrices, and 24 with the subsets) as input for divergence time estimations in MCMCTree.

Phylogenomic dating

We estimated divergence times using the MCMCTree package implemented in PAML v4.9 (Yang 2007), which can analyze genome-scale datasets in a Bayesian framework (dos Reis and Yang 2019). A drawback of MCMCTree, however, is that it cannot jointly estimate topology and divergence times, requiring instead predefined topologies as input, for which we used the 28 topologies inferred in the previous step. Because running time in MCMCTree depends more on the number of defined partitions rather than the number of genes, we divided each subset into two partitions (flanks and core UCEs). We used a total of 11 calibration points, including a secondary root calibration (Pelagiaria + Syngnatharia), six fossil calibrations, and four geological calibrations (see Appendix A, Supplementary Materials and Methods). We ran the 75% and 90% matrices for 10–50 million generations and the genomic subsets for 3–20 million generations until convergence was reached based on effective sampling size (ESS) values (>200). We used the approximate likelihood method and the HKY85 model. Prior parameters for the MCMCTree runs were as follow: independent rate relaxed-clock model, BDparas: 1, 1, 0.27; kappa_gamma: 6, 2; alpha_gamma: 1, 1; rgene_gamma: 2, 200, 1; sigma2_gamma: 2, 5, 1. We conducted two independent runs for each dataset. To check for convergence, we used Tracer v1.7.1 (Rambaut et al. 2018) to examine trace plots and ESS values for each parameter, after a 10% burn-in.

Biogeographic analyses

We ran biogeographic analyses using the R package BioGeoBEARS (Matzke 2013), which compares models of range evolution in a phylogenetic framework. We used the tree inferred in RAxML with the 75% completeness matrix ('master tree' hereafter) as the summary phylogeny. We also implemented three different biogeographic schemes to account for different levels of resolution for delimitation of biogeographic regions (Table 1). The first scheme is based on six recognized marine biogeographic regions for tropical fishes proposed by Kulbicki et al. (2013): Western Indian Ocean (WIO), Central Indo-Pacific (CIP), Central Pacific (CP), Tropical Eastern Pacific (TEP), Western Atlantic (WA), and Eastern Atlantic (EA). For the second scheme, based on Spalding et al. (2007), we added the Temperate Australasia (TA) area to the previous six-area scheme to account for species restricted to the temperate waters of Australia and New Zealand (these species were lumped into the CIP in the six-area scheme). Finally, for the third scheme, we further added an area to the seven-area scheme to discriminate species that are largely endemic to the Temperate Northeast Pacific (TNP; lumped into the EP in the 6- and 7-area schemes). We built a presence/absence matrix for each scheme by coding each extant species according to their geographic ranges primarily based on a combination of the IUCN Red List (IUCN 2019) and Ocean Biogeographic Information System (OBIS 2021) databases. We also used paleogeographic domain information as biogeographic constraints using the six fossils used to calibrate our tree. These constraints were placed in the nodes into which fossil calibration points were assigned (Table 2). Five of the six fossil species, placed in Syngnatharia, Syngnathidae/Solenostomidae, *Hippocampus*, and †Ramphosidae (the sister family of Pegasidae), were present in the western Tethys or Paratethys and thus coded as an additional area, the ancient Tethys Sea (calibrations 2–4, 6, and 7; Table 1.2 and Fig. 1.1). The sixth fossil used for Aulostomoidea (calibration 5; Table 1.2 and Fig. 1.1) was described from the Gulf of Mexico/Proto Caribbean Sea and thus coded as WA.

Table 1.1 Alternative biogeographic schemes used in BioGeoBEARS.

Scheme	**Number of areas**	**Areas**	**Maximum range size parameter**
1. Six-area	6 + Tethys Sea	WIO, CIP, CP, EP, WA, EA	6
2. Seven-area	7 + Tethys Sea	WIO, CIP, CP, TA, EP, WA, EA	7
3. Eight-area	8 + Tethys Sea	WIO, CIP, CP, TA, TNP, TEP, WA, EA	7

WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; TNP: Temperate Northeast Pacific; TEP: Tropical Eastern Pacific; WA: Western Atlantic; EA: Eastern Atlantic.

Table 1.2 Fossil geographic distributions used as biogeographic constraints in BioGeoBEARS.

Fossil calibrations	MRCA	Paleogeographic domain	Area code
(2) Syngnatharia	*Syngnathus louisianae*, *Eurypegasus draconis*	Western-central Tethys	Tet
(3) Syngnathidae/Solenostomidae	*Solenostomus cyanopterus*, *Syngnathus louisianae*	Western Tethys	Tet
(4) *Hippocampus*	*Hippocampus abdominalis*, *Hippocampus kuda*	Pannonian Basin, Central Paratethys	Tet
(5) Aulostomoidea	*Aulostomus maculatus*, *Aeoliscus strigatus*	Gulf of Mexico/Proto Caribbean Sea	WA
(6) Fistulariidae	*Fistularia corneta*, *Aulostomus maculatus*	Western Tethys	Tet
(7) Pegasidae	*Pegasus volitans, Dactylopterus volitans*	Western Tethys	Tet

Tet: Tethys Sea; WA: Western Atlantic.

We tested 12 biogeographic models in a maximum likelihood framework, previously used for marine fishes (Siqueira et al. 2019; Rincon-Sandoval et al. 2020). These models include DEC (Ree and Smith 2008), DIVA (Ronquist 1997), and BayAREA (Landis et al. 2013). Each model was run with and without the founder-speciation event (*j*) (Matzke 2014) and the dispersal matrix power exponential (*w*) parameters (Dupin et al. 2017). The *j* parameter allows the colonization of a new area by a daughter lineage while the splitting-sister lineage stays at the ancestral area (Matzke 2014). The *w* parameter is used to infer the optimal dispersal multiplier matrix, acting as an exponent on the matrix (Dupin et al. 2017). We set the *w* parameter to be free in order to allow the model to adjust the matrices according to the data. We analyzed each model using three time slices (100–12 Ma, 12–2.8 Ma, and 2.8–0 Ma) according to different geological events that span the evolutionary history of the clade. The Tethys Sea region was only added to the analyses for the first time slice (100–12 Ma) to reflect the existence of this ancient basin. A dispersal-multiplier matrix for each scheme was assembled to account for the dynamics of biogeographic barriers over time. The connectivity between areas was determined by three dispersal probability categories: 1.0 for well-connected areas, 0.05 for relatively separated areas, and 0.0001 for widely separated or disconnected areas. From 100 to 12 Ma, we allowed high dispersal probability between WIO and EA through the Tethys Seaway. The Terminal Tethyan Event (TTE), which opened and closed intermittently between 12 and 18 Ma, divided the western and the eastern Tethys Sea (Steininger and Rögl 1979; Adams et al. 1983; Rögl 1998, 1999). Thus, from 12 Ma (final closure) onwards we used a low dispersal probability value between WIO and EA to allow dispersal through southern Africa (Rocha et al. 2005). To account for the final closure of the Isthmus of Panama, which may have occurred as early as 2.8 Ma as stated above (O'Dea et al. 2016), we assigned a very low dispersal probability between WA and EP. Finally, for all time slices, we set a high dispersal probability between adjoining areas of the Indo-Pacific and a low dispersal probability between CP and EP to reflect dispersal limitations associated with the crossing of the EP Barrier (Bellwood and Wainwright 2002; Lessios and Robertson 2006). Using the 'master tree' as the input phylogeny, we calculated Akaike Information Criterion scores corrected for small sample size (AICc) for each biogeographic model and for each biogeographic scheme independently. The best-fitting model (DEC, DIVA, and BayAREA, each with a combination of ±*j* and ±*w* parameters) was then selected for each scheme (Table S4).

Accounting for uncertainty in biogeographic analyses

We assessed sensitivity of ancestral range estimations to three major sources of variation: topology and divergence times, area schemes, and biogeographic models. First, based on a recently-proposed approach (Rincon-Sandoval et al. 2020), we used the set of 28 trees inferred using the 75% and 90% completeness matrices as well as the 12 genomic subsets. Resulting trees reflect uncertainty in divergence times and phylogenetic relationships based on different underlying data. This approach fundamentally differs from the common practice of conducting comparative analyses using 'pseudo-replicated' trees obtained from a Bayesian posterior distribution estimated with a single dataset, typically consisting of a handful of genes (Huelsenbeck et al. 2000). To assess topological disparity, we estimated tree space plots for the 28 trees using a multidimensional scaling (MDS) visualization implemented in the R package *phytools* (Revell 2012). To assess the effects of phylogenetic variation in biogeographic inferences, we used a code produced by Matzke (2019) to summarize ancestral range estimates from multiple trees by selecting the 'master tree' as the topology upon which the results from all 28 trees were overlain. This approach allowed us to obtain averaged probabilities across the different trees for compatible nodes present on the 'master tree.' For comparison, we also estimated ancestral ranges based on an alternative tree estimated with ASTRAL-II using the 75% completeness matrix ('alternative tree' hereafter) and the 28-tree averaging approach described above. The final set of analyses aimed at accounting for phylogenetic uncertainty involved running BioGeoBEARS without the averaging approach, using only the 'master tree' (see also previous section) and the 'alternative tree.'

Second, we compared the biogeographic results obtained with the three different area schemes defined (6 areas, 7 areas, and 8 areas), using both the master and alternative trees with and without the 28-tree averaging approach. Finally, given recent criticisms on the implementation of the jump-dispersal parameter (*j*) (Ree and Sanmartín 2018), we interpreted different biogeographic histories based on analyses of the best-fitting models selected for different trees and area schemes, with (+*j*) and without (-*j*) this parameter (Appendix A Table S4).

In cases where colonization events of oceanic basins inferred from these different types of analyses produced incongruent results, we assessed the relative probabilities of these histories by conducting biogeographic stochastic mapping (BSM), as implemented in BioGeoBEARS (Dupin et al. 2017). For BSM analyses, a total of four possible routes were assessed (Floeter et al. 2008): (i) Tethyan relicts, (ii) lineages with Indo-Pacific origin that crossed the Tethys Seaway before its closure, (iii) lineages with Indo-Pacific origin that colonized via the Cape of Good Hope, southern Africa, and (iv) lineages with Indo-Pacific origin that crossed the EP Barrier. We simulated 1,000 stochastic histories on the 'master tree' based on the best-fit biogeographic model (with and without the *j* parameter) and calculated the probability for alternative routes. These alternatives are only considered (depicted in maps) when the probability for a major colonization event is less than 70%.

1.4 Results

Phylogenomic inference, tree uncertainty and divergence times

We conducted phylogenomic analyses using maximum likelihood (ML; RAxML) and coalescent-based (ASTRAL-II) approaches applied to the two assembled matrices—the 75% completeness matrix (932 UCEs, 268,279 sites, 11.6% missing data) and the 90% completeness matrix (346 UCEs, 119,467 sites, 6% missing data). Overall, the phylogenetic relationships among the four suborders previously defined are congruent and highly supported (>75%) based on the concatenation-based ML analyses (Fig. 1.1 and Appendix A Figs. S2, S3, and S5). However, trees inferred with ASTRAL-II using both matrices (75% and 90%; Appendix A Figs. S4 and S6) did not resolve the monophyly of Dactylopteroidei (Dactylopteridae + Pegasidae). Additionally, the suborder Callionymoidei, a clade comprising the families Draconettidae and Callionymidae, which have a strong morphological affinity (Gosline 1984; Wiley and Johnson 2010; Nelson et al. 2016), was not resolved as monophyletic with ASTRAL-II using the 90% matrix. These results suggest that lower gene coverage for Draconettidae (only 17 and 50 UCE loci present in the 90% and 75% matrices, respectively) may have affected ASTRAL-II analyses more than concatenation-based inferences. Trees estimated with RAxML had higher average bootstrap support values than those estimated with ASTRAL-II (96.8–98.3% vs. 90.5–92.4%, respectively). Likewise, trees estimated with the 75% matrix resulted in clades with higher support values relative to the 90% matrix (mean support 92.4–98.3% vs. 90.5–96.8%, respectively). All families were resolved as monophyletic in all inferred trees. Similar topologies were obtained using the additional 12 subsets (24 trees), except for the suborders Dactylopteroidei (15 trees), Callionymoidei (4 trees), and Syngnathoidei (2 trees), which were not resolved as monophyletic in some trees, mostly those estimated using ASTRAL-II (13 ASTRAL-II trees vs. 5 RAxML trees; Appendix A Fig. S7). The relative placement of the family Centriscidae, most often resolved as a sister group to the clade composed of Aulostomidae + Fistulariidae, also varies in eight subset-based trees.

The MDS plots of assessment of topological disparity between the 28 trees inferred by different methods show that, regardless of the number of genes, RAxML and ASTRAL-II trees fall in opposite areas of the tree space, never overlapping (Appendix A Fig. S8). The ASTRAL-II trees reveal, however, a greater degree of topological disparity than the RAxML trees, a pattern that is probably the result of gene-tree error affecting ASTRAL-II reconstructions. While trees inferred with more than 300 loci (75% and 90% completeness matrices) tend to be more tightly clustered in the tree space relative to subset-based trees, RAxML and ASTRAL-II topologies obtained with the same genomic dataset or subset differ substantially (see Appendix A Fig. S8).

The time-calibrated phylogeny of syngnatharians based on 11 calibration points in MCMCTree is shown in Fig. 1.1 (RAxML ‘master tree’); results obtained with the 28 trees are summarized in Appendix A Table S3 and Fig. S9. Our inferred evolutionary timescale places the origin of total group Syngnatharia at 94.1 Ma (95% highest posterior density, HPD 92.0–99.3 Ma), whereas the crown group age is dated at 86.8 Ma (HPD 84.4–94.4 Ma) in the Late Cretaceous. With the exception of the long-stemmed Mulloidei, which originated at 18.0 Ma (HPD 14.9–21.8 Ma), the divergence of all other major suborder-level clades also took place in the Late Cretaceous (~70–83 Ma), long before the Cretaceous-Palaeogene (K-Pg) mass extinction event (~66 Ma).

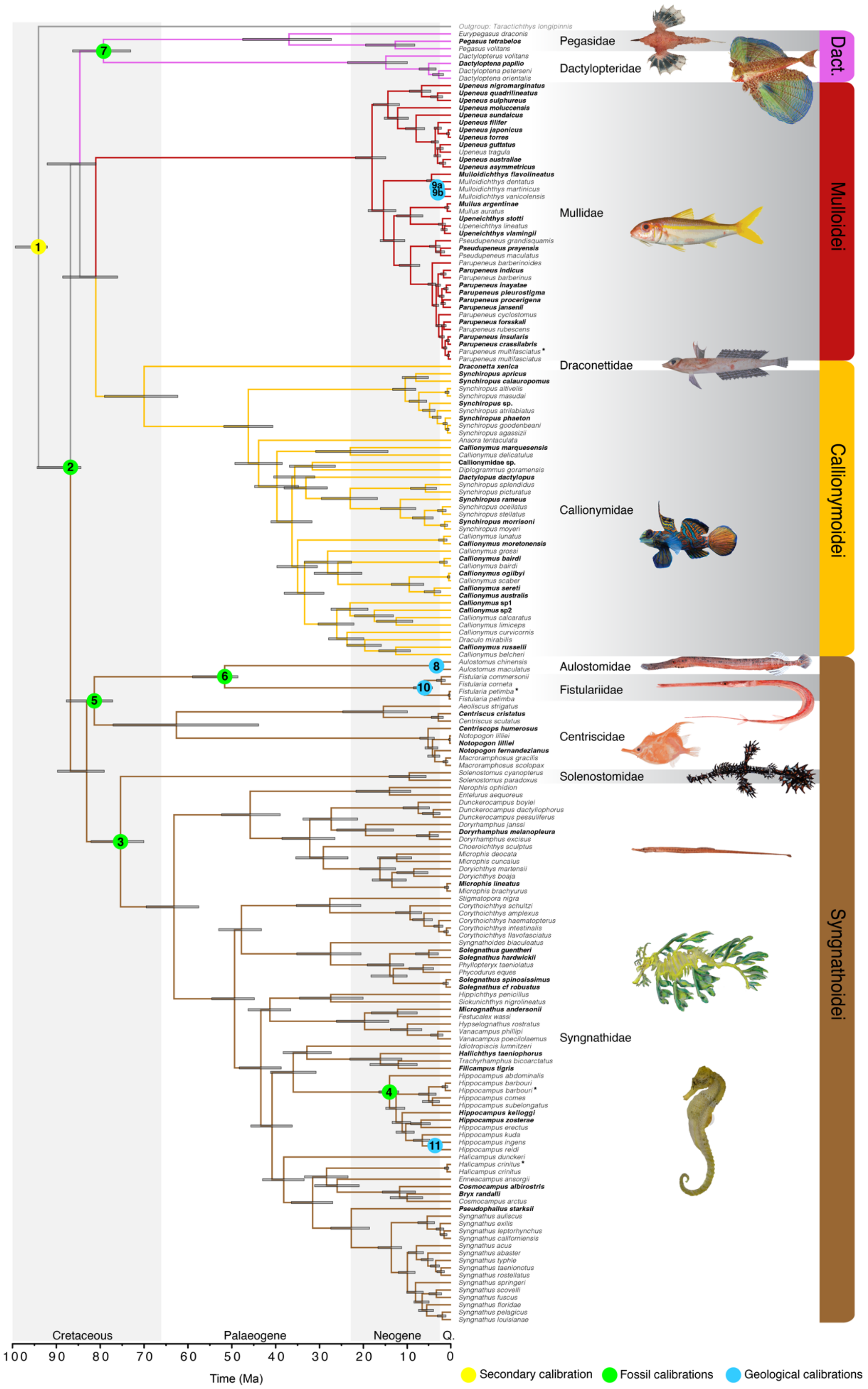

Outgroup: Taractichthys longipinnis
Eurypegasus draconis
Pegasus tetrabelos
Pegasus volitans
Dactylopterus volitans
Dactyloptena papilio
Dactyloptena peterseni
Dactyloptena orientalis
Upeneus nigromarginatus
Upeneus quadrilineatus
Upeneus sulphureus
Upeneus moluccensis
Upeneus sundaicus
Upeneus filifer
Upeneus japonicus
Upeneus torres
Upeneus guttatus
Upeneus tragula
Upeneus australiae
Upeneus asymmetricus
Mulloidichthys flavolineatus
Mulloidichthys dentatus
Mulloidichthys martinicus
Mulloidichthys vanicolensis
Mullus argentinae
Mullus auratus
Upeneichthys stotti
Upeneichthys lineatus
Upeneichthys vlamingii
Pseudupeneus grandisquamis
Pseudupeneus prayensis
Pseudupeneus maculatus
Parupeneus barberinoides
Parupeneus indicus
Parupeneus barberinus
Parupeneus inayatae
Parupeneus pleurostigma
Parupeneus procerigena
Parupeneus jansenii
Parupeneus cyclostomus
Parupeneus forsskali
Parupeneus rubescens
Parupeneus insularis
Parupeneus crassilabris
Parupeneus multifasciatus *
Parupeneus multifasciatus
Draconetta xenica
Synchiropus apricus
Synchiropus calauropomus
Synchiropus altivelis
Synchiropus masudai
Synchiropus sp.
Synchiropus atrilabiatus
Synchiropus phaeton
Synchiropus goodenbeani
Synchiropus agassizii
Anaora tentaculata
Callionymus marquesensis
Callionymus delicatulus
Callionymidae sp.
Diplogrammus goramensis
Dactylopus dactylopus
Synchiropus splendidus
Synchiropus picturatus
Synchiropus rameus
Synchiropus ocellatus
Synchiropus stellatus
Synchiropus morrisoni
Synchiropus moyeri
Callionymus lunatus
Callionymus moretonensis
Callionymus grossi
Callionymus bairdi
Callionymus bairdi
Callionymus ogilbyi
Callionymus scaber
Callionymus sereti
Callionymus australis
Callionymus sp1
Callionymus sp2
Callionymus calcaratus
Callionymus limiceps
Callionymus curvicornis
Draculo mirabilis
Callionymus russelli
Callionymus belcheri
Aulostomus chinensis
Aulostomus maculatus
Fistularia commersonii
Fistularia corneta
Fistularia petimba *
Fistularia petimba
Aeoliscus strigatus
Centriscus cristatus
Centriscus scutatus
Centriscops humerosus
Notopogon lilliei
Notopogon lilliei
Notopogon fernandezianus
Macroramphosus gracilis
Macroramphosus scolopax
Solenostomus cyanopterus
Solenostomus paradoxus
Nerophis ophidion
Entelurus aequoreus
Dunckerocampus boylei
Dunckerocampus dactyliophorus
Dunckerocampus pessuliferus
Doryrhamphus janssi
Doryrhamphus melanopleura
Doryrhamphus excisus
Choeroichthys sculptus
Microphis deocata
Microphis cuncalus
Doryichthys martensii
Doryichthys boaja
Microphis lineatus
Microphis brachyurus
Stigmatopora nigra
Corythoichthys schultzi
Corythoichthys amplexus
Corythoichthys haematopterus
Corythoichthys intestinalis
Corythoichthys flavofasciatus
Syngnathoides biaculeatus
Solegnathus guentheri
Solegnathus hardwickii
Phyllopteryx taeniolatus
Phycodurus eques
Solegnathus spinosissimus
Solegnathus cf robustus
Hippichthys penicillus
Siokunichthys nigrolineatus
Micrognathus andersonii
Festucalex wassi
Hypselognathus rostratus
Vanacampus phillipi
Vanacampus poecilolaemus
Idiotropiscis lumnitzeri
Halichthys taeniophorus
Trachyrhamphus bicoarctatus
Filicampus tigris
Hippocampus abdominalis
Hippocampus barbouri
Hippocampus barbouri *
Hippocampus comes
Hippocampus subelongatus
Hippocampus kelloggi
Hippocampus zosterae
Hippocampus erectus
Hippocampus kuda
Hippocampus ingens
Hippocampus reidi
Halicampus dunckeri
Halicampus crinitus *
Halicampus crinitus
Enneacampus ansorgii
Cosmocampus albirostris
Bryx randalli
Cosmocampus arctus
Pseudophallus starksii
Syngnathus auliscus
Syngnathus exilis
Syngnathus leptorhynchus
Syngnathus californiensis
Syngnathus acus
Syngnathus abaster
Syngnathus typhle
Syngnathus taenionotus
Syngnathus rostellatus
Syngnathus springeri
Syngnathus scovelli
Syngnathus fuscus
Syngnathus floridae
Syngnathus pelagicus
Syngnathus louisianae
Pegasidae
Dactylopteridae
Mullidae
Draconettidae
Callionymidae
Aulostomidae
Fistulariidae
Centriscidae
Solenostomidae
Syngnathidae
Dact.
Mulloidei
Callionymoidei
Syngnathoidei
1
2
3
4
5
6
7
8
9a
9b
10
11
Cretaceous
Palaeogene
Neogene
Q.
100
90
80
70
60
50
40
30
20
10
0
Time (Ma)
Secondary calibration
Fossil calibrations
Geological calibrations

Figure 1.1 Time-calibrated phylogeny for 169 species of Syngnatharia estimated in MCMCTree. The topology reflects the maximum-likelihood RAxML tree based on 932 UCEs (75% completeness matrix). Gray bars at nodes represent the 95% HPD intervals. Circles at the nodes indicate the 11 calibration points used in MCMCTree, which are colored according to the type of calibration used (see Appendix A, Supplementary Materials and Methods). For support values see Fig. S3. Dact.: Dactylopteroidei; Q.: Quaternary; Ma: millions of years. New taxa added for this study marked are shown in bold; asterisks (*) denote taxa examined by Longo et al. (2017) that were re-identified.

Sensitivity of biogeographic analyses to tree uncertainty, area schemes and the *j* parameter

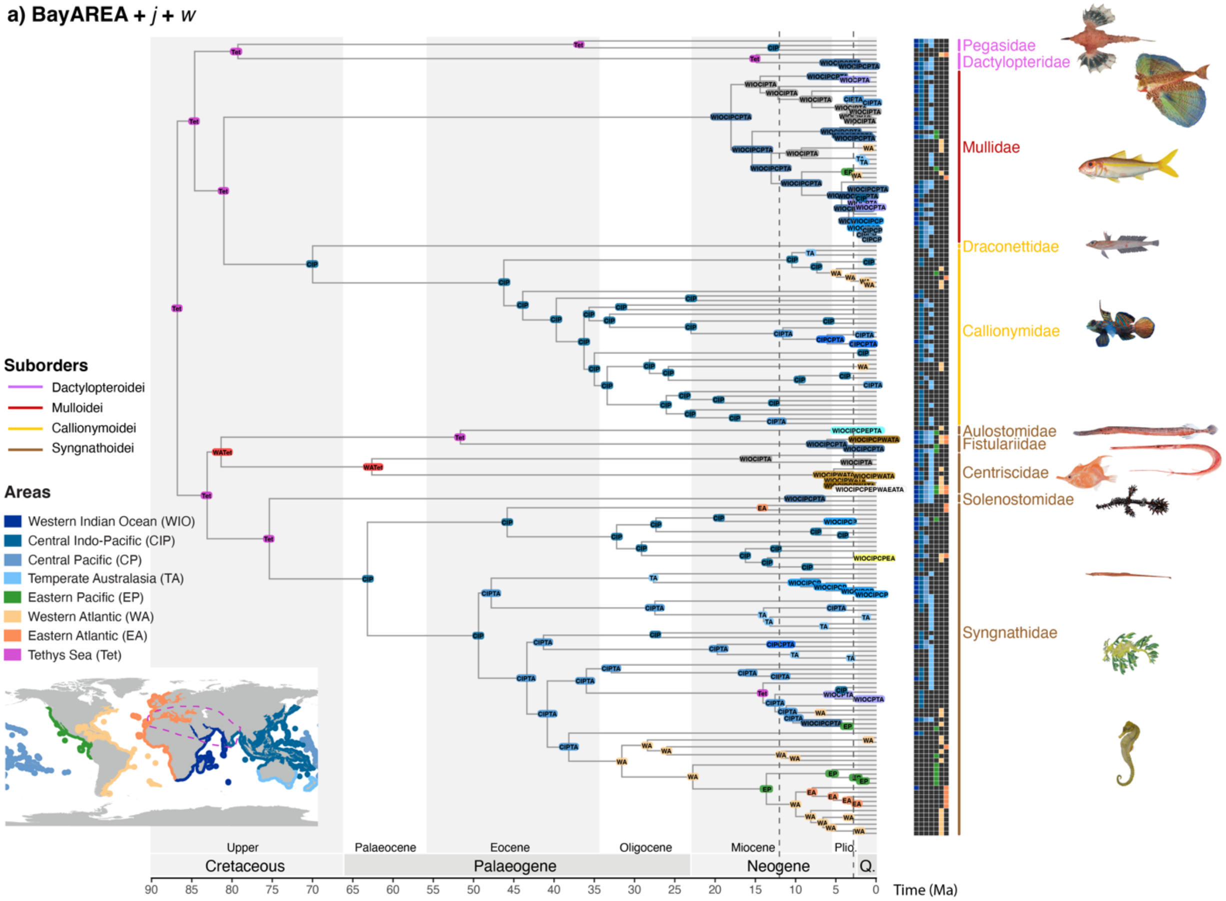
Sensitivity analyses to the three sources of variation examined reveal that the use of the *j* parameter has the strongest effect on ancestral range estimations, followed by the number of areas defined (see Fig. 1.3 and Appendix A Table S5 for details). Tree variance, by contrast, has a relatively smaller effect on the inferences (Appendix A Tables S5 and S6). The biogeographic reconstructions conducted to account for estimation error show that approximately one third of the colonization routes vary depending on the analysis (Fig. 1.3). These biogeographic patterns tend to be more similar between different area schemes using the same model rather than within each area scheme using different models (i.e., including or excluding the *j* parameter; Fig. 1.3 and Appendix A Table S5).

As expected, ancestral range estimates that use the BayAREA model along with the *j* parameter tend to identify more long-distance and recent dispersal events than those using the BayAREA model alone, which are otherwise more consistent with a Tethys Sea origin for many clades implying fewer colonization events due to widespread ancestors despite a lower model fit overall (AICc 1082–1313 for BayAREA vs. AICc 1015–1255 for BayAREA +*j*; Figs. 1.2 and 1.3). At least nine major differences in colonization routes are observed between analyses that include or exclude the *j* parameter, four of which are observed in the family Mullidae alone (Fig. 1.3). According to the -*j* inferences, this family colonized the WA from the Tethys Sea/IP at 17.6 Ma (HPD 16.1–19.2 Ma) followed by dispersal of *Mulloidichthys* from the IP to the EP via the EP Barrier at 3.9 Ma (HPD 3.2–4.6 Ma), and a subsequent dispersal at ~2.9 Ma into the WA through the Central American Seaway prior to the emergence of the Isthmus of Panama. This inference also suggests that *Pseudupeneus* colonized the EP from the WA at 2.9 Ma (HPD 2.2–3.5 Ma). In contrast, +*j* range estimates show that both *Mulloidichthys* and *Pseudupeneus* colonized the WA through the EP Barrier (3.9 Ma [HPD 3.2–4.6 Ma] and 6.6 Ma [HPD 3.5–9.6 Ma], respectively), whereas *Mullus* dispersed at 5.3 Ma (HPD 1–9.6 Ma) into that basin via the Cape of Good Hope (southern Africa; Fig. 1.3).

Ancestral range estimations also differ based on the number of areas used, but these are also largely confounded by the inclusion or exclusion of the *j* parameter. For example, biogeographic analyses based on the six- and eight-area schemes (±*j*) or the seven-area scheme (-*j*) indicate that the genera *Synchiropus* and *Hippocampus* are Tethyan relicts that colonized the WA at 27.9 Ma (HPD 10.1–45.0 Ma) and at 13.2 Ma (HPD 12.3–14.0 Ma), respectively.

Conversely, according to the seven-area +*j* inferences, *Synchiropus* took a different route via southern Africa to colonize the WA at 5.9 Ma (HPD 4.7–7.1 Ma), whereas *Hippocampus* colonized that basin in two independent dispersal events. The most recent common ancestor (MRCA) of *H. zostera* and *H. erectus* (8.8 Ma, HPD 7.1–10.5 Ma) dispersed via southern Africa, followed by the WA colonization of the MRCA of *H. reidi* and *H. ingens* (5.1 Ma, HPD 6.5–3.8 Ma), most likely through southern Africa (65.4% probability) rather than via the Central American Seaway after crossing the EP Barrier (26.9% probability). Noteworthy, while the six- and eight-area (±*j*) or the seven-area (-*j*) inferences suggest that *H. ingens* crossed the Central American Seaway and colonized the EP at ~3.8 Ma, the seven-area +*j* inference supports this colonization event but there is also a smaller probability (26.9%) that *H. reidi* could have crossed the seaway from the EP to the WA (vs. 65.4% through southern Africa). Lastly, the six- and eight-area schemes -*j* identify an additional colonization of the EP through the Central American Seaway in *Cosmocampus*. In summary, we find that for most clades the differences observed among area schemes are most striking when the *j* parameter is used (Fig. 1.3), particularly with the seven-area scheme.

Finally, analyses using the ‘master tree’ (Fig. 1.2 and Appendix A Figs. S10–S14) and the ‘alternative tree’, with and without the 28-tree averaging approach, resulted in rather similar biogeographic histories (Appendix A Tables S5 and S6) and colonization routes, with a few exceptions. For instance, in one estimation (7 areas, +*j*) the ‘alternative tree’ supports the colonization of crown Mullidae into the WA ~18 Ma, whereas the ‘master tree’ suggests that colonization of this area took place in three different mullid lineages (*Mulloidichthys, Mullus,* and *Pseudupeneus*) much later (~3.9–9.6 Ma). Likewise, only a few major differences are found between ancestral range estimates based on 28-tree averaging vs. single tree approaches. For example, inferences based on all 28 trees (summarized on either the master or the alternative trees), indicate that the Tethys Sea is the ancestral area state for the MRCA of *Dactylopterus* + *Dactyloptena*, crown *Aulostomus*, and several other lineages. In contrast, those nodes appear to be more widespread based on estimates that used either the ‘master tree’ or the ‘alternative tree’ alone (Appendix A Tables S5 and S6).


a) BayAREA + j + w
Suborders
Dactylopteroidei
Mulloidei
Callionymoidei
Syngnathoidei
Areas
Western Indian Ocean (WIO)
Central Indo-Pacific (CIP)
Central Pacific (CP)
Temperate Australasia (TA)
Eastern Pacific (EP)
Western Atlantic (WA)
Eastern Atlantic (EA)
Tethys Sea (Tet)
Pegasidae
Dactylopteridae
Mullidae
Draconettidae
Callionymidae
Aulostomidae
Fistulariidae
Centriscidae
Solenostomidae
Syngnathidae
Upper
Palaeocene
Eocene
Oligocene
Miocene
Plio.
Cretaceous
Palaeogene
Neogene
Q.
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
Time (Ma)

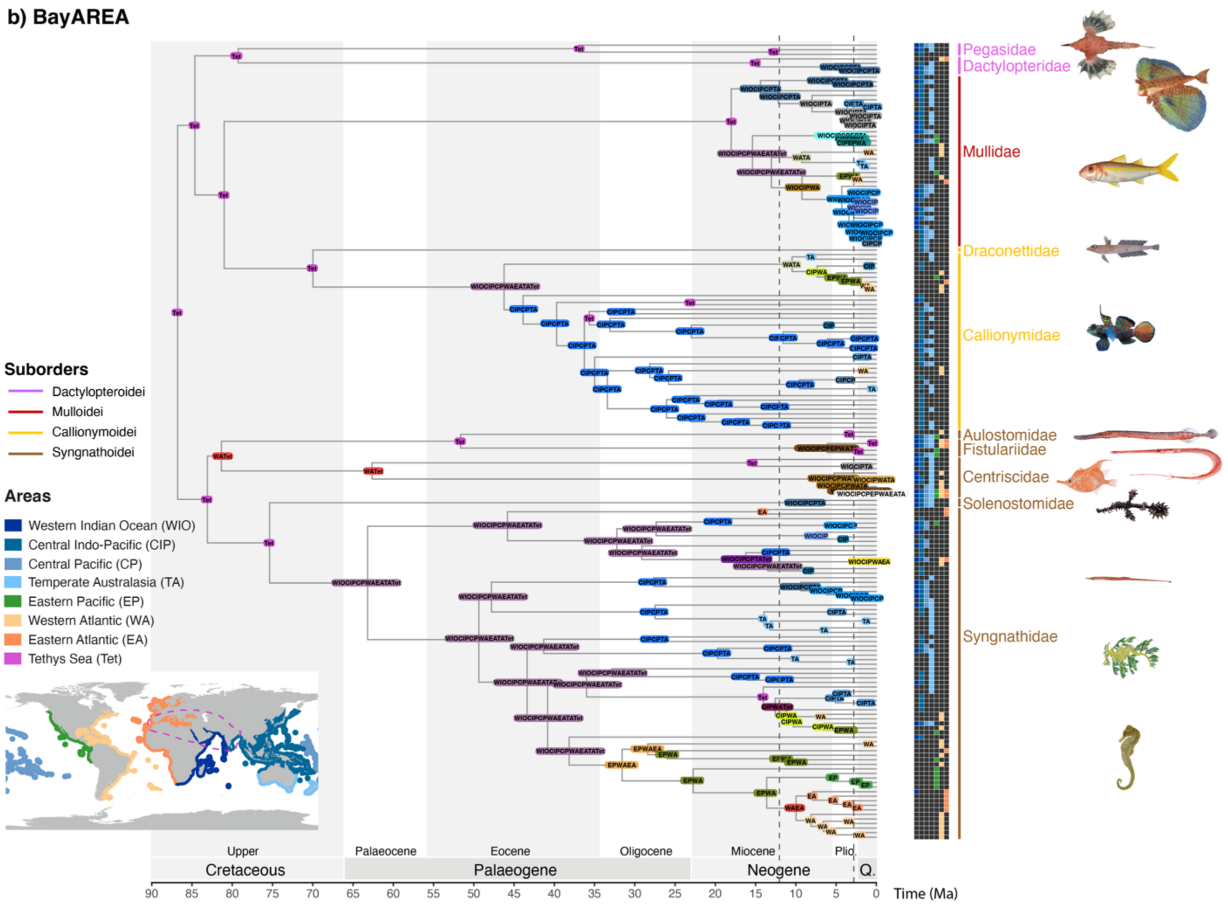


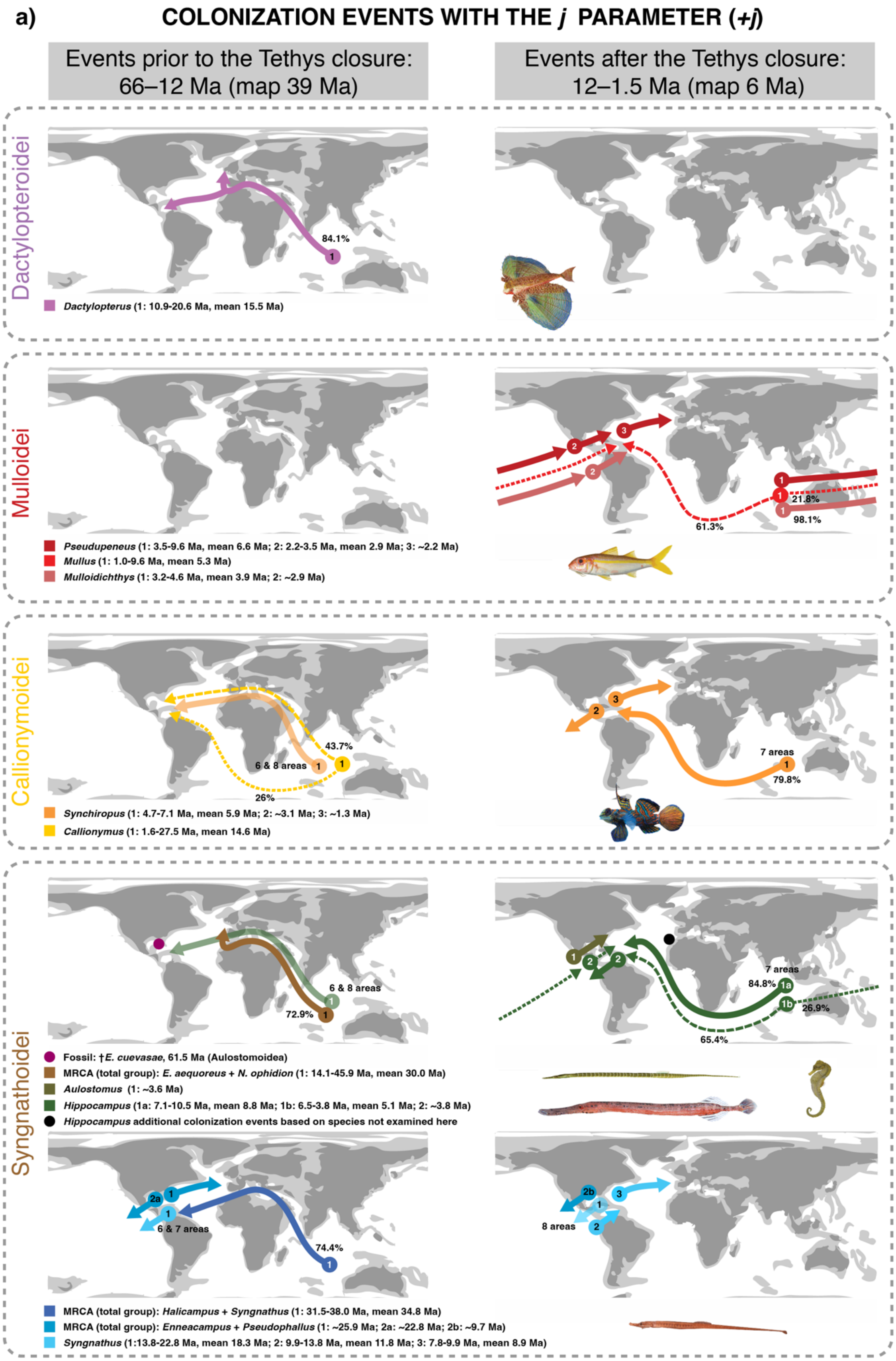
Figure 1.2 Ancestral range estimations for Syngnatharia based on the 7-area scheme applied to the 28 trees using the 'master tree' as fixed topology in BioGeoBEARS. **a)** Best-fit biogeographic model based on the BayAREA+*j*+*w* model. **b)** Given recent criticisms around the use of the *j* parameter (Ree and Sanmartín 2018) the BayAREA model is also reported here (see also Appendix A Table S4). Note that similar results were obtained with and without the *w* parameter, suggesting that this parameter alone is not a confounding factor in these comparisons. Size of boxes at the nodes are proportional to the number of areas in the estimated ancestral ranges. The map shows the seven marine biogeographic regions used to code the geographic distribution of extant species (based on Spalding et al. 2007; Kulbicki et al. 2013) and the ancient Tethys Sea. Families are color-coded by suborder. Dotted lines represent the time constraints corresponding to two major biogeographic events: the Tethys Seaway closure (12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Plio., Pliocene; Q., Quaternary; Ma: millions of years.

Ancestral range estimation and colonization of the Atlantic and eastern Pacific

Because of the uncertainties noted above, in this section we focus on identifying emergent patterns that are congruent across the different analyses to explain the biogeographic history of syngnatharians (Figs. 1.2 and 1.3). We chose to illustrate ancestral range estimates obtained on the basis of the seven-area scheme (Fig. 1.2) given that an important fraction (9.7%) of the species diversity in syngnatharians are endemic to TA (e.g., *Upeneichthys stotti*, *Solegnathus spinosissimus*, *Phycodurus eques*, *Phyllopteryx taeniolatus*; Hamilton et al. 2017). While the 8-area scheme also accounts for TA species, the additional area coded with this scheme (TNP) only includes a small proportion of endemics (1.1%).

Our analyses identified the ancient Tethys Sea as the center of origin for syngnatharians during the Late Cretaceous (86.8 Ma, HPD 84.4–94.4 Ma), followed by the origination of suborder-level lineages (~70–83 Ma) before the Cretaceous-Palaeogene (K-Pg) mass extinction event (~66 Ma). All families had an ancestor that was present in the Tethys Sea before their widespread colonization of the IP (WIO, CIP, CP, and/or TA). Syngnathidae originated in the late Cretaceous at 63.1 Ma (HPD 57.5–69.5 Ma) and started colonizing the New World in the Late Eocene. Subsequent to the origin of Centriscidae (62.6 Ma, HPD 43.9–77.1 Ma; Palaeocene/Eocene), the genera *Centriscus* and *Aeoliscus* persisted in the IP, while *Notopogon* and *Macroramphosus* dispersed into the Atlantic and EP during the Pliocene. Crown Callionymidae originated and diversified in the IP during the Eocene (46.2 Ma, HPD 40.6–51.8 Ma) while lineages in the genera *Callionymus* and *Synchiropus* colonized the Atlantic later in the Miocene. The family Dactylopteridae, which originated in the Tethys Sea/IP, also colonized the Atlantic in the Middle Miocene (genus *Dactylopterus*; 14.6 Ma, HPD 10.0–23.5 Ma). Among members of the family Mullidae (origin dated at 18.0 Ma, HPD 14.9–21.8 Ma), only the genus *Upeneus* remained restricted to the ancestral IP range, while the rest of the genera in the family colonized the Atlantic and the EP during the Miocene. Whereas the total group origin for the families Fistulariidae and Aulostomidae dates back to 51.6 Ma (HPD 48.6–58.9 Ma), their crown members diversified more recently at 6.1 Ma (HPD 4.3–8.6Ma; Miocene/Pliocene) and 3.3 Ma (HPD 2.8–4.5 Ma; Pliocene/Pleistocene), respectively. Our biogeographic analyses indicate that these two families dispersed into the three major basins (from a Tethys Sea ancestor of the total group Fistulariidae + Aulostomidae) around the Pliocene. Draconettidae (79.3 Ma; HPD 73.0–86.3 Ma; Late Cretaceous), Pegasidae (36.9 Ma; HPD 27.3–47.5 Ma; Eocene/Oligocene), and Solenostomidae (9.5 Ma; HPD 5.7–14.1 Ma; Neogene) are the only families that did not disperse outside their center of origin. Most of the genera also are inferred to have a Tethys Sea/IP (WIO, CIP, CP, or TA) origin except for two genera in the family Syngnathidae, *Enneacampus* and *Pseudophallus*, which probably had a WA origin.

Our ancestral range estimates combined with stochastic mapping suggest different routes of colonization to the EP and the Atlantic (Fig. 1.3). Except for the widespread species with circumglobal or semi-circumglobal distributions, the EP was colonized 6–8 times, with at least one event taking place eastwards across the EP Barrier (*Mulloidichthys*) and the rest occurring via the WA through the Central American Seaway before the closure of the Isthmus of Panama (*Synchiropus*, *Hippocampus*, *Pseudophallus*; Fig. 1.3). Similarly, the Atlantic was colonized 6–14 times through three different routes: (i) 6–8 lineages were either Tethyan relicts or crossed the Tethys Seaway before its closure *ca.* 12–18 Ma (Steininger and Rögl 1979; Adams et al. 1983; Rögl 1998, 1999; e.g., *Dactylopterus volitans*, *Entelurus aequoreus* + *Nerophis ophidion*), (ii) 1–4 lineages colonized the Atlantic via southern Africa (e.g., *Mullus*), (iii) and at least one lineage passed from the EP to the Atlantic prior to the emergence of the Isthmus of Panama >2.8 Ma (e.g., *Mulloidichthys*). Finally, four different syngnatharian lineages (*Pseudupeneus prayensis*, *Synchiropus phaeton*, *Enneacampus ansorgii*, and some species in *Syngnathus*) colonized the EA via a west-to-east Atlantic route (Fig. 1.3).


a)
COLONIZATION EVENTS WITH THE *j* PARAMETER (+*j*)
Events prior to the Tethys closure: 66–12 Ma (map 39 Ma)
Events after the Tethys closure: 12–1.5 Ma (map 6 Ma)
Dactylopteroidei
84.1%
1
■ *Dactylopterus* (1: 10.9-20.6 Ma, mean 15.5 Ma)
Mulloidei
1
2
3
21.8%
98.1%
61.3%
■ *Pseudupeneus* (1: 3.5-9.6 Ma, mean 6.6 Ma; 2: 2.2-3.5 Ma, mean 2.9 Ma; 3: ~2.2 Ma)
■ *Mullus* (1: 1.0-9.6 Ma, mean 5.3 Ma)
■ *Mulloidichthys* (1: 3.2-4.6 Ma, mean 3.9 Ma; 2: ~2.9 Ma)
Callionymoidei
43.7%
6 & 8 areas
26%
7 areas
79.8%
1
2
3
■ *Synchiropus* (1: 4.7-7.1 Ma, mean 5.9 Ma; 2: ~3.1 Ma; 3: ~1.3 Ma)
■ *Callionymus* (1: 1.6-27.5 Ma, mean 14.6 Ma)
Syngnathoidei
6 & 8 areas
72.9%
7 areas
84.8%
1a
1b
26.9%
65.4%
1
2
● Fossil: †*E. cuevasae*, 61.5 Ma (Aulostomoidea)
■ MRCA (total group): *E. aequoreus* + *N. ophidion* (1: 14.1-45.9 Ma, mean 30.0 Ma)
■ *Aulostomus* (1: ~3.6 Ma)
■ *Hippocampus* (1a: 7.1-10.5 Ma, mean 8.8 Ma; 1b: 6.5-3.8 Ma, mean 5.1 Ma; 2: ~3.8 Ma)
● *Hippocampus* additional colonization events based on species not examined here
2a
1
6 & 7 areas
74.4%
2b
3
8 areas
■ MRCA (total group): *Halicampus* + *Syngnathus* (1: 31.5-38.0 Ma, mean 34.8 Ma)
■ MRCA (total group): *Enneacampus* + *Pseudophallus* (1: ~25.9 Ma; 2a: ~22.8 Ma; 2b: ~9.7 Ma)
■ *Syngnathus* (1:13.8-22.8 Ma, mean 18.3 Ma; 2: 9.9-13.8 Ma, mean 11.8 Ma; 3: 7.8-9.9 Ma, mean 8.9 Ma)

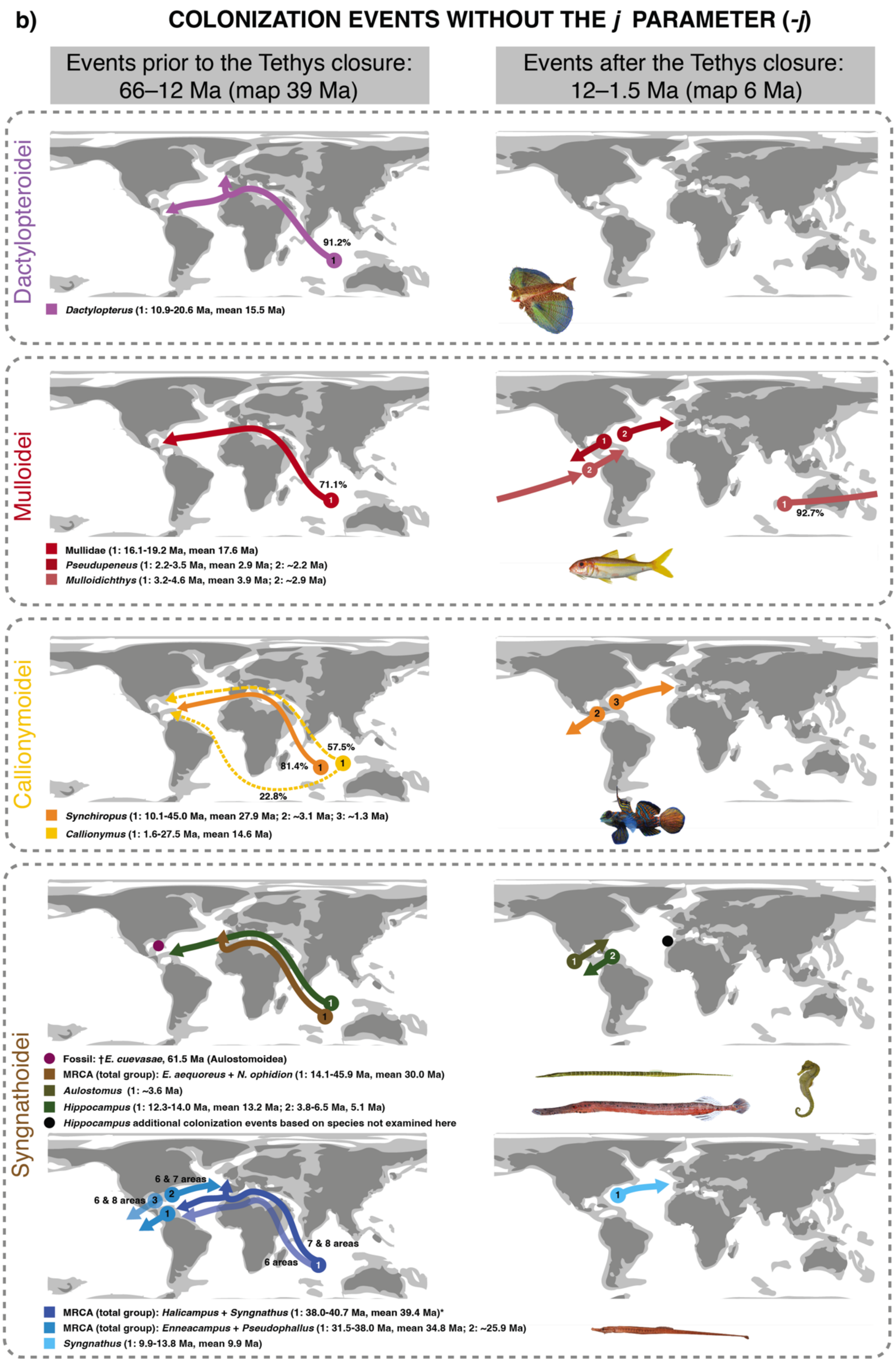
b)
COLONIZATION EVENTS WITHOUT THE j PARAMETER (-j)
Events prior to the Tethys closure: 66–12 Ma (map 39 Ma)
Events after the Tethys closure: 12–1.5 Ma (map 6 Ma)
Dactylopteroidei
91.2%
1
🟪 Dactylopterus (1: 10.9-20.6 Ma, mean 15.5 Ma)
Mulloidei
71.1%
1
2
92.7%
🟥 Mullidae (1: 16.1-19.2 Ma, mean 17.6 Ma)
🟥 Pseudupeneus (1: 2.2-3.5 Ma, mean 2.9 Ma; 2: ~2.2 Ma)
🟥 Mulloidichthys (1: 3.2-4.6 Ma, mean 3.9 Ma; 2: ~2.9 Ma)
Callionymoidei
57.5%
81.4%
22.8%
1
2
3
🟧 Synchiropus (1: 10.1-45.0 Ma, mean 27.9 Ma; 2: ~3.1 Ma; 3: ~1.3 Ma)
🟨 Callionymus (1: 1.6-27.5 Ma, mean 14.6 Ma)
Syngnathoidei
1
2
● Fossil: †E. cuevasae, 61.5 Ma (Aulostomoidea)
🟫 MRCA (total group): E. aequoreus + N. ophidion (1: 14.1-45.9 Ma, mean 30.0 Ma)
🟫 Aulostomus (1: ~3.6 Ma)
🟩 Hippocampus (1: 12.3-14.0 Ma, mean 13.2 Ma; 2: 3.8-6.5 Ma, 5.1 Ma)
● Hippocampus additional colonization events based on species not examined here
6 & 7 areas
6 & 8 areas
7 & 8 areas
6 areas
1
2
3
🟦 MRCA (total group): Halicampus + Syngnathus (1: 38.0-40.7 Ma, mean 39.4 Ma)*
🟦 MRCA (total group): Enneacampus + Pseudophallus (1: 31.5-38.0 Ma, mean 34.8 Ma; 2: ~25.9 Ma)
🟦 Syngnathus (1: 9.9-13.8 Ma, mean 9.9 Ma)

Figure 1.3 Possible dispersal routes that led to the historical colonization of the Atlantic and eastern Pacific basins (from a Tethys Sea or Indo-Pacific basin) in different clades of syngnatharian fishes as inferred using two alternative models. Ancestral ranges estimated using the BayAREA model: **a)** including the long-distance dispersal parameter (*j*), and **b)** without the *j* parameter. Solid arrows indicate ancestral range estimations using the favored 7-area scheme (see main text), and differences obtained with alternative area schemes (6 or 8) are denoted with transparency. The timing of dispersal events indicated are based on the 7-area scheme alone (Fig. 1.2; but see also Appendix A Figs. S13 and S14 for ages inferred with 6 and 8 areas). Dispersal routes mapped are macroevolutionary in scope, involving vicariant speciation events leading to the origin of at least one lineage restricted to one of the major basins. Clades including multiple widespread species with circumtropical and/or circum-temperate distributions (e.g., *Aulostomus chinensis*, *Fistularia* spp., *Doryrhamphus excisus*, *Syngnathus acus*, Centriscidae spp., and *Mulloidichthys vanicolensis*) are better examined using phylogeographic analyses (e.g., Bowen et al. 2001; Lessios and Robertson 2006) and are thus not mapped here. In cases where multiple routes are possible (e.g., *Dactylopterus volitans*), route probability is estimated based on biogeographic stochastic mapping using the 'master tree' (indicated with dashed lines; size of dashes are proportional to the probability). MRCAs indicate events for total groups (crown and stem lineages) given by the two taxa indicated in each case. Age ranges indicated per event denote the minimum and maximum ages for crown vs. stem clades obtained from the 28 inferred trees. Due to age uncertainty and overlapping, some events could be depicted on both maps; thus, selection of maps for event depiction is based on mean ages. Fish drawings are shown only for clades involved in mapped events. Maps modified from GPlate (Müller et al. 2018) represent the mean age from the following time slices: 12–66 Ma and 0–12 Ma. Ma: millions of years.

1.5 Discussion

We investigated the evolutionary and biogeographic history of marine fish species in Syngnatharia by combining genomic (UCEs), paleontological, geologic, and geographic data layers. Although the biogeographic history of a few families or genera in this group have been examined in detail––for example, Aulostomidae (Bowen et al. 2001), Syngnathidae (Hamilton et al. 2017), *Hippocampus* (Teske et al. 2004, 2007; Boehm et al. 2013; Li et al. 2021), and *Mulloidichthys* (Lessios and Robertson 2013)—this is the first biogeographic study conducted for the entire clade. Our analyses accounting for topological, temporal, and biogeographic uncertainty support a Late Cretaceous origin of syngnatharians in the Tethys Sea, with subsequent dispersal into the central Indo-Pacific and independent colonizations of the eastern Pacific and the Atlantic by most families through alternative routes.

Evolutionary relationships and divergence times

The phylogenetic relationships among major clades differ slightly depending on the methodological approach used. In agreement with other recent studies (Longo et al. 2017; Alfaro et al. 2018; Fig. 1.1 and Appendix A Figs. S3 and S5), concatenation-based analyses resolved an early split that supports the reciprocal monophyly of the long-snouted Syngnathoidei and a clade including the bottom-dwelling suborders Mulloidei, Callionymoidei, and Dactylopteroidei, most of which are short-snouted. The exception to this is Pegasidae, which like Syngnathoidei has an elongated snout due to enlargement of specialized bones of the neurocranium and suspensorium (Bergert and Wainwright 1997). The mouth in pegasids, however, is oriented ventrally (vs. terminal in syngnathoids; Pietsch 1978; Bergert and Wainwright 1997). Coalescent-based analyses, in contrast, resolved the suborder Syngnathoidei nested within a paraphyletic group comprising the bottom-dwelling suborders (Appendix A Figs. S4 and S6). Remarkably, irrespective of the reconstruction method, all analyses resolved a monophyletic Syngnathoidei, which is consistent with results from many previous studies (e.g., Near et al. 2013 and Betancur-R. et al. 2017, based on exonic data; Longo et al. 2017 and Alfaro et al. 2018, based on UCE data), but contrasts with others that examined fewer genes (Near et al. 2012; Betancur-R. et al. 2013; Song et al. 2014) or taxa (Hughes et al. 2018) rejecting the monophyly of the suborder. The relationships among major lineages of Syngnathidae, the most diverse syngnatharian family, are also largely congruent with previous studies, showing an early divergence between trunk- and tail-brooders (Wilson and Orr 2011; Hamilton et al. 2017). Finally, while this and other previous phylogenetic studies provide support for the monophyly of all genera in Mullidae (Kim 2002; Longo et al. 2017), the resolution of other intrafamilial relationships is more elusive, including the nonmonophyly of genera in Callionymidae (e.g., *Synchiropus* and *Callionymus*) and Syngnathidae (e.g., *Microphis* and *Cosmocampus*).

Our results indicate a Late Cretaceous (~86.8 Ma) origin of crown Syngnatharia, which is roughly ten million years older than the evolutionary timescales inferred by recent studies (Near et al. 2012; Betancur-R. et al. 2017; Alfaro et al. 2018; Hughes et al. 2018). These remarkable differences are likely the result of new interpretations for the age of the fossil †*Gasterorhamphosus zuppichinii* (Sorbini 1981), which implies that all previous studies underestimated the group's crown age. This interpretation is based on a recent stratigraphic analysis of the Calcari di Melissano, showing a lower Campanian (83.6 Ma; Schlüter et al. 2008) instead of a Maastrichtian (72.1 Ma; Sorbini 1981) age for this formation (see additional details in the Appendix A, Supplementary Materials and Methods). Although most family-level diversification events happened during the Cenozoic, the origin of suborders and most family-level total group predates the end of the Cretaceous (Fig. 1.1). This result runs counter to the notion that the divergence of major syngnatharian lineages is associated with the K-Pg mass extinction (Alfaro et al. 2018). While the split between Draconettidae and Callionymidae as well as the crown ages for Centriscidae and Syngnathidae are estimated to be around the K-Pg, our time-calibrated trees show no signs of diversification bursts associated with this extinction event. Instead, they reveal that the early Eocene was a period of exceptional diversification within Syngnathidae and Callionymidae. This period coincides with the early expansion of other reef-associated families (e.g., Apogonidae, Labridae, Pomacentridae; Cowman and Bellwood 2011; Bellwood et al. 2017; Fig. 1.1). The origin of crown Mulloidei, the youngest among syngnatharians suborders, has been linked to a rapid diversification process associated with extensive coral reef rearrangements as a result of the origin of the Indo-Australian-Archipelago (IAA) marine biodiversity hotspot in the Miocene (Renema et al. 2008; Bellwood et al. 2017; Siqueira et al. 2019, 2020; Fig. 1.1).

Uncertainties in biogeographic analyses

Ancestral range estimations are typically inferred using a single tree and a predefined area scheme, resolving the most probable history based on a single best-fit model (e.g., Feng et al. 2017; Tea et al. 2019; Varela et al. 2019; Batista et al. 2020). In some cases, however, a set of trees are sampled from the Bayesian posterior distribution and used to gauge the effect of alternative phylogenetic resolutions (e.g., Berger et al. 2016). Here, we inferred the biogeographic history of syngnatharians based on comprehensive approaches designed to better capture uncertainties in estimated ancestral ranges. Given that the implementation of the founder-event speciation or jump-dispersal (*j*) parameter has been suggested to favor an unparsimonious numbers of long-distance dispersal events (Ree and Sanmartín 2018), we examined the results of our best-fit biogeographic model (BayAREA), both with and without the *j* parameter. As expected, the addition of the *j* parameter resulted in a better-fit model in all cases, increasing the probability of long-distance and more recent colonization events (Figs. 1.2 and 1.3). Overall, the inclusion/exclusion of the *j* parameter had a stronger effect on our biogeographic inferences relative to the number of areas considered or the alternative topologies used, despite considerable topological discordance. Furthermore, most discrepancies were observed among different area schemes with models that incorporate the *j* parameter, indicating a confounding interaction between these two variables (Fig. 1.3).

After the concerns raised by Ree and Sanmartín (2018) a number of studies using BioGeoBEARS have omitted the *j* parameter (e.g., Dong et al. 2018; Vargas and Dick 2020), including a recent investigation of the biogeography of marine angelfishes (Baraf et al. 2019). We believe that *j* can be informative for modelling the biogeography of marine fishes in general and reef-associated fishes in particular (like most syngnatharians), which can feature long-distance dispersal during pelagic larval stages or through rafting (e.g., sargassum-associated species) aided by oceanic currents (Luiz et al. 2012). Noteworthy, the critique of Ree and Sanmartín (2018) regarding the implementation of *j* in a model-fitting framework (e.g., by comparing DEC and DEC+*j*) was more recently challenged by Klaus and Matzke (2020) on the basis of previously conducted simulations (Matzke 2014), a review of a number of empirical studies that do not seem to show inflated likelihood scores in favor of *j*, and Ree and Sanmartín's (2018) use of a small hypothetical dataset to emphasize their points (Klaus and Matzke 2020). Given these ongoing debates, we opt to focus on the similarities obtained between the two different estimations (with and without *j*), rather than their differences, to investigate the biogeographic history of syngnatharians (see next section).

Another factor of uncertainty relates to the use of alternative phylogenies to conduct ancestral range estimations. In this case, tree uncertainty appears to have a much smaller effect in this study possibly because early branching lineages that show a higher degree of topological discordance (e.g., full dataset vs. subsets; concatenation vs. multi-species coalescent analyses; Appendix A Fig. S7) are invariably estimated as having a Tethys/Indo-Pacific origin (Appendix A Tables S5 and S6). Therefore, relatively lower sensitivity to phylogenetic uncertainty is probably a factor that is case-specific to syngnatharians and should not be generalized to other groups.

Finally, to identify the alternative colonization routes that different lineages followed, we calculated their relative probabilities using biogeographic stochastic mapping or BSM (Fig. 1.3).

In some cases, the probability of a specific route was high. For example, *Mulloidichthys* dispersed through the EP Barrier with ~94% probability, and *Dactylopterus* colonized the Atlantic via the Tethys Seaway with ~87% probability. In other cases, BSM resulted in greater uncertainties, such as in *Callionymus* that colonized the western Atlantic through two possible routes (~50% via Tethys Seaway and ~25% via southern Africa). More alternative dispersal routes were available before the closure of the Tethys Seaway than after its closure, increasing the challenge to infer the pathways lineages followed. However, our analyses suggest that the most likely colonization route to the Atlantic for older dispersal events occurred via the Tethys Seaway (Fig. 1.3).

Several previous studies have examined the biogeographic history of Syngnathidae (Hamilton et al. 2017), particularly seahorses (genus *Hippocampus*, Teske et al. 2004, 2007; Li et al. 2021). These previous studies, however, did not consider biogeographic or phylogenetic variance, identifying two independent colonizations of the Atlantic by seahorse lineages—an ancient event (14.2–15.12 Ma Teske et al. 2007; 13.6–15.6 Ma Li et al. 2021) via the Tethys Seaway (MRCA *H. zosterae* + *H. erectus*) and a younger event through either The Cape of Good Hope (South Africa; 3.1–4.6 Ma Teske et al. 2007; 3.6–4.9 Ma Li et al. 2021) or the EP Barrier (3.1–4.6 Ma; MRCA *H. algiricus* + *H. ingens*, Teske et al. 2007). By contrast, although our results and those from previous studies concur in identifying an Indo-Pacific (CIP + TA) origin of seahorses (Teske et al. 2004, 2007; Hamilton et al. 2017; Li et al. 2021), we found two possible biogeographic histories for the colonization of the Atlantic (Figs. 1.2 and 1.3). These include a single colonization event (12.3–14.0 Ma) through the Tethys Seaway in the MRCA of *H. subelongatus* + *H. ingens*, and two colonization events taking place after the closure of the Tethys Seaway via southern Africa (6.8–10.3 Ma and 3.7–6.5 Ma, respectively). It is also possible that the first colonization event of the Atlantic in *Hippocampus* suggested by these previous studies (or the single colonization proposed here) happened southwestwards via The Cape of Good Hope rather than northwestwards through the Tethys Seaway, as the seaway started to close at 18 Ma (Steininger and Rögl 1979; Adams et al. 1983; Rögl 1998, 1999).

The disagreements regarding the alternative colonization routes of the Atlantic in *Hippocampus* appear to stem from conflicts associated with divergence time estimations. The timing of the second Atlantic colonization was similar among the three studies, likely an indication of the similar use of a geological calibration based on the Isthmus of Panama for the MRCA of the geminate species pair *H. ingens* and *H. reidi*. (Note that Teske et al. 2004 and Hamilton et al. 2017 did not date their trees). The major difference concerns the age of the first colonization event (see above) and the age of crown *Hippocampus* (~24 Ma in Li et al. 2021 vs. ~14 Ma in this study; note that Teske et al. 2007 did not infer an age for this node), both of which are substantially older than our estimates (Appendix A Table S7). These studies either applied a fossil calibration to the crown *Hippocampus* where it should have been instead placed in the stem lineage (see Žalohar and Hitij 2012; Li et al. 2021) or used geological calibrations only (Teske et al. 2007). In contrast, our divergence time estimates are based on both primary fossil and geological calibrations available for the entire Syngnatharia, as well as a secondary root calibration based on multiple global ray-finned fish time-calibrated trees that used dozens of primary fossil calibrations (Appendix A Tables S1 and S2).

Biogeographic history of Syngnatharia

Irrespective of the model, number of areas or tree used, all ancestral range estimations invariably reveal that the center of origin for syngnatharians was the ancient Tethys Sea. Ancestral lineages subsequently occupied the Indo-Pacific Ocean, followed by multiple independent colonization events of the eastern Pacific and the Atlantic basins, which took place via different routes both before and after the closure of the Tethys Seaway (Figs. 1.2 and 1.3 and Appendix A Figs. S10–S14). Like most marine fishes with long-distance dispersal capabilities, these patterns show that syngnatharians are successful at colonizing different oceanic realms. In fact, some species have circumglobal distributions or occur in at least two major basins (e.g. *Aulostomus chinensis*, *Fistularia* spp., *Doryrhamphus excisus*, *Syngnathus acus*, Centriscidae spp., and *Mulloidichthys vanicolensis*).

The western Tethys Sea was one of the richest regions for fossil teleost species, being a hotspot of marine biodiversity in the Eocene (Renema et al. 2008; Friedman and Carnevale 2018). In fact, the majority of syngnatharian fossils are currently described from different time horizons in that region (Sorbini 1981; Žalohar et al. 2009; Carnevale et al. 2014), suggesting that the Tethys Sea was the center of origin for the group, as it has been shown for other reef-fish families and reef-associated taxa (Renema et al. 2008; Cowman and Bellwood 2013a; Siqueira et al. 2019). The Terminal Tethyan Event (TTE), which opened and closed intermittently between 12 and 18 Ma, divided the western and the eastern Tethys Sea (Steininger and Rögl 1979; Adams et al. 1983; Rögl 1998, 1999) and created an important oceanic barrier that shaped the dispersal of marine species (Bellwood and Wainwright 2002; Barber and Bellwood 2005; Cowman et al. 2009, 2017; Cowman and Bellwood 2013b; Hou and Li 2017). Before its final closure (12 Ma), some lineages dispersed eastwards to the Indo-Pacific where diversification of most of the families occurred (e.g., Syngnathidae, Callionymidae). Rather than dispersing via The Cape of Good Hope (southern Africa), most other lineages colonized the Atlantic from the western Tethys Sea or crossed the Tethys Seaway from the Tethys Sea/Indo-Pacific (e.g., *Synchiropus*, *Callionymus*, *Dactylopterus volitans*, *Entelurus aequoreus* + *Nerophis ophidion*). During the Miocene, the syngnatharian biodiversity hotspot moved to the IAA where most reef fish clades also originated and diversified (e.g. Pomacanthidae, Baraf et al. 2019; Lutjanidae, Rincon-Sandoval et al. 2020). After the closure of the Tethys Seaway, lineages took the two remaining routes available for dispersal towards the Atlantic and the eastern Pacific: the EP Barrier or southern Africa. Like other reef fishes (Lessios and Robertson 2006), goatfish (*Mulloidichthys*) and pipefish (*Doryrhamphus*) genera crossed the EP Barrier to colonize the eastern Pacific. Other genera in these groups (e.g., *Mullus* and *Microphis*), however, most probably dispersed into the western Atlantic via southern Africa through the warm Agulhas rings that occasionally penetrates into the Atlantic (see also Rocha et al. 2005; Floeter et al. 2008). See above regarding uncertainties in dispersal routes to the Atlantic in *Hippocampus*.

During the Neogene, the rising of the Isthmus of Panama interrupted gene flow between the western Atlantic and the eastern Pacific, ultimately producing many geminate species pair (see Lessios 2008 for a review). Before its final closure at some point before 2.8 Ma (Montes et al. 2015; O'Dea et al. 2016), many lineages dispersed across the Central American Seaway in either direction (i.e., goatfishes, dragonets, trumpetfishes, seahorses, and pipefishes), but predominantly through a WA-to-EP route (at least five events vs. two events from eastern Pacific to western Atlantic). Similar asymmetric dispersal routes between these two basins have been reported for other groups—e.g., Lutjanidae (Rincon-Sandoval et al. 2020), Gobiidae (Thacker 2015), and *Apogon* (Thacker 2017). Finally, lineages that crossed the Mid-Atlantic Barrier to colonize both sides of the Atlantic either took a westwards route before the Tethys Seaways closure—probably via the North Equatorial Current (e.g., flying gurnards, pipefishes)—or dispersed eastwards no later than ~10 Ma, most likely using the Gulf Stream current (e.g., West African goatfish, Phaeton dragonet, pipefishes; see also Floeter et al. 2008; Luiz et al. 2012). While the predominant dispersal mode for these and other reef fish groups is via planktonic larvae, pipefishes and seahorses can also disperse by rafting on pelagic *Sargassum* and other macroalgae (Teske et al. 2005; Casazza and Ross 2008; Woodall 2009; Luiz et al. 2012; Boehm et al. 2013; Hamilton et al. 2017).

Aside from the Tethys Sea/Indo-Pacific region, the temperate Australasia and the Atlantic have also served as a center of origin for some genera (Fig. 1.2; Hamilton et al. 2017). The temperate Australasia region harbors significant endemicity and biodiversity of syngnatharians (particularly syngnathids) due to its extensive coastal seagrass habitats and its unique oceanographic conditions (Poore 1995; Shepherd and Edgar 2013). In fact, our analyses suggest that seadragons (*Phycodurus* and *Phyllopteryx*) and pipehorses (*Solegnathus*) originated in temperate Australasia. In the western Atlantic, the genus *Pseudophallus* is restricted to freshwater and brackish waters in Central and South America, whereas *Enneacampus* is currently distributed in the eastern Atlantic, suggesting that its ancestral lineage crossed the Mid-Atlantic Barrier eastwards (Floeter et al. 2008). A caveat of our study is that incomplete biogeographic sampling for some taxa may have affected our ancestral range estimation analyses, leading to area misplacement of lineage origin or underestimation of the number of colonization events. For instance, the genera *Bryx* and *Cosmocampus* have circumtropical distributions, but our sampling only includes species from western Atlantic and eastern Pacific/western Atlantic, respectively. Likewise, we lack representatives for the genus *Hippocampus* distributed in the eastern Atlantic, an area that *H. hippocampus* is thought to have recolonized from the western Atlantic by crossing the Mid-Atlantic Barrier eastwards via the Gulf Stream Current (Teske et al. 2007; Boehm et al. 2013; Li et al. 2021).

1.6 Conclusion

Our study uses an integrative approach in a robust phylogenomic framework to account for a number of uncertainties in phylogenetic comparative inferences to trace the biogeographic history of Syngnatharia. We identified multiple independent colonizations of the Atlantic and the eastern Pacific and inferred possible dispersal routes from the Indo-Pacific and their center of origin, the Tethys Sea. While for some lineages the biogeographic history did not change using different area schemes or including/excluding the jump-dispersal (*j*) parameter, for other clades we identified a number of alternative colonization timings and routes, particularly when different area schemes are implemented in combination with the *j* parameter. Contrary to the common practice of estimating the biogeographic history using a single tree, a predefined set of areas, and a single biogeographic model, our study highlights the importance of embracing uncertainty in ancestral range estimations. We show that the common practice can be overly simplistic, failing to capture intrinsic complexities in historical biogeographic inferences. Our results ultimately provide a robust framework to address future questions on the evolutionary history of syngnatharians, such as understanding the factors driving their evolutionary radiation and explaining the uneven richness and morphological disparity across globally distributed clades.

1.7 References

Adams C.G., Gentry A.W., Whybrow P.J. 1983. Dating the terminal Tethyan event. Utrecht, the Netherlands: Utrecht University: Utrecht Micropaleontological Bulletins Utrecht.

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Atz J.W. 1962. Does the shrimpfish swim head up or head down? Anim Kingdomq. 66:175–179.

Bannikov A.F. 2014. The systematic composition of the Eocene actinopterygian fish fauna from Monte Bolca, northern Italy, as known to date. Misc. Paleontol. 12:22–34.

Baraf L.M., Pratchett M.S., Cowman P.F. 2019. Ancestral biogeography and ecology of marine angelfishes (F: Pomacanthidae). Mol. Phylogenet. Evol. 140:106596.

Barber P.H., Bellwood D.R. 2005. Biodiversity hotspots: Evolutionary origins of biodiversity in wrasses (Halichoeres: Labridae) in the Indo-Pacific and new world tropics. Mol. Phylogenet. Evol. 35:235–253.

Batista R., Olsson U., Andermann T., Aleixo A., Ribas C.C., Antonelli A. 2020. Phylogenomics and biogeography of the world ' s thrushes (Aves , Turdus): new evidence for a more parsimonious evolutionary history. .

Bellwood D.R., Goatley C.H.R., Bellwood O. 2017. The evolution of fishes and corals on reefs: Form, function and interdependence. Biol. Rev.

Bellwood D.R., Wainwright P.C. 2002. The History and Biogeography of Fishes on Coral Reefs. Coral Reef Fishes. .

Berger B.A., Kriebel R., Spalink D., Sytsma K.J. 2016. Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. Mol. Phylogenet. Evol. 95:116–136.

Bergert B.A., Wainwright P.C. 1997. Morphology and kinematics of prey capture in the syngnathid fishes Hippocampus erectus and Syngnathus floridae. Mar. Biol. 127:563–570.

Berglund A., Rosenqvist G., Svensson I. 1986. Reversed sex roles and parental energy investment in zygotes of two pipefish (Syngnathidae) species. Mar. Ecol. Prog. Ser.

Betancur-R. R., Broughton R.E., Wiley E.O., Carpenter K., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton II J.C., Zhang F., Buser T., Campbell M., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Ortí G. 2013. The Tree of Life and a New Classification of Bony Fishes. PLoS Curr.:0–45.

Betancur-R. R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17.

Biomatters. 2019. Geneious Prime. .

Boehm J.T., Woodall L., Teske P.R., Lourie S.A., Baldwin C., Waldman J., Hickerson M. 2013. Marine dispersal and barriers drive Atlantic seahorse diversification. J. Biogeogr. 40:1839–1849.

Bowen B.W., Bass A.L., Rocha L.A., Grant W.S., Robertson D.R. 2001. Phylogeography of the trumpetfishes (aulostomus): ring species complex on a global scale. Evolution (N. Y). 55:1029.

Cantalice K., Alvarado-Ortega J. 2016. Eekaulostomus cuevasae gen. and sp. nov., an ancient armored trumpetfish (Aulostomoidea) from Danian (Paleocene) marine deposits of Belisario Domínguez, Chiapas, southeastern Mexico. Palaeontol. Electron.:1–24.

Carnevale G., Bannikov A.F. 2019. A dragonet (Teleostei, callionymoidei) from the eocene of monte bolca, italy. Boll. della Soc. Paleontol. Ital. 58:295–307.

Carnevale G., Bannikov A.F., Landini W., Sorbini C. 2006. Volhynian (early Sarmatian sensu lato) fishes from Tsurevsky, North Caucasus, Russia. 80:684–699.

Carnevale G., Bannikov A.F., Marramà G., Tyler J.C., Zorzin R. 2014. The Pesciara-Monte Postale Fossil-Lagerstätte: 2. Fishes and other vertebrates. Rend. della Soc. Paleontol. Ital. 4:37–63.

Casazza T.L., Ross S.W. 2008. Fishes associated with pelagic Sargassum and open water lacking Sargassum in the Gulf Stream off North Carolina. Fish. Bull. 106:348–363.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol.

Chen W.J., Bonillo C., Lecointre G. 2003. Repeatability of clades as a criterion of reliability: A case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. Mol. Phylogenet. Evol.

Cowman P.F., Bellwood D.R. 2011. Coral reefs as drivers of cladogenesis: Expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. J. Evol. Biol. 24:2543–2562.

Cowman P.F., Bellwood D.R. 2013a. The historical biogeography of coral reef fishes: Global patterns of origination and dispersal. J. Biogeogr. 40:209–224.

Cowman P.F., Bellwood D.R. 2013b. Vicariance across major marine biogeographic barriers: Temporal concordance and the relative intensity of hard versus soft barriers. Proc. R. Soc. B Biol. Sci. 280.

Cowman P.F., Bellwood D.R., van Herwerden L. 2009. Dating the evolutionary origins of wrasse lineages (Labridae) and the rise of trophic novelty on coral reefs. Mol. Phylogenet. Evol.

Cowman P.F., Parravicini V., Kulbicki M., Floeter S.R. 2017. The biogeography of tropical reef fishes: endemism and provinciality through time. Biol. Rev. 92:2112–2130.

Dong J., Kergoat G.J., Vicente N., Rahmadi C., Xu S., Robillard T. 2018. Biogeographic patterns and diversification dynamics of the genus Cardiodactylus Saussure (Orthoptera, Grylloidea, Eneopterinae) in Southeast Asia. Mol. Phylogenet. Evol.

Duarte-Ribeiro E., Davis A.M., Rivero-Vega R.A., Ortí G., Betancur R. 2018. Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. Proc. R. Soc. B Biol. Sci. 285.

Dupin J., Matzke N.J., Särkinen T., Knapp S., Olmstead R.G., Bohs L., Smith S.D. 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeogr. 44:887–899.

Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. 32:786–788.

Feng Y.-J., Blackburn D.C., Liang D., Hillis D.M., Wake D.B., Cannatella D.C., Zhang P. 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. Proc. Natl. Acad. Sci. 114:E5864–E5870.

Fish F.E., Holzman R. 2019. Swimming Turned on Its Head: Stability and Maneuverability of the Shrimpfish (Aeoliscus punctulatus). Integr. Org. Biol. 1.

Floeter S.R., Rocha L.A., Robertson D.R., Joyeux J.C., Smith-Vaniz W.F., Wirtz P., Edwards A.J., Barreiros J.P., Ferreira C.E.L., Gasparini J.L., Brito A., Falcón J.M., Bowen B.W., Bernardi G. 2008. Atlantic reef fish biogeography and evolution. J. Biogeogr. 35:22–47.

Fricke R., Eschmeyer W.N., Fong J.D. 2020. ESCHMEYER'S CATALOG OF FISHES. SPECIES BY FAMILY/SUBFAMILY. Available from http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp.

Friedman M., Carnevale G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. J. Geol. Soc. London.

Froese R., Pauly D. 2021. FishBase. Available from www.fishbase.org.

Gosline W.A. 1984. Structure, function, and ecology in the goatfishes (Family Mullidae). Pacific Sci.

Hamilton H., Saarman N., Short G., Sellas A.B., Moore B., Hoang T., Grace C.L., Gomon M., Crow K., Brian Simison W. 2017. Molecular phylogeny and patterns of diversification in syngnathid fishes. Mol. Phylogenet. Evol. 107:388–403.

Hou Z., Li S. 2017. Tethyan changes shaped aquatic diversification. Biol. Rev. 93:874–896.

Huelsenbeck J.P., Rannala B., Masly J.P., Huelsenbeck J.P., Rannala B., Masly J.P. 2000. Accommodating Phylogenetic Uncertainty in Evolutionary Studies Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/3075584 Linked references are available on JSTOR for this article : Accommodating Ph. 288:2349–2350.

Hughes L.C., Ortí G., Huang Y., Sun Y., Baldwin C.C., Thompson A.W., Arcila D., Betancur-R. R., Li C., Becker L., Bellora N., Zhao X., Li X., Wang M., Fang C., Xie B., Zhou Z., Huang H., Chen S., Venkatesh B., Shi Q. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. 115:6249–6254.

Hughes L.C., Ortí G., Saad H., Li C., White W.T., Baldwin C.C., Crandall K.A., Arcila D., Betancur-R. R. 2020. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. Mol. Ecol. Resour.

IUCN. 2019. The IUCN Red List of Threatened Species. Available from https://www.iucnredlist.org.

Johnson G.D. 1993. Percomorph phylogeny: progress and problems. Bull. Mar. Sci. 52:3–28.

Johnson G.D., Patterson C. 1993. Percomorph phylogeny: a survey of acanthomorphs and a new proposal. Bull. Mar. Sci.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol.

Kawahara R., Miya M., Mabuchi K., Lavoué S., Inoue J.G., Satoh T.P., Kawaguchi A., Nishida M. 2008. Interrelationships of the 11 gasterosteiform families (sticklebacks, pipefishes, and their relatives): A new perspective based on whole mitogenome sequences from 75 higher teleosts. Mol. Phylogenet. Evol. 46:224–236.

Kim B. 2002. Comparative anatomy and phylogeny of the family Mullidae (Teleostei: Perciformes). Mem. Grad. Sch. Fish. Sci. Hokkaido Univ.:1–75.

Klaus K. V., Matzke N.J. 2020. Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Syst. Biol. 69:61–75.

Kulbicki M., Parravicini V., Bellwood D.R., Arias-Gonzàlez E., Chabanet P., Floeter S.R., Friedlander A., McPherson J., Myers R.E., Vigliola L., Mouillot D. 2013. Global biogeography of reef fishes: A hierarchical quantitative delineation of regions. PLoS One. 8.

Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62:789–804.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Lessios H.A. 2008. The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. Annu. Rev. Ecol. Evol. Syst. 39:63–91.

Lessios H.A., Robertson D.R. 2006. Crossing the impassable: Genetic connections in 20 reef fishes across the eastern Pacific barrier. Proc. R. Soc. B Biol. Sci. 273:2201–2208.

Lessios H.A., Robertson D.R. 2013. Speciation on a round planet: Phylogeography of the goatfish genus Mulloidichthys. J. Biogeogr. 40:2373–2384.

Li C., Olave M., Hou Y., Qin G., Schneider R.F., Gao Z., Tu X., Wang X., Qi F., Nater A., Kautt A.F., Wan S., Zhang Y., Liu Y., Zhang H., Zhang B., Zhang H., Qu M., Liu S., Chen Z., Zhong J., Zhang H., Meng L., Wang K., Yin J., Huang L., Venkatesh B., Meyer A., Lu X. 2021. Genome sequences reveal global dispersal routes and convergent developmental mechanism in seahorse evolution. Nat. Commun.

Lin Q., Fan S., Zhang Y., Xu M., Zhang H., Yang Y., Lee A.P., Woltering J.M., Ravi V., Gunter H.M., Luo W., Gao Z., Lim Z.W., Qin G., Schneider R.F., Wang X., Xiong P., Li G., Wang K., Min J., Zhang C., Qiu Y., Bai J., He W., Bian C., Zhang X., Shan D., Qu H., Sun Y., Gao Q., Huang L., Shi Q., Meyer A., Venkatesh B. 2016. The seahorse genome and the evolution of its specialized morphology. Nature. 540:395–399.

Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48.

Luiz O.J., Madin J.S., Ross Robertson D., Rocha L.A., Wirtz P., Floeter S.R. 2012. Ecological traits influencing range expansion across large oceanic dispersal barriers: Insights from tropical Atlantic reef fishes. Proc. R. Soc. B Biol. Sci. 279:1033–1040.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Matzke N.J. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Packag. version 0.2.

Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

Matzke N.J. 2019. BioGeoBEARS - Run BioGeoBEARS on multiple trees. Available from https://github.com/nmatzke/BioGeoBEARS/blob/master/R/BioGeoBEARS_on_multiple_trees_v1.R.

McGee M.D., Faircloth B.C., Borstein S.R., Zheng J., Hulsey C.D., Wainwright P.C., Alfaro M.E. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. Proc. R. Soc. B Biol. Sci. 283:20151413.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics.

Miya M., Friedman M., Satoh T.P., Takeshima H., Sado T., Iwasaki W., Yamanoue Y., Nakatani M., Mabuchi K., Inoue J.G., Poulsen J.Y., Fukunaga T., Sato Y., Nishida M. 2013. Evolutionary Origin of the Scombridae (Tunas and Mackerels): Members of a Paleogene Adaptive Radiation with 14 Other Pelagic Fish Families. PLoS One. 8.

Montes C., Cardona A., Jaramillo C., Pardo A., Silva C., Valencia V., Ayala C., Pérez-Angel L., Rodriguez-Parra L., Ramirez V., Niño H. 2015. Middle Miocene closure of the Central American Seaway. Science (80-.). 348:226–229.

Müller R.D., Cannon J., Qin X., Watson R.J., Gurnis M., Williams S., Pfaffelmoser T., Seton M., Russell S.H.J., Zahirovic S. 2018. GPlates: Building a Virtual Earth Through Deep Time. Geochemistry, Geophys. Geosystems. 19:2243–2261.

Near T.J., Dornburg A., Eytan R.I., Keck B.P., Smith W.L., Kuhn K.L., Moore J.A., Price S.A., Burbrink F.T., Friedman M., Wainwright P.C. 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci.

Near T.J., Eytan R.I., Dornburg A., Kuhn K.L., Moore J.A., Davis M.P., Wainwright P.C., Friedman M., Smith W.L. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci.

Nelson G. 1989. Phylogeny of major fish groups. hierarchy life. Mol. Morphol. phylogenetic Anal. Proc. from Nobel Symp. 70. ICS824.

Nelson J.S., Grande T.C., Wilson M.V.H. 2016. Fishes of the World: Fifth Edition. .

Neutens C., Adriaens D., Christiaens J., De Kegel B., Dierick M., Boistel R., Van Hoorebeke L. 2014. Grasping convergent evolution in syngnathids: A unique tale of tails. J. Anat. 224:710–723.

O'Dea, Aaron, Lessios H.A., Coates A.G., Eytan R.I., Restrepo-Moreno S.A., Cione A.L., Collins L.S., de Queiroz A., Farris D.W., Norris R.D., Stallard R.F., Woodburne M.O., Aguilera O., Aubry M.-P., Berggren W.A., Budd A.F., Cozzuol M.A., Coppard S.E., Duque-Caro H., Finnegan S., Gasparini G.M., Grossman E.L., Johnson K.G., Keigwin L.D., Knowlton N., Leigh E.G., Leonard-Pingel J.S., Marko P.B., Pyenson N.D., Rachello-Dolmen P.G., Soibelzon E., Soibelzon L., Todd J.A., Vermeij G.J., Jackson J.B.C. 2016. Formation of the Isthmus of Panama. Sci. Adv. 2:1–12.

OBIS. 2021. Data from the Ocean Biogeographic Information System. Intergovernmental Oceanographic Commission of UNESCO. .

Pietsch T. 1978. Evolutionary Relationships of the Sea Moths (Teleostei : Pegasidae) with a Classification of Gasterosteiform Families Author (s): Theodore W . Pietsch Published by : American Society of Ichthyologists and Herpetologists (ASIH) Stable URL : https://w. Copeia.:517–529.

Poore G.C.B. 1995. State of the Marine Environment Report for Australia: The Marine Environment - Technical Annex: 1. Canberra, Australia.: .

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol.

Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45:741–749.

Ree R.H., Smith S.A. 2008a. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

Ree R.H., Smith S.A. 2008b. Lagrange: software for likelihood analysis of geographic range evolution. Syst. Biol.

dos Reis M., Yang Z. 2019. Bayesian molecular clock dating using genome-scale datasets. Methods in Molecular Biology. .

Renema W., Bellwood D.R., Braga J.C., Bromfield K., Hall R., Johnson K.G., Lunt P., Meyer C.P., McMonagle L.B., Morley R.J., O'Dea A., Todd J.A., Wesselingh F.P., Wilson M.E.J., Pandolfi J.M. 2008. Hopping hotspots: Global shifts in marine biodiversity. Science (80-.). 321:654–657.

Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol.

Richardson J.E., Chatrou L.W., Mols J.B., Erkens R.H.J., Pirie M.D. 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. Philos. Trans. R. Soc. B Biol. Sci. 359:1495–1508.

Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero A., Walker Jr. H.J., Carpenter K.E., Sheaves M., Ortí G., Arcila D., Betancur-R. R. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. Proc. Natl. Acad. Sci. U. S. A.

Rocha L.A., Robertson D.R., Rocha C.R., Van Tassell J.L., Craig M.T., Bowen B.W. 2005. Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. Mol. Ecol. 14:3921–3928.

Rögl F. 1999. Palaeogeographic considerations for Mediterranean and Paratethys seaways. Ann. Naturhist. Mus. Wien. 99A:279–310.

Rögl V.F. 1998. Palaeogeographic Considerations for Mediterranean and Paratethys Seaways (Oligocene to Miocene) Author (s): Fred Rögl Source : Annalen des Naturhistorischen Museums in Wien . Serie A für Mineralogie und Petrographie , Geologie und Paläontologie , Anth. Ann. Naturhist. Mus. Wien. 99:279–310.

Rokas A., Carroll S.B. 2006. Bushes in the tree of life. PLoS Biol. 4:1899–1904.

Ronquist F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

Sanciangco M.D., Carpenter K.E., Betancur-R R. 2016. Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). Mol. Phylogenet. Evol.

Schlüter M., Steuber T., Parente M. 2008. Chronostratigraphy of Campanian-Maastrichtian platform carbonates and rudist associations of Salento (Apulia, Italy). Cretac. Res. 29:100–114.

Shepherd S., Edgar G. (Eds. . 2013. Ecology of Australian Temperate Reefs: The Unique South. CSIRO Publishing, Melbourne, Australia: .

Siqueira A.C., Bellwood D.R., Cowman P.F. 2019. Historical biogeography of herbivorous coral reef fishes: The formation of an Atlantic fauna. J. Biogeogr. 46:1611–1624.

Siqueira A.C., Morais R.A., Bellwood D.R., Cowman P.F. 2020. Trophic innovations fuel reef fish diversification. Nat. Commun. 11:1–11.

Song H.Y., Mabuchi K., Satoh T.P., Moore J.A., Yamanoue Y., Miya M., Nishida M. 2014. Mitogenomic circumscription of a novel percomorph fish clade mainly comprising "Syngnathoidei" (Teleostei). Gene. 542:146–155.

Sorbini L. 1981. The Cretaceous fishes of Nardò. I°. Order Gasterosteiformes (Pisces). Boll Mus. Civ. di Stor. Nat. di Verona. 8:1–27.

Spalding M.D., Fox H.E., Allen G.R., Davidson N., Ferdaña Z.A., Finlayson M., Halpern B.S., Jorge M.A., Lombana A., Lourie S.A., Martin K.D., McManus E., Molnar J., Recchia C.A., Robertson J. 2007. Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. Bioscience. 57:573–583.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.

Steininger F., Rögl F. 1979. The paratethys history. A contribution towards the Neogene geodynamics of the alpine orogene. Ann. Geol. des Pays Hell. 3:1153–1165.

Tagliacollo V.A., Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. Mol. Biol. Evol. 35:1798–1811.

Tea Y.K., Van Der Wal C., Ludt W.B., Gill A.C., Lo N., Ho S.Y.W. 2019. Boomeranging around Australia: Historical biogeography and population genomics of the anti-equatorial fish Microcanthus strigatus (Teleostei: Microcanthidae). Mol. Ecol. 28:3771–3785.

Teske P.R., Cherry M.I., Matthee C.A. 2004. The evolutionary history of seahorses (Syngnathidae: Hippocampus): Molecular data suggest a West Pacific origin and two invasions of the Atlantic Ocean. Mol. Phylogenet. Evol. 30:273–286.

Teske P.R., Hamilton H., Matthee C.A., Barker N.P. 2007. Signatures of seaway closures and founder dispersal in the phylogeny of a circumglobally distributed seahorse lineage. BMC Evol. Biol. 7.

Teske P.R., Hamilton H., Palsbøll P.J., Choo C.K., Gabr H., Lourie S.A., Santos M., Sreepada A., Cherry M.I., Matthee C.A. 2005. Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. Mar. Ecol. Prog. Ser. 286:249–260.

Thacker C.E. 2015. Biogeography of goby lineages (Gobiiformes: Gobioidei): Origin, invasions and extinction throughout the Cenozoic. J. Biogeogr. 42:1615–1625.

Thacker C.E. 2017. Patterns of divergence in fish species separated by the Isthmus of Panama. BMC Evol. Biol. 17:1–14.

Townsend J.P., Su Z., Tekle Y.I. 2012. Phylogenetic signal and noise: Predicting the power of a data set to resolve phylogeny. Syst. Biol.

Varela L., Tambusso P.S., Mcdonald H.G., Fariña R.A. 2019. Phylogeny, Macroevolutionary Trends and Historical Biogeography of Sloths: Insights From a Bayesian Morphological Clock Analysis. Syst. Biol. 68:204–218.

Vargas O.M., Dick C.W. 2020. Diversification History of Neotropical Lecythidaceae, an Ecologically Dominant Tree Family of Amazon Rain Forest. .

Whitfield A.K. 1999. Ichthyofaunal assemblages in estuaries: A South African case study. Rev. Fish Biol. Fish.

Wiley E., Johnson G. 2010. A teleost classification based on monophyletic groups. In: Nelson JS, Schultze HP, MVH W, editors. Orig. Phylogenetic ., Interrelat. Teleosts. (Munich Verlag Dr. Friedrich Pfeil).:123–182.

Wilson A.B., Orr J.W. 2011. The evolutionary origins of Syngnathidae: Pipefishes and seahorses. J. Fish Biol. 78:1603–1623.

Woodall L. 2009. Population genetics and mating systems of the European seahorses Hippocampus guttulatus and Hippocampus hippocampus. .

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol.

York P.H., Booth D.J., Glasby T.M., Pease B.C. 2006. Fish assemblages in habitats dominated by Caulerpa taxifolia and native seagrasses in south-eastern Australia. Mar. Ecol. Prog. Ser.

Žalohar J., Hitij T. 2012. The first known fossil record of pygmy pipehorses (Teleostei: Syngnathidae: Hippocampinae) from the Miocene Coprolitic Horizon, Tunjice Hills, Slovenia. Ann. Paleontol.

Žalohar J., Hitij T., Križnar M. 2009. Two new species of seahorses (Syngnathidae, Hippocampus) from the Middle Miocene (Sarmatian) Coprolitic Horizon in Tunjice Hills, Slovenia: The oldest fossil record of seahorses. Ann. Paleontol.

Chapter 2

Colonization dynamics explain diversity patterns of syngnatharian fishes across marine realms while morphological similarities persist among them

Aintzane Santaquiteria, Elizabeth Christina Miller, Ulises Rosas-Puchuri, Carmen del R. Pedraza-Marrón, Emily M. Troyer, Mark W. Westneat, Giorgio Carnevale, Dahiana Arcila, and Ricardo Betancur-R.

2.1 Abstract

In extant marine fish groups, a clear longitudinal gradient in species richness across oceans is observed, with the Indo-Pacific exhibiting the highest levels of diversity. Three non-mutually exclusive evolutionary hypotheses have been proposed to explain this diversity gradient: time-for-speciation, center of accumulation, and *in situ* diversification rates. Using syngnatharians (seahorses, dragonets, goatfishes, and relatives) as a study system, we tested these hypotheses and additionally assessed whether patterns of morphological diversity are congruent with species richness patterns. We used well-sampled phylogenies and a suite of phylogenetic comparative methods that account for various sources of uncertainty to estimate rates of lineage diversification and morphological disparity within all three major oceanic realms (Indo-Pacific, Atlantic, and eastern Pacific), as well as within the Indo-Pacific region. We find similar diversification rates across regions, indicating that increased syngnatharian diversity in the Indo-Pacific is due to earlier colonizations from the Tethys Sea followed by in situ speciation, and more frequent colonization during the Miocene, coinciding with the formation of coral reefs. These results support both time-for-speciation and center of accumulation hypotheses. Analyses also show that body plan disparity arose early in syngnatharian history and morphological rates likewise do not vary across areas. Overall, high species richness is not necessarily coupled with high morphological disparity in different biogeographic settings.

2.2 Introduction

Understanding the evolutionary processes that shape species distributions in marine and terrestrial faunas has long been a focus of attention for evolutionary biologists, biogeographers, and ecologists (Darwin 1859; Wallace 1869). Three processes directly change the species richness of a region: speciation, extinction, and dispersal (Ricklefs 1987). Therefore, all diversity gradients (e.g., latitudinal, longitudinal, elevation, depth) are formed by variation in the rates and timing of these three processes (e.g., Wiens 2015; Vasconcelos et al. 2020; Miller et al. 2022). The mechanisms driving this variation are ultimately a combination of biotic and abiotic factors, such as climatic stability, geographical barriers, productivity, competition, and predation (Mittelbach et al. 2007; Edgar et al. 2017).

Multiple non-mutually exclusive hypotheses have been proposed to explain species richness patterns from an evolutionary process perspective (Jablonski et al. 2006; Bowen et al. 2013; Cowman 2014; Gaboriau et al. 2019). The time-for-speciation hypothesis (Stephens and Wiens 2003), sometimes referred to as the museum hypothesis, suggests that earlier colonization of regions provides lineages with more time to diversify, ultimately leading to higher species richness. The center of accumulation hypothesis (Ladd 1960; Kool et al. 2011) posits that higher diversity of a region is due to preferential colonization. A highly diverse region may act as a source, with species dispersing to new regions or serve as sink, receiving species from other area (Jablonski et al. 2006). The center of origin or *in situ* diversification rate hypothesis (Briggs 1974; Rohde 1992; Huang et al. 2018), sometimes referred to as the cradle, posits that regions with higher rates of speciation will exhibit greater species richness. Other hypotheses that remain largely untested include, for example, the center of survival, which suggests that the stability of an area may enable persistence, leading to increased species diversity (Barber and Bellwood 2005). Recent macroevolutionary studies focusing on diversity gradients have emphasized time-for-speciation as a prominent explanatory factor in various taxa, including freshwater and marine fishes, terrestrial turtles, and plants (Stephens and Wiens 2003; Cowling et al. 2017; Miller et al. 2018; García-andrade et al. 2023). Some studies have also observed lower diversification rates in regions with high species richness, as seen for marine fishes along a latitudinal diversity gradient (Rabosky et al. 2018), although the reverse is true for amphibians, mammals, and certain groups of reef fishes that have higher speciation rates in the tropics (Pyron and Wiens 2013; Rolland et al. 2014; Siqueira et al. 2016).

Species richness across the oceans exhibits a clear longitudinal gradient, with coastal marine groups being more diverse in the western Indo-Pacific (IP) compared to the eastern Pacific (EP) and the Atlantic (Atl.) oceans (Tittensor et al. 2010; Parravicini et al. 2013; Edgar et al. 2017). This pattern is particularly pronounced in the Central Indo-Pacific (CIP), which includes the Indo-Australian Archipelago (IAA) diversity hotspot (Renema et al. 2008). The four hypotheses above have individually been proposed to explain the processes driving the high biodiversity of the CIP, although it is now believed that present-day reef fish distributions are the result of a combination of accumulation, survival, and *in situ* speciation (Bowen et al. 2013; Cowman and Bellwood 2013; Bellwood et al. 2015). A study examining three diverse reef fish families (wrasses, butterflyfishes, and damselfishes; Cowman and Bellwood 2013), suggested that species accumulation (specifically colonization of the CIP from the Tethys Sea) dates back to the Palaeocene/Eocene (~66–34 Ma), followed by lineage survival from the Eocene/Oligocene (~34–23 Ma). During the Oligocene/Miocene (~23–5 Ma), geological processes (e.g., tectonic plate collisions) and climatic events (e.g., sea-level changes and temperature fluctuations) transformed the CIP into a more complex region characterized by the emergence of new islands, shallow seas, and coral reefs (Renema et al. 2008; Williams and Duda 2008; Lohman et al. 2011). These changes facilitated *in situ* lineage diversification from the Miocene onwards, followed by emigration in the Pliocene (Cowman and Bellwood 2013) and coral reef refugia for fishes in the Quaternary (Pellissier et al. 2014). A recent global-scale study found that species richness in the Coral Triangle is primarily influenced by the time-for-speciation hypothesis, supported by early lineage colonization and consistent diversification rates across the world's oceans (Miller et al. 2018).

Despite extensive research into the factors contributing to the heterogeneity of species richness across oceans, it is important to determine whether this spatial variation also correlates with differences in morphological and functional diversity (Crouch and Jablonski 2023; Diamond and Roy 2023). In regions with high species richness, species can either cluster densely in morphospace (Pigot et al. 2016; Pellissier et al. 2018) or evolve distinct morphologies which potentially enable niche partitioning and coexistence (Rosamond et al. 2020). Regions can also have functional redundancy (i.e., multiple species that perform similar functional roles), promoting resilience and adaptability and allowing the community to withstand environmental changes, disturbances, and species loss (Bellwood et al. 2006; Carturan et al. 2022). Researchers have examined morphological diversity in reef fishes from a biogeographic perspective, investigating various ecological and morphological traits like diet, habitat, body size, and body shape (Mouillot et al. 2014; Siqueira et al. 2019; Mclean et al. 2021; Diamond and Roy 2023). To date, global-scale studies on reef-associated fishes have identified trait similarities between marine realms, suggesting shared functional roles across biogeographic communities (Mouillot et al. 2014; Mclean et al. 2021). However, within individual reef fish groups (e.g., surgeonfishes, rabbitfishes, and parrotfishes), greater disparity is observed in the Indo-Pacific compared to the Atlantic (Siqueira et al. 2019), possibly due to the highly complex and structurally diverse coral formations providing more available niches in the former.

The percomorph clade Syngnatharia (669 species), which includes seahorses, pipefishes, flying gurnards, goatfishes, dragonets, and sea moths, offers an excellent opportunity to investigate the extent to which species richness and morphological patterns found on a global scale (Miller et al. 2018; Mclean et al. 2021) are also reflected within a single marine clade. This group originated in the ancient Tethys Sea (now Europe) during the Late Cretaceous (Santaquiteria et al. 2021; Stiller et al. 2022). Since then, its lineages have colonized tropical and temperate biogeographic regions across the globe and diversified in various habitats, including seagrass beds, coral and rocky reefs, and mangrove forests (Froese and Pauly 2021; Santaquiteria et al. 2021; Stiller et al. 2022). The Indo-Pacific region has the highest species richness (~535 species) compared to other oceanic realms (Atlantic: ~94 species and eastern Pacific: ~29 species). Syngnatharians also exhibit a great diversity of body plans, including elongated snouts and bodies, prehensile tails, and hyoid barbels (Neutens et al. 2014; Nash et al. 2022).

Here, we set out to examine diversification patterns and body shape morphometrics broadly across the Syngnatharia, in association with biogeographic regions. We aim to understand the evolutionary processes influencing the longitudinal diversity gradient. Specifically, we explore whether the higher species richness in the Indo-Pacific is explained by greater time-for-speciation, center of accumulation (more colonization events), and/or higher *in situ* diversification rates. The observed unevenness in species richness across oceanic basins in Syngnatharia may be ascribed to earlier colonization of the Indo-Pacific, allowing more time to diversify (Miller et al. 2018). However, if diversification rates prove to be higher in the Indo-Pacific, notably within the CIP, this could likely be attributed to the intricate geological history that has bestowed the IAA with extensive shallow-water areas that potentially acted as a diversification cradle (Bellwood and Hughes 2001; Siqueira et al. 2020). Moreover, if the rate of dispersal into the CIP surpasses that of other regions, given its elevated oceanic connectivity with adjacent regions, the CIP would assume the role of a center of accumulation (Bellwood and Hughes 2001) and potentially also a diversity source for other regions. High species richness in a region is not necessarily accompanied by high morphological or functional disparity. Syngnatharian assemblages within each ocean might have reached similar levels of body shape diversity across the globe in response to similar environmental conditions and ecological pressures, rather than distinct evolutionary trajectories within these basins (Mouillot et al. 2014; Mclean et al. 2021). However, if the greatest morphological disparity is found in the CIP, this could be attributed to increased niche availability, enabling lineages to occupy a broader morphological space (Renema et al. 2008; Parravicini et al. 2013; Sanciangco et al. 2013; Siqueira et al. 2019, 2021). To test these hypotheses, we first expanded on prior phylogenomic analyses of Syngnatharia (Santaquiteria et al. 2021) to include a total of 323 species (~50% of the species diversity). We then conducted a suite of comparative and biogeographic analyses to comprehensively examine the relationship between geographic distribution and lineage diversification and morphological evolution in this clade, while carefully considering topological, divergence time, and other methodological uncertainties.

2.3 Materials and Methods

See Appendix B, Supplementary Materials and Methods section for additional methodological details.

Taxonomic sampling, phylogenetic inference, and tree uncertainty

The phylogenetic framework for our comparative analyses builds upon two previous studies that examined the evolutionary and biogeographic history of syngnatharians based on ultraconserved elements (UCEs) data (932 loci) sequenced from 163 species (Longo et al. 2017; Santaquiteria et al. 2021). To account for tree uncertainty in downstream comparative analyses (see Santaquiteria et al. 2021) we estimated 56 backbone time trees, 28 dated using MCMCTree (dos Reis and Yang 2019) and 28 using RelTime, using as input topologies from the alternative UCE matrices analyzed with both maximum likelihood and coalescent approaches (14 RAxML and 14 ASTRAL-III trees; Stamatakis 2014; Mirarab and Warnow 2015). To improve taxonomic representation in our analyses, we expanded the sampling on the backbone trees by incorporating 160 additional syngnatharian taxa through publicly available mitochondrial sequences. This expansion covered approximately 50% of the extant diversity within the group, accounting for 323 of 669 species (see Appendix 1 available on Dryad). We rigorously vetted these sequences through a series of quality control steps and considered discrepancies in phylogenetic placement compared to a recent phylogenomic study (Stiller et al. 2022). For time calibration, we employed a combination of 13 primary calibrations, including 4 new calibrations on the backbone trees using MCMCTree and RelTime, and multiple secondary calibrations on the expanded trees using *congruification* in conjunction with treePL (Smith and O'Meara 2012).

Biogeographic history and timing of regional colonization

We estimated ancestral ranges for the new set of syngnatharian phylogenies with 323 species using the R package *BioGeoBEARS* (Matzke 2013) following the approach outlined in Santaquiteria *et al.* (2021). We built a presence/absence matrix by coding each extant species according to their geographic ranges based on a 7-area biogeographic scheme (Spalding et al. 2007; Kulbicki et al. 2013; see Fig. 2.1 for areas and abbreviations). We also used paleogeographic domain information as biogeographic constraints based on data obtained from the 10 fossils (mostly from the Tethys Sea) used to calibrate our trees. We evaluated 12 biogeographic models using three time slices (92–12, 12–2.8, and 2.8–0 Ma) and a connectivity matrix based on dispersal probabilities between regions. As the summary phylogeny, we used the "master tree" (expanded tree inferred using the 932-UCE backbone tree and dated in RelTime). Based on the sensitivity of biogeographic inferences previously identified for the group, we accounted for both topological uncertainty and the inclusion/exclusion of the *j* parameter in biogeographic models (Ree and Sanmartín 2018; Matzke 2022). We summarized ancestral range estimates from all 28 RelTime trees by overlying average probabilities across compatible nodes on the "master tree" (Matzke 2019) using the best-fit model with and without the *j* parameter.

We assessed the center of accumulation and time-for-speciation hypotheses by estimating the frequency and timing of colonization events between the seven areas. To accomplish this, we conducted biogeographic stochastic mapping (BSM) analyses by simulating 100 stochastic histories on the "master tree" based on the best-fit biogeographic models (Dupin et al. 2017). From each map, we extracted all states at every node and tip and identified all individual colonizations, their descendants, and colonization timing for each given region. For each region, we then calculated the number of cumulative lineages (due to a combination of colonization and speciation), number of independent colonization events, immigration and emigration rates, speciation rates, and extirpation rates across time by averaging over 100 histories. This approach follows the biogeographic methodology developed by Xing and Ree (2017) and implemented previously (e.g., Ding et al. 2020; Miller et al. 2022). We repeated all biogeographic analyses using the "alternative tree" (expanded tree inferred using the 932-UCE backbone tree and dated in MCMCTree) and all 28 MCMCTree trees.

Diversification rates among regions

To assess the influence of geographic distribution on lineage diversification dynamics (testing the *in situ* diversification rates hypothesis), we estimated diversification rates based on the 56 calibrated phylogenies, each with 323 tips. We also assessed the sensitivity of diversification rate analyses to 12 terminal nodes with shallow divergences (i.e., 'T-like' terminal nodes with branch lengths <0.5 Ma) which may indicate taxonomic over-splitting and can "force" models to fit extremely fast rates. For geographic-dependent analyses, we fitted 24 different area-independent and area dependent models in GeoHiSSE (Caetano et al. 2018), both with and without the *j* parameter (Appendix B Table S4). Because GeoHiSSE only allows comparisons of two regions at a time, we conducted multiple pairwise comparisons between the focal and the remaining areas: IP vs. EP+Atl., EP vs. IP+Atl., and Atl. vs. IP+EP. We then calculated the sampling fractions for each region across all comparisons (IP: 47.85%, EP: 65.52%, and Atl.: 65.96%). We calculated the AIC values for each of the models and averaged the best three models (~90–95% accumulative weight) using Akaike weights. To estimate diversification rates in BAMM, for each tree we estimated prior parameters for time-variable speciation and extinction models using the R package *BAMMTools* (Rabosky et al. 2014). After running BAMM independently for each tree, we combined all results by calculating the mean diversification rate for each tip. Finally, we also estimated tip rates using "DR statistics" function for each tree (Jetz et al. 2012).

For each analysis, we compared tip-associated lineage diversification rates between (i) all oceanic realms, and (ii) all subareas within the Indo-Pacific. We conducted these comparisons using trees that included all syngnatharian species, as well as trees within four separate suborder-level clades: Syngnathoidei, Callionymoidei, Mulloidei, and Dactylopteroidei, to account for phylogenetic scale (Clarke 2021; Miller et al. 2021) (Fig. 2.1). Residual errors from lineage diversification rates were not normally distributed (even after log transformation), therefore, to assess statistical significance of rate differences among groups, we implemented a novel phylogenetically-corrected, non-parametric Kruskal-Wallis test (see Appendix B, Supplementary Materials and Methods).

Morphological disparity and rates by region

To assess phenotypic disparity across biogeographic regions, we used 2D geometric morphometric analyses to examine morphospace occupation and evolutionary rates within major oceanic realms and within subareas of the Indo-Pacific. Landmarks and semi-landmarks were placed using photographs from 474 specimens in 171 species sourced from museum collections (e.g., Smithsonian) and online repositories (Bray and Gomon 2021; Froese and Pauly 2021). In order to accommodate seahorse, pigmy pipehorse, and sea dragon specimens with bent body structures, we created two alternative schemes: a head-only set (including these specimens) and a full-body set (excluding them; Fig. 2.4 and Appendix B Fig. S1). We used the R package *geomorph* (Adams et al. 2021) to summarize variation in syngnatharian morphology using Procrustes superimposition. Using all 56 trees as input, we then conducted a principal component analysis (PCA) as well as a phylogenetically-corrected PCA (pPCA) using *phytools* (Revell 2012). For downstream morphological analyses, we selected PCs and pPCs explaining 95% of the variation (1–4 axes for head-only and 1–6 for full-body shape; see Results).

To examine contemporary trait disparity in syngnatharians, we used the R package *dispRity* (Guillerme 2018). Using "dispRity.per.group" function, we calculated the sum of variances for each biogeographical region based on both landmark schemes using PC and pPC scores as input.

We also analyzed disparity-through-time in syngnatharians using pPC scores obtained from our "master tree" and "alternative tree", the sum of variance metric, and the "dtt.dispRity" function in *dispRity.* We then evaluated the fit of eight morphological evolutionary models using the R package *mvMORPH* (Clavel et al. 2015). To quantify the morphospace occupancy of syngnatharians in each of the major biogeographic regions, we calculated the overlap (Jaccard and Sørenson statistics) between each region using the R package *hypervolume* (Blonder et al. 2018). We created four-dimensional hypervolumes for the head-only scheme and six-dimensional hypervolumes for the full-body scheme. Finally, to assess morphological evolutionary rates within areas, we estimated rates for each syngnatharian lineage across regions in BAMM, with the caveat that each PC/pPC needed to be analyzed separately (Uyeda et al. 2015). We ran BAMM independently for all 56 trees. We then combined the MCMC results from all selected PCs and pPCs and calculated mean rates for each lineage across the 28 RelTime trees and the 28 MCMCTree trees. Lastly, we compared the statistical significance of morphological rates between realms and Indo-Pacific regions using the modified Kruskal-Wallis test as explained for diversification rates (see above).

2.4 Results

See Appendix B, Supplementary Results for extended results.

Biogeographic history and timing of regional colonization

The best-supported biogeographic model in our ancestral range reconstruction analyses was BAYAREA+*j* (Appendix B Table S5). In light of recent criticisms on the implementation of the jump-dispersal parameter (*j*), we conducted the biogeographic analyses and interpreted the results with and without this parameter. Ancestral range reconstructions are similar to those obtained in previous studies for Syngnatharia (Fig. 2.1A; Santaquiteria et al. 2021; Stiller et al. 2022; see also Appendix B Fig. S37). The ancestral syngnatharian lineage originated in the Tethys Sea in the Late Cretaceous around 87 Ma. Syngnathids (seahorses and pipefishes) subsequently dispersed eastwards into the IP around 52 Ma, right after the Paleocene-Eocene Thermal Maximum (PETM). The remaining families dispersed into the IP more recently, during the Miocene and the IAA rearrangement (20 Ma onwards). Finally, all major families except pegasids (sea moths) colonized the Atl. and the EP multiple times via alternative routes (Fig. 2.1A and Appendix B Fig. S6; see also (Santaquiteria et al. 2021). The BAYAREA results without the *j* parameter show similar patterns but with more widespread ancestral distributions (Appendix B Fig. S7). Finally, results obtained using MCMCTree trees are similar, except that ancestral colonizations are inferred to have occurred earlier (Appendix B Figs. S8 and S9).

Based on analyses of biogeographic stochastic histories, we find that over 50% of syngnatharian lineages have similar dispersal rates from the Tethys Sea into the WIO, CIP, and WA regions (0.037–0.045), with the highest dispersal rate observe into CIP (0.045; Fig. 2.2A). Lineages began colonizing the IP more frequently around 23 million years ago, coinciding with the diversification of corals; however during this period, tectonic activity also played an important role in the formation of coastal habitats (Fig. 2.2B; Bellwood et al. 2017). More than 60% (~250 species) of lineages dispersed mostly out of CIP into different subareas within the IP. These events mostly occurred around when the IAA rearrangement concluded (~5 Ma; Lohman et al. 2011). Fewer dispersal events (~40% lineages) happened outside the remaining (non-CIP) areas and mostly occurred after the IAA (Fig. 2.2A). These findings are supported by the cumulative lineage plot over time (Fig. 2.1B), which shows that syngnatharian lineages have primarily accrued throughout their history in the IP, particularly within the CIP. Notably, lineages started dispersing to the CIP more frequently after the Tethys Seaway closure (Fig. 2.1B). We also observed that syngnatharian lineages have tended to disperse at high rates among adjacent areas (Fig. 2.2A). Dispersal rates are higher within IP subareas (0.028 to 0.046) than between oceanic realms (0.005 to 0.015). Within the IP, WIO is the region that receives the highest number of colonizations (22 independent lineages), followed closely by TA (20), while colonizations into CIP and CP are roughly half of that (11 and 10, respectively). Colonization events over time into the Atl. and EP tend to be more recent, most concentrated around the closure of the Panama Isthmus (Fig. 2.2B). Fewer lineages have colonized these two realms compared to the IP (Fig. 2.2). Dispersal rates are highest from EA into WA (0.057), followed by from EP into WA (0.050), and from WA into EP (0.030) and EA (0.043). Dispersal rates from these three areas into IP are overall lower, ranging from 0.016 to 0.037 (Fig. 2.2A).

Speciation rates per area over time tend to correlate with major geological and climatic events (Fig. 2.1C). We observe three major peaks: one marks the early Tethyan origin of all main (suborder level) lineages in Syngnatharia (~80 Ma), a second peak within the IP after the PETM (~56 Ma), and a third peak in all areas, mostly after the beginning of the IAA rearrangement (~35 Ma). Thereafter, speciation rates had similar dynamics within each ocean showing a decrease towards the present, particularly after the closure of the Isthmus of Panama (2.8 Ma). Finally, extirpation rates across regions are rarely constant over time, although the WA shows higher rates until ~60 Ma (Appendix B Fig. S13). When comparing all these results with those obtained using the BAYAREA model without the *j* parameter, we find similar patterns (Appendix B Fig. S14), although colonization rates tend to be higher as a result of including jump dispersal in the model (Appendix B Fig. S10). Analyses conducted on the “alternative tree” produced similar results to those obtained from the “master tree” (Appendix B Figs. S11 and S12), except the sequence of events began slightly earlier in MCMCTree analyses (Appendix B Figs. S15 and S16).

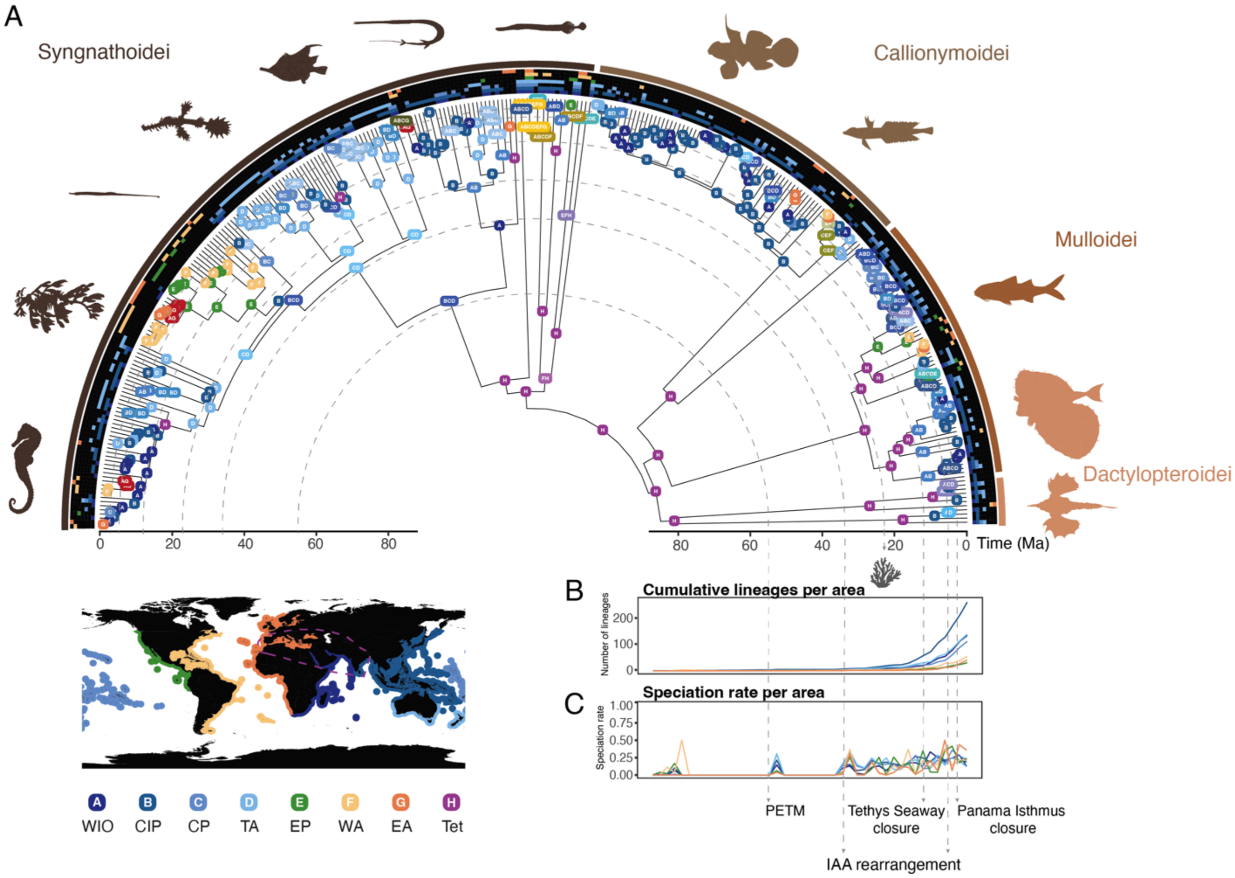
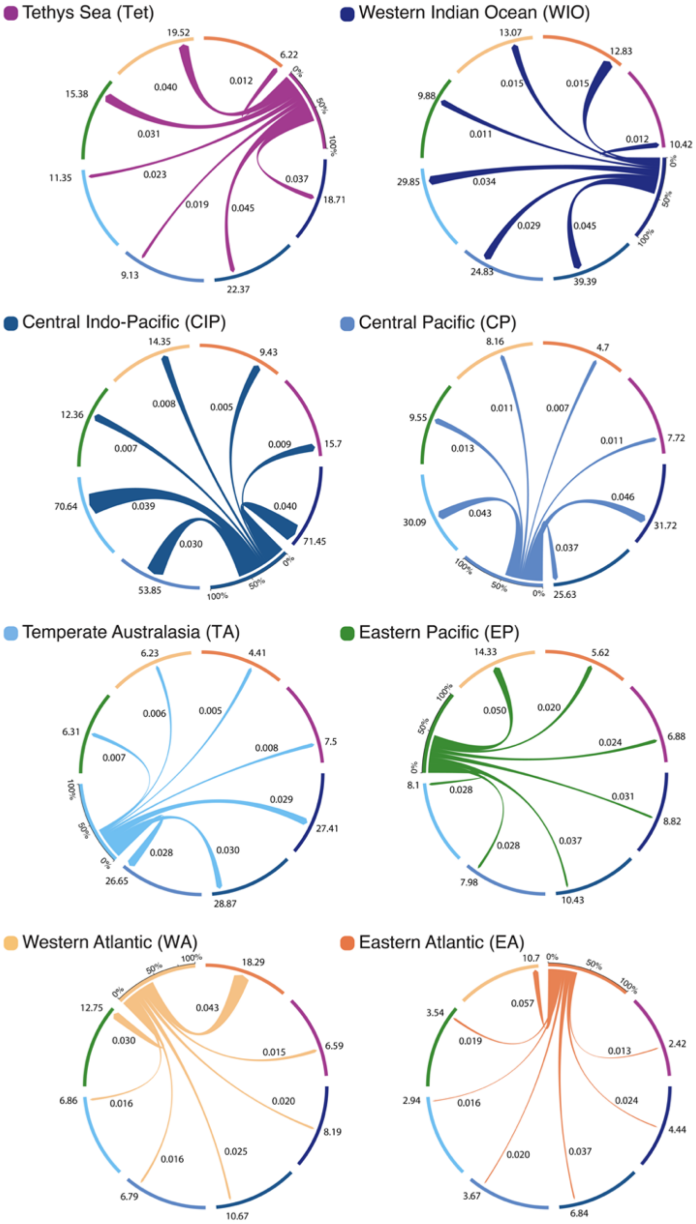


Figure 2.1 A) Syngnatharian phylogeny, biogeographic history, B) regional speciation rate and C) lineages through time based on ancestral range inferences and biogeographic stochastic mapping. CIP has accrued the largest number of lineages, experiencing a steep increase after the closure of the Tethys Seaway. Speciation rates are synchronous across areas, with two notable increases, one following the PETM and another after the IAA rearrangement. Major geological events are depicted with dotted lines: PETM (Paleocene-Eocene Thermal Maximum; ~56 Ma), IAA (Indo-Australian Archipelago; 33.9–5.3 Ma) rearrangement (associated with the expansion of modern coral reef formations), Tethys Seaway closure (12 Ma), and the closure of the Isthmus of Panama (2.8 Ma). Tet: Tethys Sea; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: eastern Pacific; WA: Western Atlantic; and EA: Eastern Atlantic. For complementary analyses see Appendix B Figs. S5–S9 and S13–S16.

A Dispersal events per area



B Colonization events per area over time

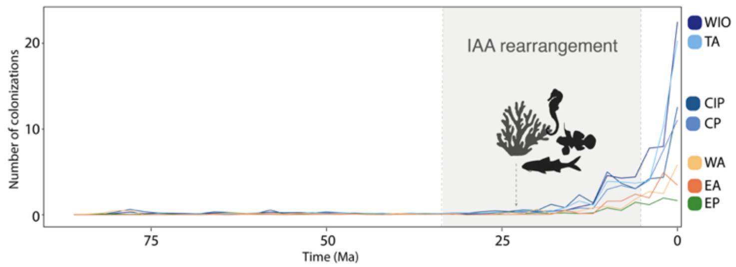


Figure 2.2. Tempo and mode of dispersal and colonization events between different biogeographic regions. The Indo-Pacific acts as a center of accumulation of syngnatharian lineages, a process that correlates with the IAA rearrangement and expansion of modern coral reef formations. CIP acts as a source of lineages while the Atlantic (EA and WA) acts as a sink. **A)** Chord diagrams for dispersal events outside each region; line width represents the percentage of lineages dispersing from a focal area to the rest of the areas. Number of lineages dispersed, and dispersal rates are depicted outside and inside the diagrams, respectively. **B)** Number of colonizations per area over time. For complementary analyses see Appendix B Figs. S10–S16.

Diversification rates among regions

After accounting for topological and divergence time uncertainties, lineage diversification analyses show no major differences whether using all taxa (323 tips) or after excluding potential instances of taxonomic over-splitting (311 tips; see Appendix B Fig. S17). Thus, here we report the results obtained using the complete taxonomic dataset. The CIP, particularly in New Caledonia, Philippines, Taiwan, and Japan, is the region with the highest species richness, while the Atl. and EP have the lowest (Fig. 2.3C). BAMM and DR analyses show no significant differences in lineage diversification rates across the three major realms, nor within the IP subareas ($p>0.05$; Fig. 2.3A). However, GeoHiSSE±*j* estimates using model averaging identified the highest rates in the EP (median 0.12), followed by the IP (0.008), and the lowest in the Atl. (0.004; Fig. 2.3A). Species with widespread distributions tend to have higher diversification rates based on GeoHiSSE±*j* (Fig. 2.3A and Appendix B Fig. S18). Clades that undergo burst of speciation also exhibit higher rates in GeoHiSSE±*j* and BAMM. Specifically, this pattern is observed in *Syngnathus* lineages within the EP (e.g., the most recent common ancestor [MRCA] of *S. auliscus* and *S. californiensis*) and the Atl. (e.g., MRCA of *S. schlegeli* and *S. pelagicus*) regions, consistent with findings reported by Stiller *et al.* (2022). Additionally, similar trends are observed in *Callionymus* species within the IP (e.g., MRCA of *C. valenciennei* and *C. planus*). Using tip-associated rates calculated from GeoHiSSE, we find significant differences between Atl. and IP (p=7.6e^{-6} and 9.50e^{-8}, with or without *j*) and between Atl. and EP (p=0.003 +*j* and 0.047 -*j*), whereas all analyses show constant diversification rates within IP subareas (median ~0.008; Fig. 2.3A and Appendix B Table S6). At the suborder level, Syngnathoidei and Mulloidei also show major rate differences between Atl. and IP and between Atl. and EP in GeoHiSSE±*j* analyses (Appendix B Fig. S18 and Table S6). Other suborders, however, show rate constancy across geographies regardless of the method used (Appendix B Fig. S18). When comparing these results with those obtained using MCMCTree trees, analyses based on all species also show no significant differences among biogeographic regions (Appendix B Fig. S19 and Table S6). However, we find significant differences in the suborder Callionymoidei between the Atl. and EP, Atl. and IP, as well as within the IP subareas (between CIP and CP in GeoHiSSE+*j* and CIP and TA in GeoHiSSE-*j*) that were not apparent when using RelTime trees (Appendix B Table S6). We also find significance in Mulloidei between the Atl. and IP across all the methods except for DR (Appendix B Fig. S20 and Table S6). Results of DR analyses after accounting for topological incongruences with respect to the Stiller *et al.* (Stiller et al. 2022) phylogeny are also similar (Appendix B Fig. S38 and Table S6).

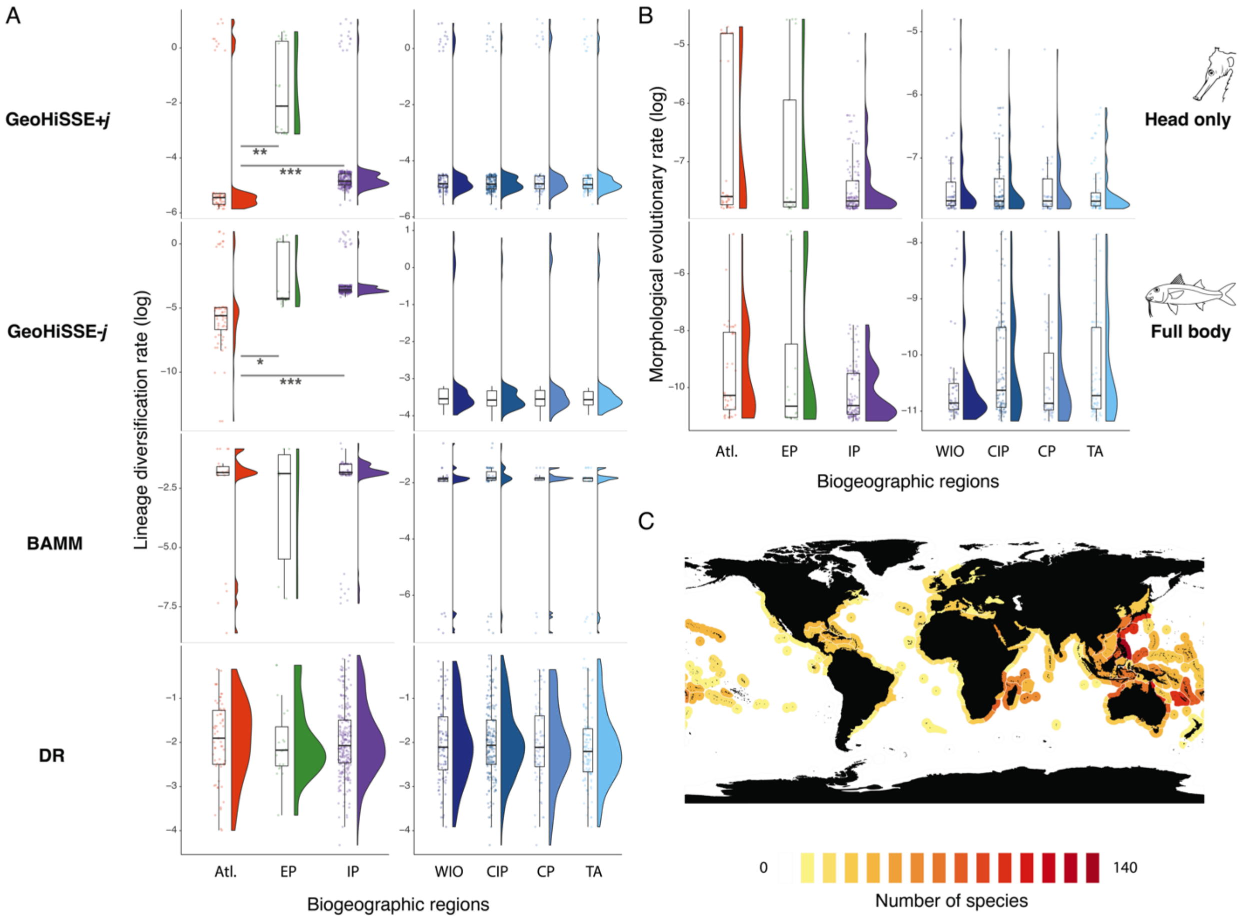


Figure 2.3 Lineage and morphological diversification rates, and species richness across biogeographic regions. The high species richness in the Indo-Pacific cannot be attributed to rate differences: both rates of lineage diversification and shape evolution are similar across both oceans regions, except for GeoHiSSE analyses which show lower rates in the Atlantic. **A)** Average log-transformed rates of lineage diversification as estimated based on three different approaches (GeoHiSSE±*j*, BAMM, and DR) and depicted using raincloud plots (half-violin plots and boxplots). **B)** Average log-transformed rates of shape evolution as estimated in BAMM with pPC scores and depicted using raincloud plots for both datasets (head-only and full-body shape). Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$). **C)** Map shows species richness for syngnatharians; colors are proportional to the number of species based on data from Marine Ecoregions of the World (MEOW). See Fig. 2.1 for other area abbreviations; for more detailed analyses see Appendix B Figs. S18–S10 and S33–S36.

Morphological disparity and rates by region

We found that disparity in head morphology (including seahorses) is higher in the IP (0.055) and EP (0.057) than in the Atl. (0.048; Fig. 2.4). However, disparity in body morphology is higher in the Atl. (0.039) and EP (0.045) than in the IP (0.030; Fig. 2.4). Suborder-level analyses reveal idiosyncratic patterns of disparity across biogeographic regions that do not necessarily reflect results for Syngnatharia as a whole (Appendix B Fig. S21). These findings hold when using different tree sources (RelTime and MCMCTree) and PC versus pPC scores (Appendix B Figs. S21–S23). Morphospace analyses show overlap in major realms and within the IP for both head and body shape datasets (Fig. 2.4 and Appendix B Fig. S24). However, head morphology exhibits higher similarity in both morphospace occupation and disparity within IP subareas, particularly between CIP and TA (Fig. 2.4A, see Appendix 2 available on Dryad). Body shape overlap is generally lower than that of head shape (Fig. 2.4B, see Appendix 2 available on Dryad).

Multivariate disparity-through-time analyses indicate that a significant proportion of morphological variation originated early in the history of Syngnatharia (ca. 87–78 Ma), followed by a steady reduction in disparity during most of the Cenozoic with a small peak towards the present (Fig. 2.4 and Appendix B Fig. S29). For head-only morphology, BMOUi or EBOUi models were best supported (AICw values 0.61 and 0.36, respectively; Appendix B Fig. S30). Most of the analyses of full-body morphology favored BM (AICw 0.82–0.96; Appendix B Fig. S31). Morphospace clustering tends to show adaptive peaks that differentiate suborder- or family-level lineages (Appendix B Fig. S32). Morphological evolutionary rates are similar across major realms and within subareas of the IP (head: median ~0.0005, body: 0.000034; EP and IP: 0.000024; Fig. 2.3B). No significant differences are found between biogeographic regions overall (Appendix B Table S7). Within Syngnathoidei, there are significant differences in head-only between the Atl. and IP (*p*=0.011; Appendix B Fig. S33 and Table S7). We observe similar results using uncorrected PC scores, with significance in head-only morphology between WIO and TA subareas (*p*=0.035) and between the Atl. and IP for full-body morphology in Syngnathoidei (*p*=0.011; Appendix B Fig. S34 and Table S7). Analyses using MCMCTree trees are largely similar (Appendix B Fig. S35), with PC scores showing significance in body morphology between the Atl. and EP (*p*=0.041) and the Atl. and IP (*p*=0.0002) for Syngnathoidei (Appendix B Fig. S36 and Table S7).

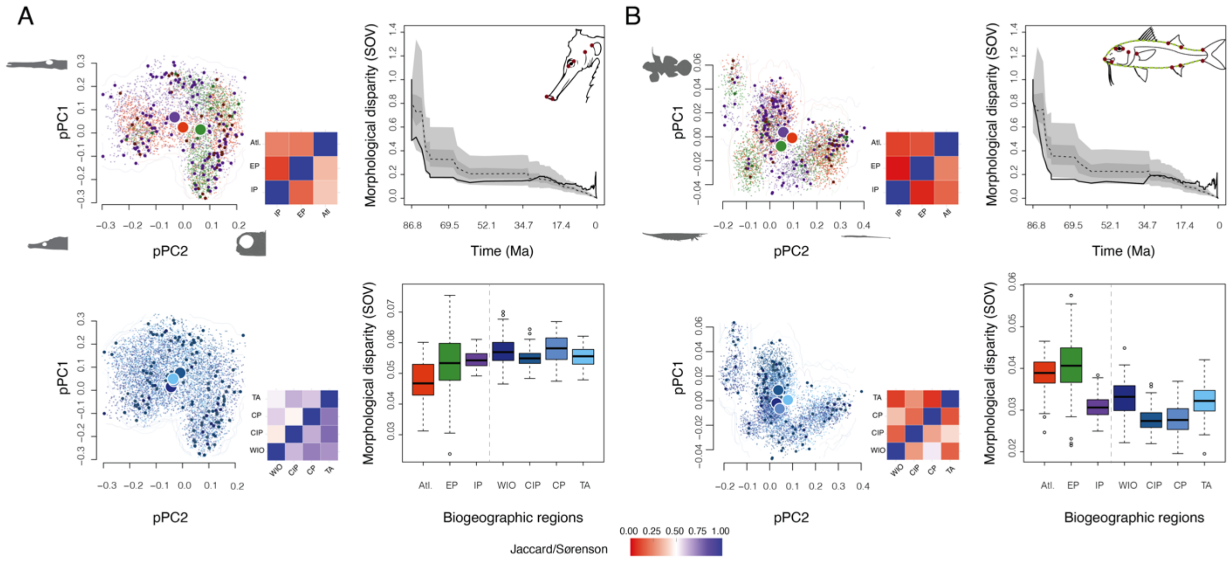


Figure 2.4 Patterns of morphospace occupation (hypervolumes) and disparity across regions and through time for syngnatharians as estimated using sum of variances (SOV) for A) head-only and B) full-body datasets. Disparity-through-time analyses show an early burst of shape evolution, while patterns of contemporary disparity across regions are relatively similar, except for full body disparity which is higher in the Atlantic and the eastern Pacific. Two-dimensional hypervolumes of pPCs 1 and 2 for head and body shape show a greater overlap within Indo-Pacific regions. Morphospace overlap/similarity is calculated based on hypervolume overlap statistics using Jaccard (above the diagonal of the heatmap) and Sørenson (below the diagonal of the heatmap) indices, where colder and warmer colors correspond to higher and lower overlap between areas, respectively. Graphic illustrations denote landmark schemes used for 2D geometric morphometric analyses. Red dots represent homologous landmarks while green dotted lines are semi-landmarks. See Fig. 2.1 for area abbreviations; for complimentary analyses see Appendix B Figs. S21–S29.

2.5 Discussion

In this study, we conducted integrative comparative analyses within a robust phylogenomic framework to test three hypotheses that may explain the high species richness of syngnatharians in the Indo-Pacific and the evolution of body plans in the group. Of these three, we found the strongest support for the center of accumulation and time-for-speciation hypotheses. Syngnatharians originated in the Tethys Sea during the Late Cretaceous, approximately 87 Ma, and it was in this region that their various body plans evolved. The higher species richness of the Indo-Pacific today is attributed to its relatively earlier colonization (after Tethys reorganization) compared to the eastern Pacific and Atlantic, plus a burst of speciation early on with diversification rates near the present remaining constant across all three realms. Furthermore, there have been many more colonizations of the Indo-Pacific than of the Atl. and EP, supporting the center of accumulation hypothesis. These colonizations also facilitated the spread of body plans to different regions, explaining the similar morphological disparity of assemblages in each region today. This similarity is not the result of independent evolution of the same body plans within each biogeographic region, but rather due to the early burst of shape evolution that likely occurred in the Tethys.

We found that the timing of the colonizations played an important role in why syngnatharian species richness is presently highest in the Indo-Pacific. Our findings indicate that after colonizing the Indo-Pacific region from the Tethys Sea, lineages underwent *in situ* speciation. A burst of speciation coincides with the PETM paleoclimatic event ~56 Ma, during which global warming and ocean acidification occurred (McInerney and Wing 2011). The fossil record shows this event also led to increased extinction in some marine fish groups (Arcila and Tyler 2017), potentially enabling syngnatharians to expand into vacant niches (Fig. 2.1C). The establishment of tropical reefs primarily composed of scleractinian corals around the same period (Wallace and Rosen 2006; Mihaljević et al. 2014; Santodomingo et al. 2015) might also have helped in accelerating speciation in syngnatharians. The early Eocene environmental conditions, which encompassed coastal biomes like seagrasses, played an important role in facilitating the diversification and rapid expansion of different reef fish clades into new ecological niches (Bellwood 2003; Goatley et al. 2010; Wainwright et al. 2012). An increase of colonization events of syngnatharian lineages into the different Indo-Pacific subareas from the CIP took place during the Miocene (23 Ma onwards). It was also during this period that major reef fish groups, such as wrasses and damselfishes, began to diversify across various regions globally, particularly flourishing within the CIP region, as observed in syngnatharians (Cowman and Bellwood 2011; Siqueira et al. 2020). Notably, this diversification coincided with the rearrangement of the IAA, a recognized hotspot for marine reef fishes which belongs to the CIP. This rearrangement occurred after the collision of the Australia and New Guinea plates, which created new shallow water habitats (Renema et al. 2008; Leprieur et al. 2016). Additionally, this timeframe witnessed the expansion of modern coral reef formations mainly by acroporids (Cowman and Bellwood 2011; Bellwood et al. 2017; Siqueira et al. 2021), zooxanthellate corals, and gastropods (Williams and Duda 2008). Consequently, coral reefs have played a dual role as both evolutionary cradles and ecological refuges for a wide array of tropical marine lineages (Bellwood et al. 2015).

While we observe that syngnatharian diversity is highest in the CIP, lineages originating within this region show high dispersal rates to adjacent areas due to its central position within the Indo-Pacific (Woodland 1983). Consequently, the CIP is acting as a source to the rest of the Indo-Pacific, as syngnatharians are expanding their range and colonizing the WIO, the CP, and TA. The high dispersal rates among these biogeographic regions suggest that the presence of soft barriers (e.g., currents) does not hinder syngnatharians from dispersing (Hughes et al. 2002; Prazeres et al. 2020). Syngnatharian species that successfully dispersed from the Indo-Pacific to both the Atlantic and eastern Pacific oceans likely did so facilitated by oceanic currents through various routes mostly during their planktonic larval stage or rafting on the pelagic sargassum and other macroalgae (Teske et al. 2005; Casazza and Ross 2008; Hamilton et al. 2017). However, most lineages that colonized either side of the Atlantic and the eastern Pacific tended to remain within their respective regions, explaining the low number of dispersal events and rates that we observe out of these regions. As a result, the Atlantic and eastern Pacific regions have generally functioned as sinks. These regions have also experienced higher extirpation rates compared to the Indo-Pacific (Appendix B Fig. S13), primarily due to a series of geologic and climatic events—the closure of the Tethys Seaway and Isthmus of Panama, and the Middle Miocene Climate Transition—that caused environmental changes, including sea-level and temperature fluctuations, ocean current circulation oscillation, salinity variations, and shifts in primary productivity (Leprieur et al. 2016; Super et al. 2020; Steinthorsdottir et al. 2021). These patterns align with some results from our lineage diversification rate analyses using GeoHiSSE±*j*, which indicate higher net diversification rates in the Indo-Pacific compared to the Atlantic. Noteworthy, however, this finding might not be a result of elevated *in situ* speciation in the Indo-Pacific but rather due to the higher extirpation rates in the Atlantic as explained above. Furthermore, it is crucial to consider the taphonomic bias affecting this group and other reef fishes, as evidenced by the significant number of fossil representatives from the Eocene Bolca Lagerstätten (western Tethys Sea (Bellwood 1996; Friedman and Carnevale 2018), in contrast to the scarcity of fossils from the Atlantic. The Atlantic exceptions are fossils described from the Tenejapa–Lacandón Formation in Palenque, Mexico, dated at 63 Ma and located close to the location where the meteorite "Chicxulub impactor" impacted three million years before. During the Palaeocene, this region could have been a marine hotspot as important as the Tethys Sea (Cantalice et al. 2022).

Although a few body plans in Syngnatharia are restricted to particular regions—for example, the bat shape of the seamoths, the macroalgae- or octocoral-like appearance of ghost pipefishes, or the camouflage-mimicking seaweed morphology of seadragons in the Indo-Pacific—what characterizes the trajectory of morphological evolution in this group is an early diversification in body plans, followed by the evolution of multiple adaptive peaks within each subclade. This pattern is intriguing as the subclade diversifications in body shape peak later at the Miocene-Pliocene transition in seahorses and goatfishes (Nash et al. 2022). Syngnatharian morphometric divergence is not primarily driven by evolution within any single region, as most body plans are represented in all the biogeographic regions: the laterally compressed bodies of snipefishes, the elongated snouts of pipefishes, seahorses, ghost pipefishes, trumpetfishes, and cornetfishes, or the fusiform shape of flying gurnards and goatfishes. In fact, we see a high morphospace overlap, similar disparities and morphological rates between the three major realms and within the Indo-Pacific when examining the shape of the head and the full body. Shallow reefs have been formed through consistent environmental and ecological pressures (Mclean et al. 2021). Despite the large differences in species richness between realms and taking into account the evolutionary history of syngnatharian clades in particular and reef fishes in general, these patterns suggest that the functional space richness is similar across regions and that regions share similar functional roles (Mouillot et al. 2014; Mclean et al. 2021). Since syngnatharians originated in the Late Cretaceous and radiated into different regions of the morphospace, all body plans had already evolved prior to the K-Pg mass extinction event. It was during this period (90–66 Ma) that acanthomorph fishes displayed an expansion in their general head shape (Sallan and Friedman 2012), despite the low complexity of reefs at this time (Kiessling 2009). However, we observe multiple adaptive peaks evolve within syngnatharians after the K-Pg, particularly as they colonized the Indo-Pacific and exploited vacant empty niches. These niches possibly emerged following the extinction of competitors, causing a burst of not only diversification of lineages but also morphological innovations and specializations in acanthomorphs (e.g., Friedman 2010; Wainwright and Longo 2017; Alfaro et al. 2018; Ribeiro et al. 2018). In fact, the higher disparity found in the head shape compared to full-body shape of syngnatharian lineages in the Indo-Pacific could be due to the trophic specializations in response to the adaptation of these new habitats (Siqueira et al. 2020). These findings seemingly conflict with a recent study which suggests that this happened instead due to a gradual accumulation of morphological disparity following the K-Pg (Ghezelayagh et al. 2021). Overall, the high species richness of coexisting syngnatharian species in the Indo-Pacific does not seem to correlate with body plan morphology, as levels of disparity are similar across biogeographic regions.

Our results align with trends observed at larger phylogenetic scales in numerous families of reef fishes. Research at these scales has emphasized the significance of time and colonization in shaping regional species richness (Miller et al. 2018), along with the presence of analogous functional traits across an oceanic longitudinal gradient (Mouillot et al. 2014; Mclean et al. 2021). Additionally, the age of Syngnatharia could influence the observed patterns, and testing these hypotheses in younger clades may yield complementary results (Clarke 2021; Miller et al. 2021; Nash et al. 2022; Diamond and Roy 2023). We did observe different levels of morphological disparity among regions at the suborder level; for example, the CP showed lower disparity in Callionymoidei, Dactylopteroidei, and Mulloidei (in both head-only and full-body analyses), while the eastern Pacific exhibited the highest disparity in Syngnathoidei (in head-only; Appendix B Fig. S21).

Uncertainties in macroevolutionary inferences and study caveats

Using an exhaustive and integrative phylogenomic approach to account for topological, divergence times, and method uncertainty is crucial in macroevolutionary analyses (e.g., Henao Diaz et al. 2019; Rincon-Sandoval et al. 2020; Goswami et al. 2022; Troyer et al. 2022). Many of our analyses have used a sample of 28 trees inferred based on independent gene subsets to address these uncertainties in historical biogeography, lineage diversification, and morphological evolution analyses. In biogeographic analyses, the major impact was between the dating method used and the implementation or omission of the jump-dispersal parameter (Ree and Sanmartín 2018; Klaus and Matzke 2020; Matzke 2022; Budd and Mann 2023), rather than topology (Santaquiteria et al. 2021). A single tree ("master tree" in this study) was required as input in biogeographic stochastic mapping analyses, and we observed that the dating method significantly impacted speciation, extinction, and colonization patterns over time (Schwartz and Mueller 2010). Conversely, for diversification rate analyses, where we calculated average tip rates across trees, the choice of methods played a crucial role, as some methods supported one hypothesis while others did not. As a result, we present our results comprehensively by considering all the methods used. While in morphological analyses we observed that pPC and PC scores produced similar results, dating methods affect the selection of the best-fit morphological evolution model in body plans. For many comparative analyses, tree uncertainty had no major effect. However, our model-fitting results on morphological evolution show substantial variation across trees (Appendix B Fig. S30), suggesting that choosing a single tree could lead to biased estimates. Despite these insights, many studies still tend to rely on a single tree and one approach to estimate lineage diversification rates (e.g., Feng et al. 2017; Xing and Ree 2017). To address uncertainty comprehensively, we advocate for adopting best practices by employing multiple trees and implementing alternative methodologies (e.g., Economo et al. 2018; Title and Rabosky 2019).

Some caveats of our study are worth mentioning. For certain groups, we may have underestimated morphological disparity due to the lack of semi-landmarks on fins (e.g., dorsal, caudal) or ornaments (e.g., the leaf-like protrusions in leafy sea dragons). Additionally, the preservation of seahorse specimens with bent structures makes their inclusion in our full-body shape analyses unfeasible. Furthermore, the limitations of 2D images prevent us from capturing the globular shapes of several syngnatharian species (e.g., flying gurnards, dragonets, seahorses), highlighting the need for future work in collecting CT scan data for 3D geometric-morphometric analyses (Buser et al. 2018; Evans et al. 2021, 2023). Further studies would benefit from exploring additional biotic variables, such as different functional traits, along with environmental factors like sea level, temperature, and primary productivity. These investigations will further enhance our understanding of evolutionary processes and their effects along this longitudinal gradient of marine fishes.

2.6 Conclusions

We find that the high diversity of syngnatharian species in the Indo-Pacific is primarily due to older colonizations followed by *in situ* speciation in the Palaeogene, right after the PETM, and lineage accrual in the Miocene at the onset of the IAA rearrangement. In contrast, the eastern Pacific and the Atlantic feature lower regional diversities due to either more recent colonization and diversification onset, with lineages in these regions mostly accruing during the Miocene. Overall, these findings best support the time for speciation and center of accumulation hypotheses. We also observe both disparity and rates of morphological evolution to be similar across areas, with clade-specific variations. Our analyses demonstrate that a considerable portion of syngnatharian morphological variation emerged early in their evolutionary history in the Tethys Sea, followed by a gradual reduction in subclade disparity punctuated by the origin of multiple adaptive peaks, especially in head morphology. This study advances our understanding of the evolutionary processes that have shaped the diversity and morphology of marine fishes in general and syngnatharians in particular, underscoring the importance of considering multiple factors affecting historical biogeographic and macroevolutionary inferences.

2.7 References

Adams D.C., Collyer M.L., Kaliontzopoulou A., Baken E. 2021. Geomorph: Software for geometric morphometric analyses. R package version 3.3.2. See https//cran.r-project.org/package=geomorph.

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Arcila D., Tyler J.C. 2017. Mass extinction in tetraodontiform fishes linked to the Palaeocene-Eocene thermal maximum. Proc. R. Soc. B Biol. Sci. 284.

Barber P.H., Bellwood D.R. 2005. Biodiversity hotspots: Evolutionary origins of biodiversity in wrasses (Halichoeres: Labridae) in the Indo-Pacific and new world tropics. Mol. Phylogenet. Evol. 35:235–253.

Bellwood D.R. 1996. The Eocene fishes of Monte Bolca: The earliest coral reef fish assemblage. Coral Reefs.

Bellwood D.R. 2003. Origins and Escalation of Herbivory in Fishes : A Functional Perspective. 29:71–83.

Bellwood D.R., Goatley C., Cowman P.F. 2015. 6 The evolution of fishes on coral reefs : fossils , phylogenies , and functions. .

Bellwood D.R., Goatley C.H.R., Bellwood O. 2017. The evolution of fishes and corals on reefs: Form, function and interdependence. Biol. Rev.

Bellwood D.R., Hughes T.P. 2001. Regional-scale assembly rules and biodiversity of coral reefs. Science (80-.). 292:1532–1534.

Bellwood D.R., Hughes T.P., Hoey A.S. 2006. Sleeping Functional Group Drives Coral-Reef Recovery. Curr. Biol. 16:2434–2439.

Blonder B., Morrow C.B., Maitner B., Harris D.J., Lamanna C., Violle C., Enquist B.J., Kerkhoff A.J. 2018. New approaches for delineating n-dimensional hypervolumes. Methods Ecol. Evol. 9:305–319.

Bowen B.W., Rocha L.A., Toonen R.J., Karl S.A. 2013. The origins of tropical marine biodiversity. Trends Ecol. Evol. 28:359–366.

Bray D.J., Gomon M.F. 2021. Fishes of Australia. Museums Victoria and OzFishNet. Available from http://fishesofaustralia.net.au/.

Briggs J.C. 1974. Marine zoogeography. New York, NY McGraw-Hill B. Company.

Budd G.E., Mann R.P. 2023. Two Notorious Nodes: a Critical Examination of Relaxed Molecular Clock Age Estimates of the Bilaterian Animals and Placental Mammals. Syst. Biol.

Buser T.J., Sidlauskas B.L., Summers A.P. 2018. 2D or Not 2D? Testing the Utility of 2D Vs. 3D Landmark Data in Geometric Morphometrics of the Sculpin Subfamily Oligocottinae (Pisces; Cottoidea). Anat. Rec. 301:806–818.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution (N. Y). 72:2308–2324.

Cantalice K.M., Alvarado-Ortega J., Bellwood D.R., Siqueira A.C. 2022. Rising from the Ashes: The Biogeographic Origins of Modern Coral Reef Fishes. Bioscience. 72:769–777.

Carturan B.S., Parrott L., Pither J. 2022. Functional Richness and Resilience in Coral Reef Communities. Front. Ecol. Evol. 10:1–19.

Casazza T.L., Ross S.W. 2008. Fishes associated with pelagic Sargassum and open water lacking Sargassum in the Gulf Stream off North Carolina. Fish. Bull. 106:348–363.

Clarke J.T. 2021. Evidence for general size-by-habitat rules in actinopterygian fishes across nine scales of observation. Ecol. Lett. 24:1569–1581.

Clavel J., Escarguel G., Merceron G. 2015. mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol. Evol.

Cowling R.M., Bradshaw P.L., Colville J.F., Forest F. 2017. Levyns' Law: explaining the evolution of a remarkable longitudinal gradient in Cape plant diversity. Trans. R. Soc. South Africa. 72:184–201.

Cowman P.F. 2014. Historical factors that have shaped the evolution of tropical reef fishes: A review of phylogenies, biogeography, and remaining questions. Front. Genet. 5:1–15.

Cowman P.F., Bellwood D.R. 2011. Coral reefs as drivers of cladogenesis: Expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. J. Evol. Biol. 24:2543–2562.

Cowman P.F., Bellwood D.R. 2013. The historical biogeography of coral reef fishes: Global patterns of origination and dispersal. J. Biogeogr. 40:209–224.

Crouch N.M.A., Jablonski D. 2023. Is species richness mediated by functional and genetic divergence? A global analysis in birds. Funct. Ecol. 37:125–138.

Darwin C. 1859. The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. .

Diamond J., Roy D. 2023. Patterns of functional diversity along latitudinal gradients of species richness in eleven fish families. Glob. Ecol. Biogeogr. 32:450–465.

Ding W.N., Ree R.H., Spicer R.A., Xing Y.W. 2020. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. Science (80-.). 369:578–581.

Duarte-Ribeiro E., Davis A.M., Rivero-Vega R.A., Ortí G., Betancur R. 2018. Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. Proc. R. Soc. B Biol. Sci. 285.

Dupin J., Matzke N.J., Särkinen T., Knapp S., Olmstead R.G., Bohs L., Smith S.D. 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeogr. 44:887–899.

Economo E.P., Narula N., Friedman N.R., Weiser M.D., Guénard B. 2018. Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nat. Commun. 9:1–8.

Edgar G.J., Alexander T.J., Lefcheck J.S., Bates A.E., Kininmonth S.J., Thomson R.J., Duffy J.E., Costello M.J., Stuart-Smith R.D. 2017. Abundance and local-scale processes contribute to multi-phyla gradients in global marine diversity. Sci. Adv. 3.

Evans K.M., Larouche O., Gartner S.M., Faucher R.E., Dee S.G., Westneat M.W. 2023. Beaks promote rapid morphological diversification along distinct evolutionary trajectories in labrid fishes (Eupercaria: Labridae). Evolution (N. Y).:1–15.

Evans K.M., Larouche O., Watson S.J., Farina S., Habegger M.L., Friedman M. 2021. Integration drives rapid phenotypic evolution in flatfishes. Proc. Natl. Acad. Sci. U. S. A. 118:1–10.

Feng Y.-J., Blackburn D.C., Liang D., Hillis D.M., Wake D.B., Cannatella D.C., Zhang P. 2017. Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous–Paleogene boundary. Proc. Natl. Acad. Sci. 114:E5864–E5870.

Friedman M. 2010. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. Proc. R. Soc. B Biol. Sci. 277:1675–1683.

Friedman M., Carnevale G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. J. Geol. Soc. London.

Froese R., Pauly D. 2021. FishBase. Available from www.fishbase.org.

Gaboriau T., Albouy C., Descombes P., Mouillot D., Pellissier L., Leprieur F. 2019. Ecological constraints coupled with deep-time habitat dynamics predict the latitudinal diversity gradient in reef fishes. Proc. R. Soc. B Biol. Sci. 286.

García-andrade A.B., Tedesco P.A., Carvajal-quintero J.D., Arango A., Villalobos F., Evolutiva L.D.M., Evolutiva R.D.B., Ecología I. De, García-andrade A.B. 2023. Same process , different patterns : pervasive effect of evolutionary time on species richness in freshwater fishes. .

Ghezelayagh A., Harrington R.C., Burress E.D., Campbell M.A. 2021. Prolonged morphological expansion of spiny-rayed fishes following the end-Cretaceous. .

Goatley C.H.R., Bellwood D.R., Bellwood O. 2010. Fishes on coral reefs: changing roles over the past 240 million years. Paleobiology. 36:415–427.

Goswami A.A., Noirault E., Coombs E.J., Clavel J. 2022. Attenuated evolution of mammals through the Cenozoic. Science (80-.). 378:277–383.

Guillerme T. 2018. dispRity: A modular R package for measuring disparity. Methods Ecol. Evol. 9:1755–1763.

Hamilton H., Saarman N., Short G., Sellas A.B., Moore B., Hoang T., Grace C.L., Gomon M., Crow K., Brian Simison W. 2017. Molecular phylogeny and patterns of diversification in syngnathid fishes. Mol. Phylogenet. Evol. 107:388–403.

Henao Diaz L.F., Harmon L.J., Sugawara M.T.C., Miller E.T., Pennell M.W. 2019. Macroevolutionary diversification rates show time dependency. Proc. Natl. Acad. Sci. U. S. A. 116:7403–7408.

Huang D., Goldberg E.E., Chou L.M., Roy K. 2018. The origin and evolution of coral species richness in a marine biodiversity hotspot*. Evolution (N. Y). 72:288–302.

Hughes T.P., Bellwood D.R., Connolly S.R. 2002. Biodiversity hotspots, centres of endemicity, and the conservation of coral reefs. Ecol. Lett. 5:775–784.

Jablonski D., Kaustuv R., Valentine J.W. 2006. Out of the Tropics : Evolutionary Diversity Gradient. Science (80-.). 314:102–106.

Jetz W., Thomas G.H., Joy J.B., Hartmann K., Mooers A.O. 2012. The global diversity of birds in space and time. Nature. 491:444–448.

Kiessling W. 2009. Geologic and biologic controls on the evolution of reefs. Annu. Rev. Ecol. Evol. Syst. 40:173–192.

Klaus K. V., Matzke N.J. 2020. Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Syst. Biol. 69:61–75.

Kool J.T., Paris C.B., Barber P.H., Cowen R.K. 2011. Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. Glob. Ecol. Biogeogr.

Kulbicki M., Parravicini V., Bellwood D.R., Arias-Gonzàlez E., Chabanet P., Floeter S.R., Friedlander A., McPherson J., Myers R.E., Vigliola L., Mouillot D. 2013. Global biogeography of reef fishes: A hierarchical quantitative delineation of regions. PLoS One. 8.

Ladd H.S. 1960. Origin of the Pacific Island molluscan fauna. Am. J. Sci. 258A:137–150.

Leprieur F., Descombes P., Gaboriau T., Cowman P.F., Parravicini V., Kulbicki M., Melian C.J., De Santana C.N., Heine C., Mouillot D., Bellwood D.R., Pellissier L. 2016. Plate tectonics drive tropical reef biodiversity dynamics. Nat. Commun. 7:1–8.

Lohman D.J., de Bruyn M., Page T., von Rintelen K., Hall R., Ng P.K.L., Shih H.-T., Carvalho G.R., von Rintelen T. 2011. Biogeography of the Indo-Australian Archipelago. Annu. Rev. Ecol. Evol. Syst.

Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48.

Matzke N.J. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Packag. version 0.2.

Matzke N.J. 2019. BioGeoBEARS - Run BioGeoBEARS on multiple trees. Available from https://github.com/nmatzke/BioGeoBEARS/blob/master/R/BioGeoBEARS_on_multiple_trees_v1.R.

Matzke N.J. 2022. Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. J. Biogeogr. 49:1805–1824.

McInerney F.A., Wing S.L. 2011. The paleocene-eocene thermal maximum: A perturbation of carbon cycle, climate, and biosphere with implications for the future. Annu. Rev. Earth Planet. Sci. 39:489–516.

Mclean M., Stuart-smith R.D., Villéger S., Auber A., Edgar G.J. 2021. Trait similarity in reef fish faunas across the world ' s oceans. Proc. Natl. Acad. Sci.:1–10.

Mihaljević M., Renema W., Welsh K., Pandolfi J.M. 2014. Eocene-miocene shallow-water carbonate platforms and increased habitat diversity in Sarawak, Malaysia. Palaios. 29:378–391.

Miller E.C., Hayashi K.T., Song D., Wiens J.J. 2018. Explaining the ocean's richest biodiversity hotspot and global patterns of fish diversity. Proc. R. Soc. B Biol. Sci. 285.

Miller E.C., Martinez C.M., Friedmand S.T., Wainwright P.C., Price S.A., Luke Tornabene. 2022. Alternating regimes of shallow and deep-sea diversification explain a species-richness paradox in marine fishes. Proc. Natl. Acad. Sci.:1–12.

Miller E.C., Mesnick S.L., Wiens J.J. 2021. Sexual dichromatism is decoupled from diversification over deep time in fishes. Am. Nat. 198:232–252.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics.

Mittelbach G.G., Schemske D.W., Cornell H. V., Allen A.P., Brown J.M., Bush M.B., Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R., McDade L.A., McPeek M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F., Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. Ecol. Lett. 10:315–331.

Mouillot D., Villéger S., Parravicini V., Kulbicki M., Arias-González J.E., Bender M., Chabanet P., Floeter S.R., Friedlander A., Vigliola L., Bellwood D.R. 2014. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. Proc. Natl. Acad. Sci. U. S. A. 111:13757–13762.

Nash C.M., Lungstrom L.L., Hughes L.C., Westneat M.W. 2022. Phylogenomics and body shape morphometrics reveal recent diversification in the goatfishes (Syngnatharia: Mullidae). Mol. Phylogenet. Evol. 177:107616.

Neutens C., Adriaens D., Christiaens J., De Kegel B., Dierick M., Boistel R., Van Hoorebeke L. 2014. Grasping convergent evolution in syngnathids: A unique tale of tails. J. Anat. 224:710–723.

Parravicini V., Kulbicki M., Bellwood D.R., Friedlander A.M., Arias-Gonzalez J.E., Chabanet P., Floeter S.R., Myers R., Vigliola L., D'Agata S., Mouillot D. 2013. Global patterns and predictors of tropical reef fish species richness. Ecography (Cop.). 36:1254–1262.

Pellissier L., Leprieur F., Parravicini V., Cowman P.F., Kulbicki M., Litsios G., Olsen S.M., Wisz M.S., Bellwood D.R., Mouillot D. 2014. Quaternary coral reef refugia preserved fish diversity. Science (80-.). 344:1016–1019.

Pellissier V., Barnagaud J.Y., Kissling W.D., Şekercioğlu Ç., Svenning J.C. 2018. Niche packing and expansion account for species richness–productivity relationships in global bird assemblages. Glob. Ecol. Biogeogr. 27:604–615.

Pigot A.L., Trisos C.H., Tobias J.A. 2016. Functional traits reveal the expansion and packing of ecological niche space underlying an elevational diversity gradient in passerine birds. Proc. R. Soc. B Biol. Sci. 283.

Prazeres M., Morard R., Roberts T.E., Doo S.S., Jompa J., Schmidt C., Stuhr M., Renema W., Kucera M. 2020. High dispersal capacity and biogeographic breaks shape the genetic diversity of a globally distributed reef-dwelling calcifier. Ecol. Evol. 10:5976–5989.

Pyron R.A., Wiens J.J. 2013. Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. Proc. R. Soc. B Biol. Sci. 280.

Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.

Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5:701–707.

Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45:741–749.

dos Reis M., Yang Z. 2019. Bayesian molecular clock dating using genome-scale datasets. Methods in Molecular Biology. .

Renema W., Bellwood D.R., Braga J.C., Bromfield K., Hall R., Johnson K.G., Lunt P., Meyer C.P., McMonagle L.B., Morley R.J., O'Dea A., Todd J.A., Wesselingh F.P., Wilson M.E.J., Pandolfi J.M. 2008. Hopping hotspots: Global shifts in marine biodiversity. Science (80-.). 321:654–657.

Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol.

Ricklefs R.E. 1987. Community Diversity : relative roles of local and regional processes. Science (80-.). 235:167–171.

Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero A., Walker Jr. H.J., Carpenter K.E., Sheaves M., Ortí G., Arcila D., Betancur-R. R. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. Proc. Natl. Acad. Sci. U. S. A.

Rohde K. 1992. Latitudinal Gradients in Species Diversity: The Search for the Primary Cause. Oikos.

Rolland J., Condamine F.L., Jiguet F., Morlon H. 2014. Faster Speciation and Reduced Extinction in the Tropics Contribute to the Mammalian Latitudinal Diversity Gradient. PLoS Biol. 12.

Rosamond K.M., Kent C.M., Hunter S.C., Sherry T.W. 2020. Morphological traits influence prey selection by coexisting species of New World warblers (Parulidae). J. F. Ornithol. 91:393–408.

Sallan L.C., Friedman M. 2012. Heads or tails: Staged diversification in vertebrate evolutionary radiations. Proc. R. Soc. B Biol. Sci. 279:2025–2032.

Sanciangco J.C., Carpenter K.E., Etnoyer P.J., Moretzsohn F. 2013. Habitat Availability and Heterogeneity and the Indo-Pacific Warm Pool as Predictors of Marine Species Richness in the Tropical Indo-Pacific. PLoS One. 8.

Santaquiteria A., Siqueira A.C., Duarte-Ribeiro E., Carnevale G., White W.T., Pogonoski J.J., Baldwin C.C., Ortí G., Arcila D., Ricardo B.R. 2021. Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences. Syst. Biol. 70:1145–1162.

Santodomingo N., Wallace C.C., Johnson K.G. 2015. Fossils reveal a high diversity of the staghorn coral genera Acropora and Isopora (Scleractinia: Acroporidae) in the Neogene of Indonesia. Zool. J. Linn. Soc. 175:677–763.

Schwartz R.S., Mueller R.L. 2010. Branch length estimation and divergence dating: Estimates of error in Bayesian and maximum likelihood frameworks. BMC Evol. Biol. 10.

Siqueira A.C., Bellwood D.R., Cowman P.F. 2019. The evolution of traits and functions in herbivorous coral reef fishes through space and time. Proc. R. Soc. B Biol. Sci. 286.

Siqueira A.C., Morais R.A., Bellwood D.R., Cowman P.F. 2020. Trophic innovations fuel reef fish diversification. Nat. Commun. 11:1–11.

Siqueira A.C., Morais R.A., Bellwood D.R., Cowman P.F. 2021. Planktivores as trophic drivers of global coral reef fish diversity patterns. Proc. Natl. Acad. Sci. U. S. A. 118.

Siqueira A.C., Oliveira-Santos L.G.R., Cowman P.F., Floeter S.R. 2016. Evolutionary processes underlying latitudinal differences in reef fish biodiversity. Glob. Ecol. Biogeogr. 25:1466–1476.

Smith S.A., O'Meara B.C. 2012. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 28:2689–2690.

Spalding M.D., Fox H.E., Allen G.R., Davidson N., Ferdaña Z.A., Finlayson M., Halpern B.S., Jorge M.A., Lombana A., Lourie S.A., Martin K.D., McManus E., Molnar J., Recchia C.A., Robertson J. 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. Bioscience. 57:573–583.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.

Steinthorsdottir M., Coxall H.K., de Boer A.M., Huber M., Barbolini N., Bradshaw C.D., Burls N.J., Feakins S.J., Gasson E., Henderiks J., Holbourn A.E., Kiel S., Kohn M.J., Knorr G., Kürschner W.M., Lear C.H., Liebrand D., Lunt D.J., Mörs T., Pearson P.N., Pound M.J., Stoll H., Strömberg C.A.E. 2021. The Miocene: The Future of the Past. Paleoceanogr. Paleoclimatology. 36.

Stephens P.R., Wiens J.J. 2003. Explaining species richness from continents to communities: The time-for-speciation effect in emydid turtles. Am. Nat. 161:112–128.

Stiller J., Short G., Hamilton H., Saarman N., Longo S., Wainwright P., Rouse G.W., Simison W.B. 2022. Phylogenomic analysis of Syngnathidae reveals novel relationships, origins of endemic diversity and variable diversification rates. BMC Biol. 20:1–21.

Super J.R., Thomas E., Pagani M., Huber M., O'Brien C.L., Hull P.M. 2020. Miocene Evolution of North Atlantic Sea Surface Temperature. Paleoceanogr. Paleoclimatology. 35:1–15.

Teske P.R., Hamilton H., Palsbøll P.J., Choo C.K., Gabr H., Lourie S.A., Santos M., Sreepada A., Cherry M.I., Matthee C.A. 2005. Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. Mar. Ecol. Prog. Ser. 286:249–260.

Title P.O., Rabosky D.L. 2019. Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? Methods Ecol. Evol. 10:821–834.

Tittensor D.P., Mora C., Jetz W., Lotze H.K., Ricard D., Berghe E. Vanden, Worm B. 2010. Global patterns and predictors of marine biodiversity across taxa. Nature. 466:1098–1101.

Troyer E.M., Betancur-R. R., Hughes L.C., Westneat M., Carnevale G., White W., Pogonoski J., Tyler J.C., Baldwin C.C., Ortí G., Brinkworth A., Clavel J., Arcila D. 2022. Cope's and Bergmann's Rules in Ectotherms: The Impact of Paleoclimatic Changes on Body Size Evolution in Marine Fishes. Proc. Natl. Acad. Sci. 119:1–9.

Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative Analysis of Principal Components Can be Misleading. Syst. Biol. 64:677–689.

Vasconcelos T.N.C., Alcantara S., Andrino C.O., Forest F., Reginato M., Simon M.F., Pirani J.R. 2020. Fast diversification through a mosaic of evolutionary histories characterizes the endemic flora of ancient Neotropical mountains. Proc. R. Soc. B Biol. Sci. 287.

Wainwright P.C., Longo S.J. 2017. Functional Innovations and the Conquest of the Oceans by Acanthomorph Fishes. Curr. Biol. 27:R550–R557.

Wainwright P.C., Smith W.L., Price S.A., Tang K.L., Sparks J.S., Ferry L.A., Kuhn K.L., Eytan R.I., Near T.J. 2012. The Evolution of Pharyngognathy: A Phylogenetic and Functional

Appraisal of the Pharyngeal Jaw Key Innovation in Labroid Fishes and beyond. Syst. Biol. 61:1001–1027.

Wallace A.R. 1869. The Malay Archipelago: The Land of the Orang-utan and the Bird of Paradise; a Narrative of Travel, with Studies of Man and Nature (Macmillan, 1869). .

Wallace C.C., Rosen B.R. 2006. Diverse staghorn corals (Acropora) in high-latitude Eocene assemblages: Implications for the evolution of modern diversity patterns of reef corals. Proc. R. Soc. B Biol. Sci. 273:975–982.

Wiens J.J. 2015. Explaining large-scale patterns of vertebrate diversity. Biol. Lett. 11:10–13.

Williams S.T., Duda T.F. 2008. Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? Evolution (N. Y).

Woodland D.J. 1983. Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns (Indo-Pacific). Bull. Mar. Sci.

Xing Y., Ree R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci. U. S. A. 114:E3444–E3451.

Chapter 3

Evolution of planktivory in extant and fossil acanthuriforms and the genomic basis of this trophic transition

Aintzane Santaquiteria, Giorgio Carnevale, Melissa Rincón-Sandoval, Fernando Meléndez-Vazquez, Willian T. White, Carol C. Baldwin, Guillermo Ortí, Matthew D. McGee, Mark W. Westneat, James C. Tyler, Dahiana Arcila, and Ricardo Betancur-R.

3.1 Abstract

Independent and recurrent transitions from benthic to pelagic habitats are often associated with the evolution of novel trophic strategies. Among the different trophic guilds, planktivory is a frequent evolutionary destination in trophic transition among freshwater and marine fishes. Given their exceptional fossil record, relatively small genomes, and well-documented trophic strategies, surgeonfishes, and allies (Acanthuriformes) provide an opportunity to investigate the evolutionary outcomes of dietary transitions by integrating diverse datasets from fossil and living species. Using phylogenetic and comparative genomic analyses, we examined the ecological and molecular drivers of the independent evolutionary transitions from non-planktivory to planktivory diet among fossil and extant acanthuriforms. By combining genomic data for 80 species (~93% of extant diversity) with morphological characters for 32 fossil taxa, we constructed a set of time-calibrated phylogenies using tip-dating approaches that incorporated both extant and extinct species. We then estimated the ancestral ranges and the evolutionary dynamics of each trophic guild. Our analyses show that acanthuriforms originated ~64 million years ago in the aftermath of the K-Pg mass extinction event. We found at least seven transitions to planktivory from non-planktivory lineages followed by at least four reversals to a non-planktivory dietary condition. Most of these transitions occurred in the Indo-Pacific, or the ancient Tethys Sea. The evolution of planktivory is driven by a confounding effect from cool past climatic temperatures and the phylogenetic signal in the trees. Contrary to our expectation of a potential evolutionary dead end in the relatively homogeneous water column environment, where opportunities for diversification may be more limited, we found that, while speciation rates are similar across trophic guilds, extinction is slightly lower in planktivores. The asymmetry in transition rates is unexpected, as there is a higher rate of transitions from planktivores to non-planktivores compared to the reverse direction; however, this pattern is mostly driven by recent reversals in unicornfishes. We also generated a chromosome-level genome and short-read genomes for 45 species to identify genes under positive selection underlying transitions to planktivory. Phylogenetic genotype-to-phenotype analyses found 91 genes that convergently evolved under positive selection in at least two planktivore lineages and three that are unique to planktivores. The most common functions of these genes are involved in metabolic processes and body shape adaptations. This study represents an important step towards understanding the macroevolutionary dynamics and genetic basis of transitions along the benthic-pelagic axis in marine fishes.

3.2 Introduction

Convergent adaptations often arise in response to shared ecological challenges or opportunities present in similar environments (Losos 2011). The transition from benthic habitats to the water column represents a pivotal shift in the ecological dynamics of aquatic organisms, often paralleled by corresponding morphological and trophic specializations (Cooper et al. 2010; Burress et al. 2017; Friedman et al. 2020). Among these specializations, the evolution of planktivory emerges as a deterministic transition, shaping the feeding strategies, ecological roles as well as functional traits of various freshwater and marine fish taxa. A well-known example is the recent multiple benthic-pelagic transitions observed in African and Central American cichlids, whitefish, and three-spined stickleback, where benthic forms specialized for bottom-dwelling invertebrate prey have transitioned to pelagic forms adapted for planktonic feeding (Rundle et al. 2000; Hulsey et al. 2013; Præbel et al. 2013; Elmer et al. 2014). While these sympatric lineages reveal strong evolutionary forces driving divergence along this vertical axis, there are also evident transitions occurring at deeper evolutionary scales. For example, lutjanids (snappers and fusiliers; Rincon-Sandoval et al. 2020), haemulids (grunts; Tavera et al. 2018) and acanthurids (surgeonfishes; Friedman et al. 2016) have transitioned from benthic to midwater habitats multiple times over the last 60 Ma.

Among reef fishes, trophic shifts are also a recurring phenomenon, with dietary identity emerging as one of the most important explanatory variables in reef ecosystems (Siqueira et al. 2020). As fishes adapt to varying dietary preferences, their impact on ecological niches and overall diversity within reef ecosystems becomes increasingly evident. In fact, planktivorous fishes are characterized by unique attributes closely associated with their feeding behavior. While the larval and early juvenile stages of many fish species exhibit planktivory, only those possessing suitable adaptations typically persist as obligate planktivores in the adult stage (Hobson 1991). Previous studies have shown that marine diversity is higher in benthic (bottom-living) compared to water column environments (Gray 1997; Duarte-Ribeiro et al. 2018), suggesting that the adaptation to a planktivore lifestyle may present an evolutionary dead end. While planktivores may initially diversify due to the absence of competition for resource in vacant niches, in the long term, maintaining this diet in an resource-poor and homogeneous environment could potentially lead to reduced speciation rates or elevated extinction rates (Van Valkenburgh et al. 2010; Rincon-Sandoval et al. 2020). Understanding trophic shifts enables us to gain insight into species' adaptive capacity and evolutionary responses to environmental changes, encompassing both historical (paleoclimatic) and potential future effects of climate change, and how these impacts subsequently affect the ecosystem. Past climatic ocean temperatures have been shown to also influence the diversity of marine taxa, with drastic temperature changes causing extinction (Harnik et al. 2012; Reddin et al. 2019) while also promoting diversification (Davis et al. 2016). To our knowledge, no studies examining the effect of temperature in trophic transitions have been conducted.

While ecological factors commonly drive these transitions, the potential role of genetic factors has received minimal attention (but see De-Kayne et al. 2022; Ahi et al. 2023). A fundamental question concerns the extent to which adaptive convergence in trophic specializations is caused by convergent changes at the molecular sequence level (Parker et al. 2013; Vizueta et al. 2019; Chen et al. 2023; Eastment et al. 2024). Genomic innovations have been found to likely contribute to fuel different ecological specializations, such as habitat transitions (Aristide and Fernández 2023), vision (Musilova et al. 2019), or diet (Roycroft et al. 2021; Singh et al. 2022). Determining which genes or regulatory regions are responsible for a dietary change and understanding how these DNA regions evolved convergently across distantly related species are important steps to shed light on the underlying mechanisms of adaptation. Previous research, primarily in mammals and insects, suggests that dietary evolution is associated with gene family expansions (Seppey et al. 2019), changes in gene copy number (Li and Zhang 2014), or loss of gene function (Hecker et al. 2019) involved in metabolism. Genes linked to appetite regulation like anatomical compartments of gastrointestinal tract and digestive enzymes have been identified in zebrafish (Ahi et al. 2022). As transitions to the water column are associated with both a shift to a planktivore diet and morphological changes, previous studies in cichlids, whitefish, or sticklebacks have also identified genes linked to body elongation and caudal fin shape (Ahi et al. 2023), gill raker counts (Glazer et al. 2014; De-Kayne et al. 2022) and jaw length (Shapiro et al. 2009). As mentioned above, most studies evaluating genomic signatures linked to trophic and habitat transitions have been conducted in freshwater fishes. However, such studies are rare for marine fishes, primarily due to the limited availability of high-quality assemblies for reef species, with only a few representatives available from public repositories [e.g., butterflyfish Zhang et al. (2023), wrasse Liu et al. (2021), and parrotfish Tea et al. (2024)]. For example, a recent study examining the genome of a parrotfish, known for its highly specialized herbivorous dietary niche facilitated by abrasion-resistant biomineralized teeth, found and expansion of detoxification gene families (Tea et al. 2024).

Reef fishes in the order Acanthuriformes, which includes the charismatic surgeonfishes, louvars, and zanclids, provide an opportunity to investigate dietary transitions among both extant and fossil species, as well as the genomic underpinnings of these transitions in living species. This is made possible by their relatively small genomes (*ca.* 700 Mb), well-documented ecological life histories, and exceptional fossil record (over 30 fossils) that can be used as a proxy to investigate how their diet has been changing over time (Bellwood et al. 2014; Friedman and Carnevale 2018). While much has been learned from comparative analyses about the role of morphology and diversification dynamics associated with trophic transitions in acanthuriforms (Klanten et al. 2004; Friedman et al. 2016; Siqueira et al. 2020; Tebbett et al. 2022), integrating data from fossils and extant species offers an avenue to explore the macroevolutionary and biogeographic history of this clade. A study using a fraction of known fossil and living species (Bellwood et al. 2014) proposed that the emergence of surgeonfishes took place in two distinct phases: first during the Paleocene/Eocene (~66–34 Ma) following the K-Pg mass extinction event, potentially due to the emergence of vacant niches; and second in the Oligocene/Miocene (~23–5 Ma), coinciding with the establishment of large coral reef habitats (Flügel and Kiessling 2002). Trophically, acanthuriforms can be divided in herbivores and detritivores, which eat turf algae, macroalgae, and detritus (non-planktivores hereafter); and planktivores that feed on zooplankton or small crustaceans (Siqueira et al. 2020). While this group is widely distributed across the different tropical realms, planktivore species are predominantly found in the Indo-Pacific Ocean (Siqueira et al. 2019; Tebbett et al. 2022). Previous studies focused on surgeonfish trophic transitions have suggested that the planktivore lifestyle independently originated at least five times (Friedman et al. 2016; Tebbett et al. 2022). However, these studies used a phylogenetic framework that relied on a restricted set of mitochondrial and nuclear genes, without integrating paleontological data. Additionally, our understanding of the genomic basis related to trophic transitions in marine fishes remains exceedingly rare when examined in a macroevolutionary framework.

This study aims to investigate the tempo and mode of the evolution of planktivory in fossil and extant acanthuriform species, as well as the genomic basis of this trophic transition. More specifically, we estimated the timing and locations of trophic transitions in acanthuriforms, along with their frequency. We also investigated whether planktivory constitutes an evolutionary dead end, examined the correlation between diet shifts and paleoclimatic changes, and explored the connection between genes subject to diversifying selection and transitions to planktivory. To address these questions, we first estimated a set of calibrated trees using a total-evidence dating approach, combining genome-wide data from nearly a thousand loci for 80 extant species and morphological data (107 characters) coded from 32 fossil and 20 extant species. Based on these phylogenies, we estimated the ancestral trophic and geographic states, performed diet-dependent diversification analyses by estimating speciation and extinction rates, and evaluated the influence of ocean paleoclimatic temperatures on trophic shifts. To explore the genomic signatures of planktivore transitions, we sequenced and assembled a chromosome-level genome from a surgeonfish representative as well as short-read genomes for 45 acanthuriform species. We employed phylogenetic genotype-to-phenotype (PhyloG2P) approaches to identify genes under positive selection across planktivore lineages and along branches where a transition to planktivory occurred. We hypothesize that the evolution of trophic strategies in acanthuriform fishes has been driven by past geological events and ocean temperature changes, with the transition to a planktivorous diet potentially representing an evolutionary dead end.

3.3 Materials and Methods

See Appendix C, Supplementary Materials and Methods section for additional methodological details.

DNA extractions, exon capture, sequencing, and assembly

We generated new exon capture data from tissue samples extracted from museum voucher specimens for a total of 57 acanthuriforms (*Acanthurus tractus* duplicated) and 9 outgroups from closely related families. High quality DNA extractions were sent to Arbor Biosciences for library preparation and exon capture to target the 1,105 single-copy exons developed for the FishLife project (Hughes et al. 2018) using the Eupercaria-specific probe set (Hughes et al. 2020), that also includes PCR-based 29 legacy markers (mtDNA and nuclear genes) commonly used for fish phylogenetics (e.g., Li et al. 2007; Betancur-R et al. 2013; Broughton et al. 2013). Enriched libraries were sequenced using one lane of the Illumina HiSeq 4000 platform with paired-end 100bp at the University of Chicago Genomics facility. The raw sequence data was assembled and aligned using the bioinformatic pipeline developed by Hughes et al. (2020; available at https://github.com/lilychughes/FishLifeExonCapture/). The final step of the pipeline generate alignments in their correct reading frames for each exon using MACSE v. 2.03 (Ranwez et al. 2018). After implementing a set of quality control steps, the final reduced molecular matrix assembled consists of 998 genes for 56 ingroup species (including a duplicated *A. tractus*) out of the 86 extant acanthuriforms (~65.1%), representing all genera and families, along with 9 outgroup species as outlined above.

Taxonomic sampling augmentation

To increase the number of species, we generated an expanded matrix combining the 91 newly sequenced species (reduced matrix) with sequences for up to 29 markers obtained from GenBank. We first individually aligned each legacy marker from GenBank using MACSE. We then aligned them with their corresponding legacy marker sequenced on our reduced matrix. The final expanded molecular matrix comprises of 1002 genes for 80 ingroup species (~93%) and 9 outgroup species (65 spp. with FishLife exons and 25 spp. with legacy markers).

Phylogenomic inference

For each assembled molecular matrix, reduced and expanded, we inferred maximum likelihood (ML) trees and multispecies coalescent species trees. First, we determined the best-fitting partition scheme for each matrix using PartitionFinder2 (Lanfear et al. 2017) based on a priori by-codon partitions for each protein-coding marker, and two partitions for each of the ribosomal markers (12S and 16S). We estimated concatenation-based ML trees in RAxML v. 8.2.11 (Stamatakis 2014) using the best-fit partitioning schemes selected via the Bayesian Information Criterion (BIC) and the GTRGAMMA model. Using the *raxml-ng* (extension of RAxML for supercomputers; Kozlov et al. 2019) we ran 30 independent ML searches and used 100 nonparametric bootstrapping to assess edge support. To infer species trees while accounting for incomplete lineage sorting (ILS), we initially estimated individual gene trees in RAxML using by-codon partitions. All mtDNA markers were grouped into a single locus alignment, with by-codon partitions applied specifically for protein-coding genes and two partitions for 12S and 16S, as explained above. After inferring best trees from multiple runs and bootstrap support (BS) values, we collapsed gene tree branches with low BS (<33%). We then conducted multispecies coalescent species-tree analyses with multi-locus bootstrapping in ASTRAL-III (Zhang et al. 2017) using collapsed gene trees as input to generate a species tree for each matrix. We also assessed gene concordance factors (Minh et al. 2020) by calculating the percentage of gene trees in the data matrix that support a specific branch in the concatenation-based (RAxML) and multispecies coalescent-based (ASTRAL-III) species trees inferred for both datasets (Minh et al. 2020).

Integration of fossils and extant species

We newly coded a morphological matrix consisting of 107 characters for 32 fossil and 19 extant acanthuriform species plus 5 extant outgroups (see below for the list of osteological characters and character states). To assess the phylogenetic placement of each species based on morphology, we inferred trees based on parsimony and ML approaches. We estimated the parsimony tree in TNT v. 1.5 (Goloboff and Catalano 2016) using a driven-search strategy (sectorial ratchet, tree-fusing methodologies) with default parameters. The ML morphological tree was estimated using the MULTIGAMMA and Mk models with 30 iterations in RAxML. We combined the morphological and the expanded molecular matrices for a total of 112 fossil and extant ingroup species and nine outgroups. We estimated the combined matrix in RAxML using the MULTIGAMMA and Mk models, 100 bootstraps and six partitions: five for the molecular sequences (one for each codon position of all nuclear and mtDNA protein-coding markers, plus two for 12S and 16S) and one for the morphological dataset. Taxa with polymorphic character states were coded as missing ("?") for RAxML, which cannot handle polymorphic characters.

Total-evidence dating analyses and phylogenetic uncertainty

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 (Ronquist et al. 2012). To account for topological uncertainty, we assembled largely independent subsets (randomly subsampled from the expanded matrix with genes only), each with enough genes to overcome sampling error. Each genomic subset was combined with the morphological dataset with fossil and extant taxa to include a total of 112 taxa. The ages of each fossil used to estimate divergence times are provided in Appendix C Table S2. After independently estimating phylogenies based on datasets with morphology only and combining molecules with morphology, we noticed that there were two fossils, †*Gazolaichthys vestenanovae* and †*Padovathurus gaudryi*, that had incongruent placements in the resulting trees. In the combined matrix these fossils appear as the sister group of Zanclidae+Acanthuridae (Scheme 1, Appendix C Fig. S7), while on the morphology-only dataset (Appendix C Fig. S6) and a previous study (Siqueira et al. 2019) their placement was sister to Acanthuridae (Scheme 2). To address this ambiguity, we chose to include these two schemes as constraints in our dating analyses, conducting a total of 10 analyses by running MrBayes for each scheme on every subset. Each analysis was run with eight independent runs and four Monte Carlo Markov chains (MCMCs) for over 350 million generations each, sampling every 10,000 generations. We used a sample probability of 0.94 and a relaxed clock model with the clock rate prior following a log normal distribution and independent gamma rate (IGR). The first 10% of trees sampled were discarded as relative burn-in and convergence of the MCMC was verified using the estimated sample size (ESS) criterion for each parameter in TRACER v. 1.7 (Rambaut et al. 2018). After more than 14 months of total runtime, we found that 9 (of the 10) analyses reached convergence where ESS values were close to or above 200 (Appendix C Table S4). Because these analyses ran for over a year, we removed the Subset 2 based on the Scheme 2 for all downstream analyses. We sampled ~2000 trees for Scheme 1 and ~2500 trees for Scheme 2 evenly distributed along the posterior distribution from each subset to have a total of 10000 trees. For each scheme independently, we inferred a Maximum Clade Credibility tree (MCC tree) in TreeAnnotator v. 2.7.5 (Drummond and Rambaut 2007). To obtain a posterior distribution (PD) of trees for phylogenetic comparative analyses, depending on the type of analysis and their computational time, we sampled either 100 or 4 trees from each subset, resulting in a total of 500 and 20 trees for each scheme.

Diet and biogeographic data

We compiled a discrete diet database for extant and fossil taxa (i.e., non-planktivore vs. planktivore). While extant species were coded based on the diet composition from existing literature, fossil diet is typically determined based on their tooth morphology, paleoecology, and paleoenvironment (Purnell et al. 2012; DeSantis 2016). However, for surgeonfishes, this task is particularly daunting given the complex tooth morphology in the group. We thus examined if there is a correlation between tooth morphology and diet for extant species only by codifying three types of teeth: conical, multi-denticulate, and brush-like. We then conducted a phylogenetic regression to determine their relationship, finding no significant evolutionary correlation (p = 0.842; Appendix C Fig. S11). We find cases where both non-planktivore and planktivore fishes displayed multi-denticulate or conical tooth morphologies. This result indicates that factors other than tooth morphology should be used to determine the diet of acanthuriform fossils. Therefore, we focused on the paleoenvironment of each fossil, categorizing fossils derived from oceanic sediments as planktivores and from limestones as non-planktivores (Marramà et al. 2016; Friedman and Carnevale 2018). Two fossils (†*Gazolaichthys vestenanovae* and †*Padovathurus gaudryi*) were coded as ambiguous.

We also gathered geographic distribution data and built a presence/absence matrix by coding each extant and fossil species according to their geographic ranges primarily based on the IUCN Red List (IUCN 2021), Ocean Biogeographic Information System (OBIS 2021), and Paleobiology database (http://paleodb. org) as well as from the primary literature. We used a seven-region biogeographic scheme (based on Spalding et al. 2007 and Kulbicki et al. 2013): Western Indian Ocean (WIO), Central Indo-Pacific (CIP), Central Pacific (CP), Tropical Eastern Pacific (TEP), Western Atlantic (WA), Eastern Atlantic (EA), and Tethys Sea (Tet).

Ancestral diet and ancestral range estimates

To examine the frequency of shifts to planktivory and their timing, we conducted ancestral diet reconstruction analyses using “make.simmap” function in the R package *phytools* (Revell 2012). We fixed the root state to non-planktivory, considering that close acanthuriform relatives primarily feed on algae and benthic invertebrates, such as Chaetodontidae (Hodge et al. 2021) and Pomacanthidae (Baraf et al. 2019). We first assessed whether these trophic transitions fit to a model where shifts between non-planktivory (state 0) and planktivory (state 1), and vice versa, exhibit equal rates (ER), or if instead these transitions are different (ARD; (Revell 2012). We then conduced stochastic character mapping using the best-fit model on the MCC tree and across 20 trees of each scheme. Transitions were counted when a nodal pie is >50% of state 0 and one of its descendant branches is >50% for state 1 and vice versa. Exceptions were made following parsimony principle when the nodal pie with one state was between two nodes with the alternative state, resulting in fewer transition count. For comparison purposes only, we also ran the analysis excluding all the fossils for MCC tree from Scheme 1.

We also investigated where trophic transitions occurred. To this end we conducted ancestral range reconstruction analyses in the R package *BioGeoBEARS* (Matzke 2013) following the approach outlined in Santaquiteria et al. (2021). We evaluated 12 biogeographic models combining DEC (Ree and Smith 2008), DIVA (Ronquist 1997), and BayAREA (Landis et al. 2013), with and without the jump-dispersal or founder-speciation event (j) (Matzke 2014) and the dispersal matrix power exponential (w) parameters (Dupin et al. 2017). We analyzed each model using three time slices based on two major geological events: (i) prior to the closure of the Tethys Seaway (65–16 or 12 Ma), (ii) after the closure of the Tethys Seaway and prior to the last rising of the Panama Isthmus (16 or 12–2.8 Ma), and (iii) after the last rising of the Panama Isthmus (2.8–0 Ma; O'Dea et al. 2016). Due to the ambiguous age of the total closure of the Tethys Seaway, the terminal Tethyan event, we decided to ran separate biogeographic analyses assuming 16 Ma (Adams et al. 1983) and 12 Ma (Steininger and Rögl 1979, 1984; Rögl 1999) as the final closure. We also accounted for connectivity between areas by implementing three different dispersal probability categories: 1 (high connectivity), 0.5 (intermediate separation), and 0.0001 (wide separation/no connectivity). We ran these analyses for both schemes using the MCC trees with fossils as input phylogenies. We then selected the best-fit biogeographic model based on the Akaike Information Criterion scores corrected for small sample size (Appendix C Table S5). The best-fit model was implemented in 20 time trees and the results were summarized by overlying average probabilities across compatible nodes on each MCC tree (Matzke 2019).

Diet-dependent diversification

We estimated diversification and transition rates between non-planktivory and planktivory diets, by conducting state-dependent speciation and extinction analyses (SSE) in *HiSSE* (Beaulieu and O'Meara 2016). Because SSE analyses cannot handle non-ultrametric trees, for these analyses we used extant-only trees (but see below). We fitted a total of five SSE models with and without unknown ('hidden') states: null BiSSE model (rates independent to the trait), full BiSSE or BiSSE like HiSSE (rates dependent to the trait), (iii) full HiSSE model (rates dependent to the trait with hidden states), CID-2 model (character-independent model with two hidden states), and CID-4 model (character-independent model with four hidden states). We defined sampling fractions for each trophic guild to account for missing taxa; non-planktivores 91% and planktivore 100%. As with ancestral diet reconstruction analyses, we also fixed the root to non-planktivory. We ran this analysis on the MCC tree and across 20 trees of each scheme. To incorporate uncertainties in model choice, we then model-averaged rates for all tips and nodes in the trees calculated under five models using AIC weights. We then used the R package *ggplot2* (Wickham 2016) to generate boxplots plots and assess differences in diversification rates between two the states for all tips and nodes. Finally, we statistically compared tip-associated lineage diversification rates between the trophic guilds conducting a phylogenetic ANOVA in *phytools* (Revell 2012) with Bonferroni correction.

To incorporate fossil data into our diversification analyses, we estimated the number of lineages over the evolutionary history of this group for the two trophic guilds. We generated lineage-through-time plots in *phytools* (Revell 2012) using the 500 trees for each scheme. For comparison purposes only, we also ran this analysis excluding all the fossils for MCC tree from Scheme 1 with 100 simulations.

Paleoclimate-dependent diet evolution

We compiled paleoclimatic temperature curves spanning the last ~64 Ma of acanthuriform's evolutionary history, to examine the association of planktivory to two alternative temperatures: the global average ocean temperatures obtained from oxygen isotope data and the tropical ocean temperatures based on sea surface temperatures from tropical latitudes (Scotese et al. 2021). We assessed the association of the global and tropical paleoclimatic temperature curves with the evolution of planktivory using a modified regression threshold model (Felsenstein et al. 2012). We fitted a total of four different evolutionary models for discrete data separately for both temperature curves: three climate-independent models, Brownian motion, early burst, and phylogenetic signal or lambda, and the climate-dependent model recently developed (Melendez-Vazquez et al., in prep.). We used 100 integrations (N=100) and a climatic spline interpolation function based on 500 degrees of freedom (df= 500). For each curve we fitted all these models using both the MCC tree and the 500 trees of each scheme.

Sequencing and assembly of chromosome-level genome

We generated a phased chromosome-level genome and transcriptome for *Acanthurus chirurgus*. We obtained flash-frozen muscle tissues from an individual caught using a hand net in the Florida Keys (24°59.564 N, 80°25.753 W), US, by Phillip Rauch on the 29th of January of 2023. The voucher specimen is deposited at Scripps Institution of Oceanography (SIO), collection number SIO 24-10. Subsequently, we outsourced the DNA and RNA extractions, library preparations, sequencing, assembly, and annotation to Cantata Bio LLC. Briefly, high molecular weight DNA extraction was followed by PacBio library construction and sequencing. The resulting PacBio reads were assembled into scaffolds using Hifiasm v. 0.15.4-r347 (Cheng et al. 2022), yielding one de novo assembly for each haplotype. To improve genome architecture through scaffolding and read orientation, Dovetail Omni-C library sequencing was performed on an Illumina HiSeqX platform at ~30x sequence coverage. The *de novo* assembly and Dovetail OmniC library reads were then assembled with for Omni-C HiRise (Putnam et al. 2016). Transcriptomic data was obtained by extracting RNA, preparing libraries, and sequencing on a NovaSeq6000 platform. Genome completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.0.0 (Simão et al. 2015) based on the single-copy orthologs for ray-finned fishes database (actinopterygii_odb9). Scaffolds were assigned chromosomal numbers based on their length as no other surgeonfish genome was available at the time of assembly. For the genome annotation, first repeat families found in the genome assembly of *A. chirurgus* were identified *de novo* and classified using the software package RepeatModeler v. 2.0.1 (Flynn et al. 2020). The custom repeat library obtained from RepeatModeler were used to discover, identify, and mask the repeats in the assembly file using RepeatMasker v. 4.1.0 (Smit et al. 2013–2015). Finally, the coding sequences across the genome were predicted using AUGUSTUS (v. 2.5.5; Stanke et al. 2006), SNAP (v. 2006-07-28; https://github.com/ KorfLab/SNAP), and MAKER2 (Holt and Yandell 2011) software.

Sequencing and assembly of short-read genomes

Using the libraries prepared at Arbor Biosciences, we sequenced genomes for a total of 48 acanthuriform species (average genome size 0.8 Gbp). We sequenced short reads at 30x coverage using the Illumina NovaSeq S4 PE150 platform at the Oklahoma Medical Research Foundation (OMRF) sequencing facility. We pre-assembled the raw data in MaSuRCA v. 4.0.8 (Zimin et al. 2013) and removed three species due to bad quality. We then enhanced scaffolding of each genome by using as a reference the chromosome-level genome for *A. chirurgus* in RagTag v. 2.1.0 (Alonge et al. 2022) to reorder, reorient, and fill gaps. Finally, we assessed genome completeness for the 45 species in BUSCO (Simão et al. 2015) using the same procedure as for chromosome-level genome. This analysis also identified single-copy and duplicated genes within each species based on the ray-finned fishes database.

Variations in transposable elements linked to trophic shifts

First, we used Tandem Repeats Finder for the identification of tandem repeats (Benson 1999). We then conducted both homology-based and *de novo* methods to detect transposable elements (TEs). The homology-based TE identification for all species was carried out using RepeatMasker and based on the Repbase library (Bao et al. 2015). Additionally, the *de novo* TE annotation was conducted using RepeatModeler with default settings to create a *de novo* repeat library for each assembled genome. Subsequently, we employed the *de novo* repeat library in conjunction with RepeatMasker to predict soft-masked repeats for each species. The identified repeat elements were categorized into ten distinct class/subfamilies, encompassing long interspersed nuclear elements (LINE), short interspersed nuclear elements (SINE), long terminal repeats (LTR), transposons, rolling circle elements, small RNAs, satellites, simple repeats, low complexity repeats, and unclassified elements. To evaluate the correlation between the quantity of TEs (total and for each class/subfamily) and trophic guilds, we conducted a phylogenetic ANOVA in *phytools* (Revell 2012) using Bonferroni correction.

Identification of ortholog genes

We identified single-copy genes among species based on the genes retained by BUSCO in OrthoFinder (Emms and Kelly 2019). We pruned the MCC tree from Scheme 1 to match the species with genomes and use it as the guide tree. OrthoFinder identified 1269 ortholog genes among the 45 species. Each gene was aligned in MACSE (Ranwez et al. 2018) using the same approach as for the exon capture data. As positive selection analyses are sensitive to alignment error (Jordan and Goldman 2012), we cleaned the edge of the alignments by only allowing the maximum proportion of gaps to be 20% in MACSE and then removed genes with less than 70% of the species (31 or less) and shorter than 200bp. We then visually inspected all the alignments to adjust the reading frames, remove poor-quality reads and correct misaligned sections in Geneious (Kearse et al. 2012), retaining 1250 single copy orthologs.

Positive selection linked to trophic shifts

To identify instances of positive selection associated with planktivory, we used the dN/dS metric (the ratio of non-synonymous substitutions or dN to synonymous substitutions or dS) implemented in HyPhy (Kosakovsky Pond et al. 2020) to interrogate the 1250 single-copy genes. The MCC tree from Scheme 1 was pruned to retain the species present in each gene alignment, and the respective gene alignments were used as input. We first ran BUSTED-E to screen each alignment for regions that might exhibit 'odd' variation patterns (ω>100 and weight <1%), which can be indicative of false signals of positive selection arising from sequencing or alignment error. To do that we selected the branches with planktivore diet as foreground. This analysis identified 126 high confidence positively selected genes (PSGs), which we retained for downstream analyses. Subsequently, we assessed the incidence of positive selection occurring exclusively on the branches where a transition to planktivory occurred, using the branch-site model with adaptive branch-site random effects likelihood (aBSREL; Smith et al. 2015). For the aBSREL analysis, we selected the branches that experienced transitions to planktivory based on ancestral diet reconstruction analyses (five total, see Fig. 3.3A). This test thus assesses each specific branch, facilitating the identification of convergence signals when distinct branches featuring a dietary transition independently display positive selection in the same set of genes. To illustrate the number of convergent genes in each transition to planktivory, we created an upset plot using the function UpSet in the *UpSetR* R package (Conway et al. 2017). Additionally, we examined whether selection that occurred in any lineages is linked to a specific trophic guild (non-planktivore or planktivore) using BUSTED-PHenotype (BUSTED-PH; Spond 2022). In this analysis, we designated both stem and crown lineages with a planktivore state as foreground, non-planktivores as background, and reversals to a non-planktivore diet as nuisance, which are not considered in the analysis (Fig. 3.3D). We identified positively selected genes (PSGs) after correcting for a false discovery rate (FDR) p-value <0.05 for aBSREL and BUSTED-PH analyses. Lastly, we determined gene ontology (GO) information (e.g., GO terms, gen family name, and biological functions) for each PSG using the Protein Analysis Through Evolutionary Relationships (PANTHER) database in the PANTHER Classification system (https://www.pantherdb.org; Thomas et al. 2022).

3.4 Results

Phylogenomic inference, total-evidence dating, and tree uncertainty

The time-calibrated MCC trees, derived from total-evidence dating analyses of five independent subsets that integrate morphological and non-overlapping molecular data for 112 species (32 fossil, 80 extant), generally exhibit stable phylogenetic relationships. This stability holds true when compared with different phylogenies obtained by previous molecular (Clements et al. 2003; Klanten et al. 2004; Sorenson et al. 2013a; Friedman et al. 2016; Betancur-R. et al. 2017; Siqueira et al. 2019; Ghezelayagh et al. 2022) and morphological (Winterbottom 1993) studies. All three families are monophyletic, with Luvaridae being the sister group of Zanclidae plus Acanthuridae.

While the topologies inferred from exon and legacy markers, using both reduced and expanded datasets with concatenation ML (Appendix C Figs. S1 and S3) and summary multispecies coalescent approaches (Appendix C Figs. S2 and S4), along with the combined matrix in RAxML (Appendix C Fig. S7), consistently show similar placements, phylogenies based exclusively on morphological data reveal the non-monophyly of Zanclidae and Acanthuridae. Additionally, these morphological-based phylogenies do not position Luvaridae as the earliest branching family (Appendix C Fig. S6). All genera, except for the fossil genus *Tylerichthys* and the extant genera *Ctenochaetus* and *Acanthurus*, are monophyletic, with the latter two being paraphyletic, a result that is also consistent with previous studies (Clements et al. 2003; Sorenson et al. 2013b; Siqueira et al. 2019). Additionally, gene concordance factor analyses based on only molecular data reveal several clades with consistent gene tree and species tree relationships (e.g., the genera *Naso*, *Zebrasoma* or *Ctenochaetus*), with few exceptions such as the placement of *Acanthurus thompsoni* or the clade containing the most recent common ancestor of *A. dussumieri* and *A. grammoptilus* (see Appendix C Fig. S5).

The placement of the †*Gazolaichthys vestenanovae* and †*Padovathurus gaudryi* fossils, initially suggested to be within the family Acanthuridae based on morphological data alone (Appendix C Fig. S6 and a previous study by Siqueira et al. 2019), is now placed outside of this clade, and identified as the sister group of Zanclidae and Acanthuridae (Appendix C Fig. S7). Although two constraints were applied for tip-dating analyses, we support their position in Scheme 1 over Scheme 2, given Tyler's (1970) conclusion that the osteological differences between the families Acanthuridae and Zanclidae are so limited that they could potentially be placed within a single family. From an anatomical perspective, the demarcation of two distinct families primarily stems from historical inertia and the taxonomic *status quo*. Thus, here we report the phylogenetic comparative results based on the time trees obtained from Scheme 1 and from Scheme 2 in the Appendix C, Supplementary Materials and Methods. Total evidence dating analyses suggest that acanthuriforms originated in the Paleocene at 63.6 Ma (95% highest posterior density [HPD]: 64–63.3 Ma), soon after the K-Pg mass extinction event, rather than the previously suggested 80 Ma based on more limited dataset that lacked a morphological matrix for fossil placement (Siqueira et al. 2019). Luvars originated 62.1 Ma (95% HPD: 63.8–60.6 Ma), followed by unicornfishes, surgeonfishes and tangs (Acanthuridae) at 61.7 Ma (95% HPD: 62.9–59.3 Ma), and zanclids at 57.8 Ma (95% HPD: 62.1–54.6 Ma). At the genus level within acanthurids, most fossil genera originated between 60 and 50 Ma (e.g., *Avitoluvarus*, *Tauichthys*, and *Proacanthurus*), with *Naso* diversifying at 26.4 Ma (95% HPD: 28.5–23.5 Ma), *Prionurus* at 14.4 Ma (95% HPD: 29.1–12.2 Ma), *Paracanthurus* at 25.0 Ma (95% HPD: 27.5–12.9 Ma), *Zebrasoma* at 17.9 Ma (95% HPD: 23.2–9.8 Ma), *Acanthurus* at 23.2 Ma (95% HPD: 31.7–18.5 Ma), and the clade including *Acanthurus* + *Ctenochaetus* at 11.6 Ma (95% HPD: 17.4–8.2 Ma).

Tempo, mode, and geography of planktivory evolution

We reconstructed the diet of acanthuriforms to investigate the number of times the planktivory lifestyle evolved, as well as the regions and timing of these transitions. Ancestral state reconstruction analyses were conducted based on the best-fit, all rates different (ARD) model (AICw 0.62 over equal rates 0.38). Our analysis reveals 6 transitions to planktivory using the extant-only tree (Fig. S12 and S15). However, when incorporating fossils in the phylogeny, our analyses identified that planktivory evolved independently at least 7 times (7–9 based on PD trees), including 1 time (1–3 for PD trees) in extinct lineages, 3 times (2–3 for PD trees) in stem clades that include fossil and extant species, and 3 times (3–4 for PD trees) in extant-only clades (Fig. 3.1A and Appendix C Figs. S13 and S16). Additionally, we observe 3 reversals to non-planktivory when fossils are excluded (Appendix C Fig. S12), and 4 (3–5 for PD trees) reversals when fossils are included (Fig. 3.1A and Appendix C Figs. S13 and S16). Notably, we observe trophic transitions spanning most of acanthuriform's history, ranging from 62.1 Ma to 3.6 Ma (Fig. 3.1A). Biogeographic analyses conducted using either 16 (Adams et al. 1983) or 12 (Steininger and Rögl 1979, 1984; Rögl 1999) Ma as the final closure of the Tethys Seaway yielded similar results (Fig. 3.2 and Appendix C Fig. S18). Based on the best-fit model, BAYAREALIKE+w (Appendix C Table S5), we initially identified that acanthuriforms originated in the Tethys Sea, with 5 lineages (from all genera, except *Ctenochaetus*) subsequently expanding into the Indo-Pacific, and one (by *Eonaso*) to the WA. From the Indo-Pacific, at least 6 dispersal events occurred towards the TEP, 1 towards the WA, and 2 towards the EA. Then 2 lineages crossed the Atlantic: (i) the ancestor of *Acanthurus chirurgus*, *A. bahianus,* and *A. tractus*, which dispersed from EA to WA; and (ii) an ancestral *Prionurus biafraensis* lineage that went from the TEP to the WA. This biogeographic history aligns with a previous study on surgeonfishes and many other circumglobal shallow marine fish groups (Baraf et al. 2019; Siqueira et al. 2019; Rincon-Sandoval et al. 2020; Santaquiteria et al. 2021). These biogeographic inferences also suggest that early transitions to planktivory (4–5 events) occurred in the Tethys Sea, while more recent transitions (3–4 events) predominantly took place in the Indo-Pacific region: 2–3 widespread in WIO, CIP and CP; and 1 in CIP and CP. No transitions were identified in the WA, EA or TEP (Fig. 3.2).

We assessed the effect of diet on acanthuriforms' diversification by running HiSSE analysis using the MCC tree with only extant species (80 tips). To incorporate uncertainties in model choice, we model-averaged rates for all tips and nodes in the trees calculated under five models using AIC weights (Appendix C Table S6). Our results indicate that both net diversification (*p*-values = 0.893 for the MCC tree and 0.001–0.995 for PD trees) and speciation rates (*p*-values = 0.893 for the MCC tree and 0.001–0.996 for PD trees) are similar between non-planktivore and planktivore species, but extinction is slightly lower in planktivores (*p*-values = 0.001 for the MCC tree and 0.001–0.895 for PD trees; Fig. 3.1C and Appendix C Table S7). Transition rates from non-planktivory to planktivory (0.010 for the MCC tree and 0.007–0.011 for PD trees) are lower than the reverse (0.026 for the MCC tree and 0.013–0.034 for PD trees; Appendix C Table S7). LTT analyses conducted on both fossil and extant species show that non-planktivore species exhibited greater diversity, characterized by increased species origination, during the aftermath of the K-Pg in the Paleocene and early Eocene (66–50 Ma) compared to planktivore species. However, during the Middle and Late Eocene (44–30 Ma), planktivore species had higher origination rates than their non-planktivore counterparts. From the Oligocene onwards (30 Ma–present), both trophic guilds resulted in an increase in speciation, with non-planktivores exhibiting a higher rate (Fig. 3.1A).

We further evaluated whether paleoclimatic temperatures influenced the diversification of planktivore species. We observe a pattern where non-planktivore lineages tend to originate during periods of temperature increase, such as in the aftermath of the K-Pg mass extinction event and the formation of the Indo-Australian Archipelago in the Miocene. We also see a slight increase in planktivore lineages when temperatures decrease after 50 Ma (Fig. 3.1B). After running the paleoclimatic model on the MCC tree, we find that both the phylogenetic signal as estimated with the lambda model (AICw=0.55 vs. 0.51) and the climate dependent model (AICw=0.44 vs. 0.48) have a better fit in both temperature curves (global average and tropical temperatures, respectively) than climate independent models like Brownian motion (AICw=0.005 vs. 0.004) and early burst (AICw=0.002 vs. 0.002, Fig. 3.1B and Appendix C Table S8). However, when we test the model over the 500 trees from the posterior distribution, the lambda model is slightly better supported than the climatic model, with global average temperature curve AICw average of 0.34 compared to 0.63 for lambda, and the tropical temperature curve at 0.15 compared to 0.82 for lambda (Fig. 3.1B and Appendix C Table S8). Finally, the same set of phylogenetic comparative analyses conducted on Scheme-2 time trees show identical or very similar results, as the placement of the †*Gazolaichthys vestenanovae* and †*Padovathurus gaudryi* fossils did not affect the outcomes of the analyses. See Appendix C Figs. S14, S17, S20, S21, and S22 and Tables S5–S7.

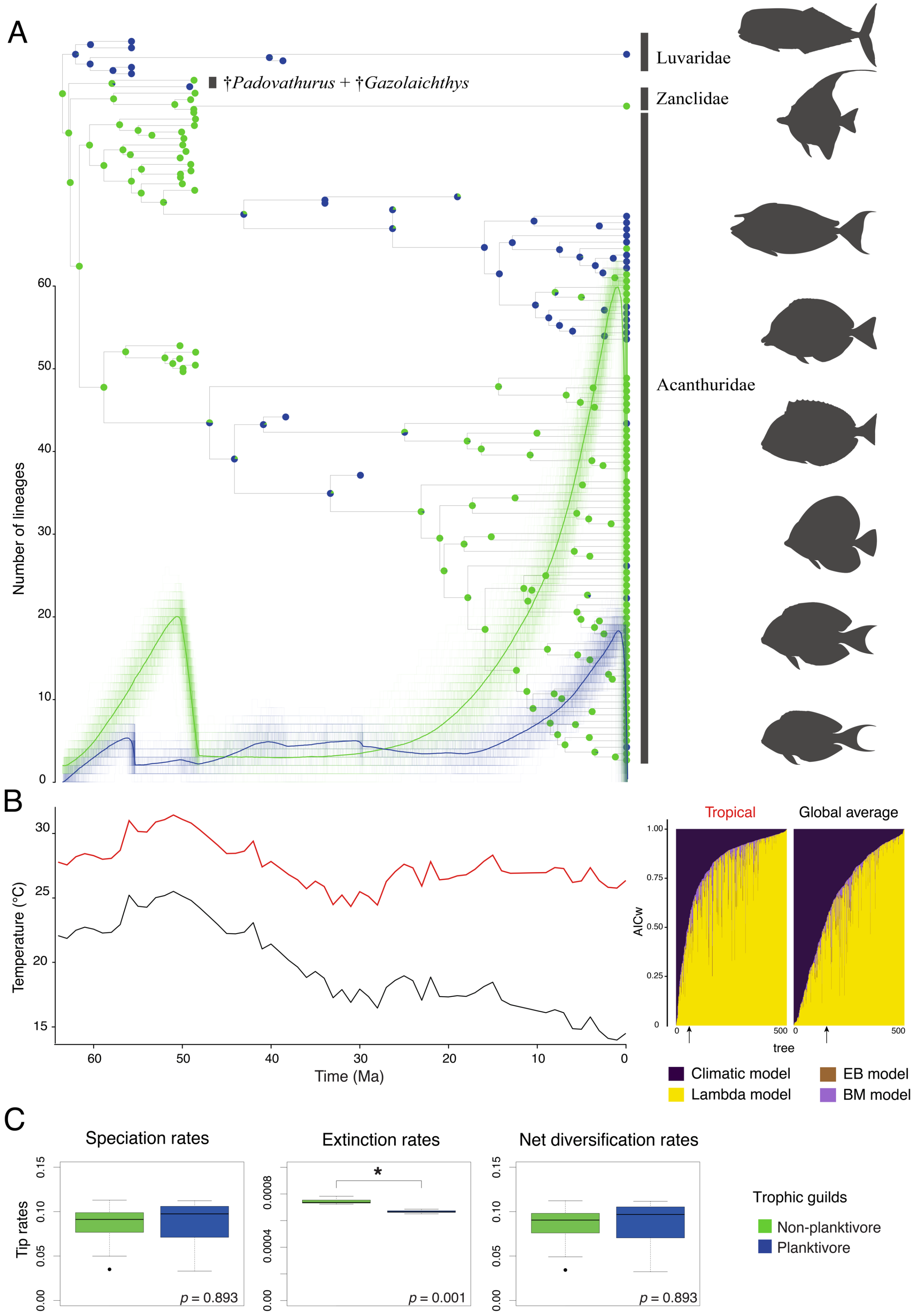

A
†Padovathurus + †Gazolaichthys
Luvaridae
Zanclidae
Acanthuridae
Number of lineages
60
50
40
30
20
10
0
B
Temperature (°C)
30
25
20
15
60
50
40
30
20
10
0
Time (Ma)
Tropical
Global average
AICw
1.00
0.75
0.50
0.25
0
0
500
0
500
tree
Climatic model
Lambda model
EB model
BM model
C
Speciation rates
Extinction rates
Net diversification rates
Tip rates
0.15
0.10
0.05
0.00
0.0008
0.0004
0.00
*
p = 0.893
p = 0.001
p = 0.893
Trophic guilds
Non-planktivore
Planktivore

Figure 3.1 Drivers of the evolution of planktivory in acanthuriforms. A) Ancestral diet reconstruction (stochastic mapping based on asymmetric rate model) and lineage through time (LTT) plot for each trophic guild based on a posterior distribution of 500 trees, evenly sampled from five independent genomic subsets (total of 999 exon alignments) from Scheme 1 trees for 112 species (32 fossil, 80 extant). The LTT plot illustrates the number of non-planktivore and planktivore lineages across time (the darker line depicts the mean of the 500 trees). Colors indicate each trophic guild. **B)** Relationship between diet and sea surface temperature for tropical and global temperature averages during the evolutionary history of acanthuriforms. AIC weights for climate dependent (tropical in red and global average temperature in black) and independent models based on the 500 alternative trees as well as the MCC tree (indicated with a black arrow). **C)** Box plots show model averaged tip rates estimated with HiSSE for speciation, extinction, and net diversification rates in extant non-planktivore and planktivore lineages based on the MCC tree. Phylogenetic ANOVA significance values are also shown at the bottom right of each panel. EB: early burst; BM: Brownian motion; EB: early burst; BM: Brownian motion; Ma: millions of years.

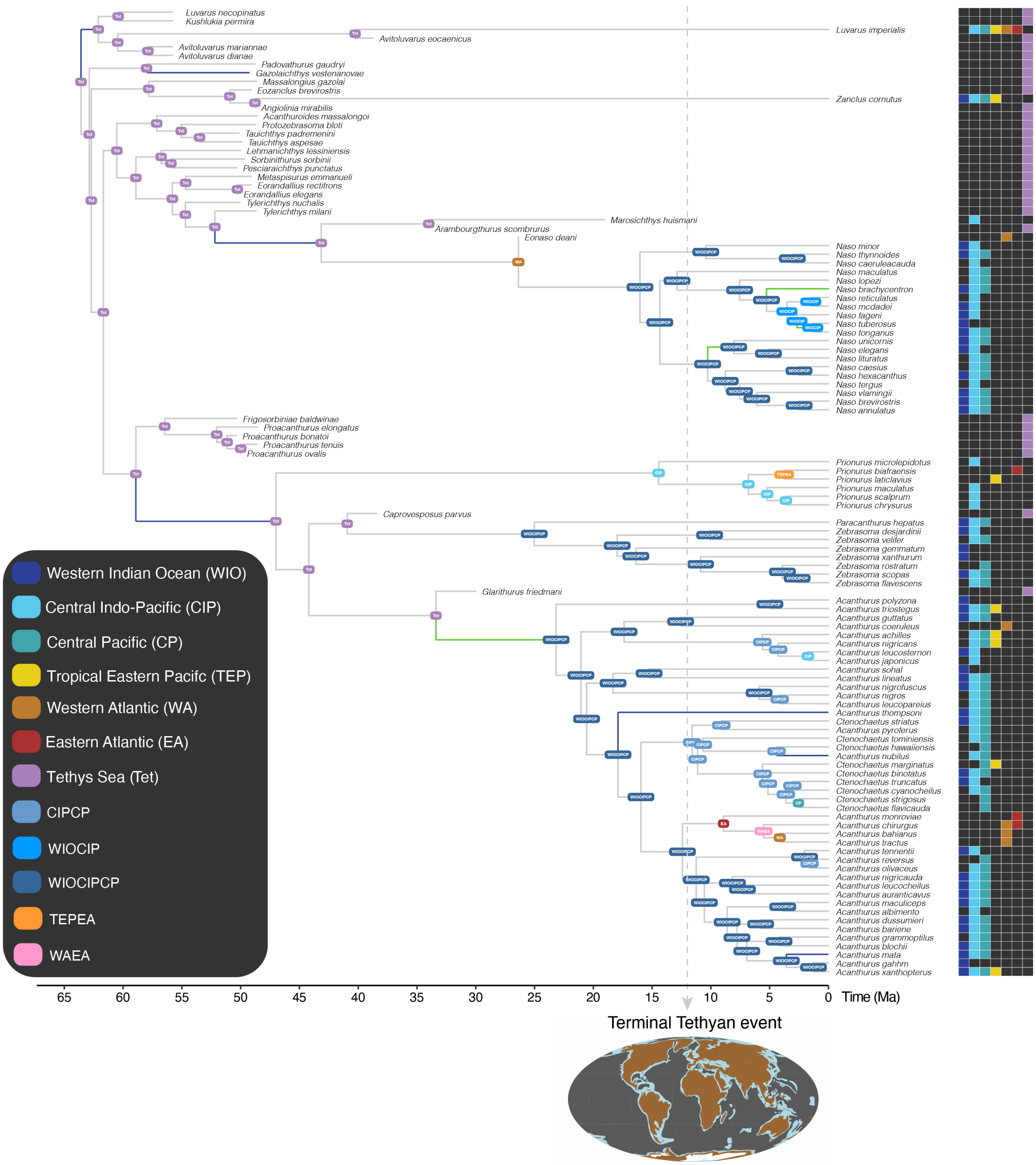


Figure 3.2 Historical biogeography of acanthuriforms and the geography of diet transitions. Ancestral range estimations in *BioGeoBEARS* using the best-supported biogeographic model (BAYAREALIKE+w) applied to 20 trees subsampled across the five subsets from Scheme 1 and using the MCC tree as fixed topology. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. Branches undergoing a transition to planktivory are indicated in blue, while those reverting to non-planktivory are shown in green. The vertical dotted line represents the Terminal Tethyan event that occurred 12 Ma. Ma: millions of years.

Signatures of positive selection associated with transitions to planktivory

We assembled and annotated a chromosome-level reference genome of *Acanthurus chirurgus* (Doctorfish Tang, Family Acanthuridae) using PacBio, Illumina, and Hi-C sequencing technologies. This resulted in a high-quality phased genome with 15 putative chromosomes for each haplotype. Haplotype 1 has a size of 776.87 Mb, with scaffold N50 of 47.48 Mb, while haplotype 2 has a genome size of 664.96 Mb, scaffold N50 of 47.66 Mb (Appendix C Fig. S22). Before annotation, the BUSCO analysis showed a completeness score of 98.8% (Fig. 3.3), which decreased to 92.2% on the annotated genome. We found that 16.39% of the genome is comprised of repetitive sequences and predicted 23548 protein-coding genes. The completeness of the 45 assembled short-read genomes ranged from 74 to 98% (Fig. 3.3), with repetitive elements ranging from 13.8% and 21.3%. We investigated the association between the percentage of repetitive elements in each trophic guild and found that neither transposable elements (p=0.989; Appendix C Fig. S23) nor any of the ten repeat element categories (p=0.178–0.975; Appendix C Fig. S24) show a significant association with trophic guilds.

We tested for positive selection associated with diet by using single-copy genes obtained from the single copy BUSCOs from the assemblies generated for the chromosome-level and short-read genomes. After excluding genes affected by sequencing or alignment error using BUSTED-E, we retained 126 high-confidence positively selected genes (PSGs). These filtered genes were subsequently examined using aBSREL (Fig. 3.3A) to identify branches that experienced transitions to planktivory, and BUSTED-PH (Fig. 3.3D) to scrutinize lineages associated with specific trophic guilds (see Materials and Methods). These analyses were conducted based on the time-calibrated MCC tree with ancestral diet reconstructions. We identified a total of 118 positively selected genes (PSGs) along the branches associated with planktivory transitions, of which 91 exhibited convergent evolution in at least two of the examined transitions to planktivory (Fig. 3.3A), most of them involved in cellular and metabolic processes. A total of eight genes were found to have convergently evolved in all five transitions. The CMTR1, TPCN1, and DUCF2428 genes appear to be linked with metabolism, NOTCH2 and PTPN13 with body size and elongation, and TRRAP, DYNC2H1, and UNC80 with both processes (Fig. 3.3E and Appendix C Table S9). In the BUSTED-PH analysis, we found 124 PSGs in planktivore species and 122 non-planktivores. While cellular process is the most common biological function associated with planktivory including 56 genes, metabolism is represented by 30 genes, biological regulation by 20 genes, and localization by 20 genes (Fig. 3.3D). We also found PSGs that are unique to individual trophic guilds: ERG28, EGR2, and SLC25A23 in planktivores and KasgA/Erm in non-planktivores (Fig. 3.3D and Appendix C Table S9). The function of these genes is mainly associated with metabolism, except for EGR2, which is involved in the elongation and size of their body (Fig. 3.3E and Appendix C Table S9). Refer to Table S9 for additional information on the biological and molecular functions of these 12 PSGs.

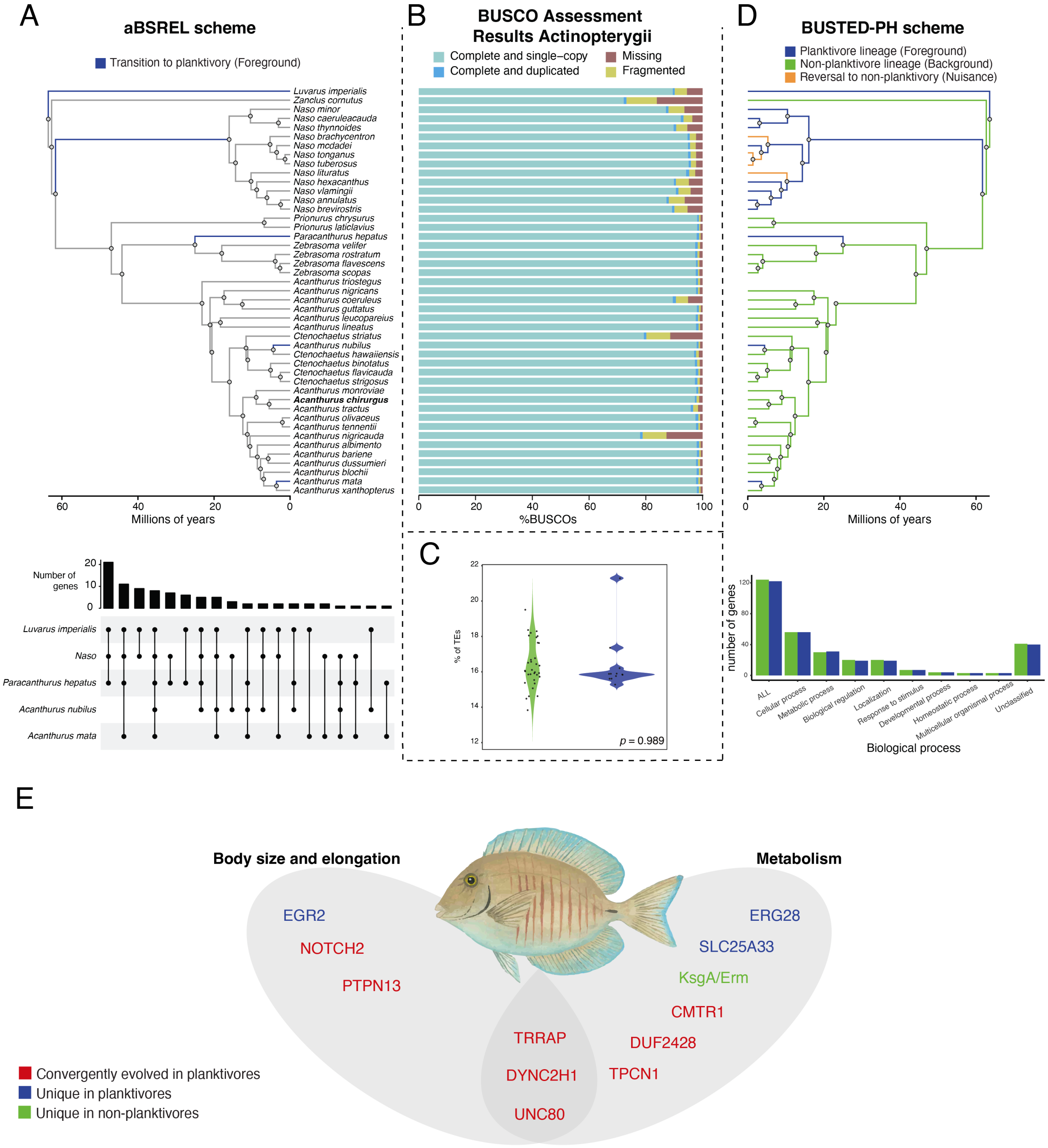

A
aBSREL scheme
Transition to planktivory (Foreground)
Luvarus imperialis
Zanclus cornutus
Naso minor
Naso caeruleacauda
Naso thynnoides
Naso brachycentron
Naso mcdadei
Naso tonganus
Naso tuberosus
Naso lituratus
Naso hexacanthus
Naso vlamingii
Naso annulatus
Naso brevirostris
Prionurus chrysurus
Prionurus laticlavius
Paracanthurus hepatus
Zebrasoma velifer
Zebrasoma rostratum
Zebrasoma flavescens
Zebrasoma scopas
Acanthurus triostegus
Acanthurus nigricans
Acanthurus coeruleus
Acanthurus guttatus
Acanthurus leucopareius
Acanthurus lineatus
Ctenochaetus striatus
Acanthurus nubilus
Ctenochaetus hawaiiensis
Ctenochaetus binotatus
Ctenochaetus flavicauda
Ctenochaetus strigosus
Acanthurus monroviae
Acanthurus chirurgus
Acanthurus tractus
Acanthurus olivaceus
Acanthurus tennentii
Acanthurus nigricauda
Acanthurus albimento
Acanthurus bariene
Acanthurus dussumieri
Acanthurus blochii
Acanthurus mata
Acanthurus xanthopterus
60
40
20
0
Millions of years
B
BUSCO Assessment Results Actinopterygii
Complete and single−copy
Complete and duplicated
Missing
Fragmented
0
20
40
60
80
100
%BUSCOs
D
BUSTED-PH scheme
Planktivore lineage (Foreground)
Non-planktivore lineage (Background)
Reversal to non-planktivory (Nuisance)
0
20
40
60
Millions of years
Number of genes
20
10
0
Luvarus imperialis
Naso
Paracanthurus hepatus
Acanthurus nubilus
Acanthurus mata
C
22
20
18
16
14
12
% of TEs
p = 0.989
number of genes
120
80
40
0
ALL
Cellular process
Metabolic process
Biological regulation
Localization
Response to stimulus
Developmental process
Homeostatic process
Multicellular organismal process
Unclassified
Biological process
E
Body size and elongation
Metabolism
EGR2
NOTCH2
PTPN13
ERG28
SLC25A33
KsgA/Erm
CMTR1
DUF2428
TPCN1
TRRAP
DYNC2H1
UNC80
Convergently evolved in planktivores
Unique in planktivores
Unique in non-planktivores

Figure 3.3 Signatures of positive selection associated with transitions to planktivory. A) aBSREL scheme used to find positively selected genes (PSGs) that evolved among lineages that convergently transitioned to planktivory, including an upset plot (below) that illustrates the number of shared PSGs across planktivore lineages. Genome completeness, as assessed by BUSCO scores, for both chromosome-level (Doctorfish tang, *Acanthurus chirurgus* in bold) and short-read (remaining species) genomes. **C)** Violin plots depict the distribution of the percentage of TEs across non-planktivore and planktivore lineages, with dots representing the raw scores for each species. Phylogenetic ANOVA significance value is shown at the bottom right of the violin plot. **D)** BUSTED-PH scheme used to find PSGs with a planktivore diet and below the number of families for each of the biological process found for PSGs. **E)** PSGs that convergently evolved in all five planktivory transitions (red), unique in planktivore species (blue), and unique in non-planktivore species (green), and their putative adaptive function. Fish illustration was retrieved from https://marinewise.com.au/ website.

3.5 Discussion

We conducted a suite of comparative analyses to assess the evolutionary dynamics of trophic transitions in fossil and extant acanthuriform fishes, examining their association with diversification rates, paleoclimatic changes, and other genomic factors. By integrating morphological and genome-wide data for both fossil and living acanthuriform species, we inferred the most comprehensive tip-dated phylogeny to date for the group, accommodating uncertainties in both topology and divergence time estimates. Based on this phylogenomic/total evidence framework, we found that planktivores evolved at least seven times, with at least four transitions occurring in extinct lineages early in the Cenozoic following the group's origin. Planktivore lineages displayed higher extinction rates compared to non-planktivores, indicating potential evolutionary constraints linked to this trophic guild. We found that both paleoclimatic temperatures and phylogenetic signal have influenced diet shifts in the group. In fact, we observe a correlation between the increased diversification of water column planktivores and cooler temperatures, and conversely, a correlation between the decreased diversification of benthic herbivores and warmer temperatures. Furthermore, our analyses of positive selection associated with transitions to planktivory based on whole genome data identified several genes undergoing convergent evolution. These genes are associated with metabolic processes and changes in body shape, potentially linked to adaptations for a planktivore diet and dwelling in the water column.

During the last 64 Ma of acanthuriform's evolutionary history, ocean temperatures have been fluctuating (Scotese et al. 2021). These temperature variations have been attributed to gradual changes in paleogeography, plate tectonics, and paleoceanographic conditions (Huber and Caballero 2011). After the K-Pg mass extinction event, acanthuriforms gradually started diversifying, particularly benthic acanthuriform lineages which fed on filamentous or macrophytic algae as well as detritus and microbes. The early Eocene period (56–48 Ma) marked a significant phase of elevated temperatures (McInerney and Wing 2011), with ocean temperatures surging by 6 °C (Scotese et al. 2021). This warm interval, called Paleocene-Eocene Thermal Maximum (PETM), coincides with the Monte Bolca formation in the Tethys Sea, a rich site of reef fish fossils. The PETM has been associated with heightened extinction in certain marine fish groups (Arcila and Tyler 2017), a phenomenon attributed to a coral reef crisis induced by ocean acidification and subsequent reductions in reef growth rates (Kiessling and Simpson 2011). During this time frame, several benthic acanthuriform lineages went extinct (Bellwood et al. 2014; Friedman and Carnevale 2018). However, for zooplankton eaters, who mostly inhabit the water column, this temperature variation did not have a significant effect. The cooling period began after the PETM (~50 Ma) when the collision of India with south-central Asia resulted in an increased influx of calcium into the world's oceans, forming limestone and reducing CO_2 levels in the atmosphere (Raymo and Ruddiman 1992). Consequently, plankton diversity (e.g., foraminifera, diatoms and radiolarians) increased (Lowery et al. 2020), potentially enabling planktivores to take advantage and diversify (Siqueira et al. 2020). It is important to consider the taphonomic bias affecting this group and other reef fishes, as evidenced by the significant number of fossil representatives from the Eocene Bolca Lagerstätten (Bellwood 1996; Friedman and Carnevale 2018). The presence of mostly non-planktivore acanthuriform fossils from Monte Bolca might lead to an overestimation of the temperature's role in extinction. It was not until the early Miocene, around ~20 Ma, when acanthuriform planktivores and reef fishes in general, experienced increases in diversification rates. This coincided with new reef configurations and diversification of corals (leading to an expansion of suitable habitat), as well as changes in water hydrodynamics of the Indo-Australian Archipelago at the Indo-Pacific Ocean (Bellwood et al. 2017; Siqueira et al. 2020, 2021, 2023; Tebbett et al. 2022). Furthermore, the constant warm temperatures in the tropics for the last ~20 Ma (Scotese et al. 2021; Steinthorsdottir et al. 2021) have increased algal turf cover, resulting in greater food availability and a higher abundance of herbivorous reef fishes (Foo et al. 2022).

The recurrent phenomenon of trophic transitions from an herbivore/detritivore diet to a planktivore one is not limited to surgeonfishes and louvars but extends to various other reef fish groups, including damselfishes, groupers, snappers, triggerfishes, and wrasses (Hobson 1991), as well as freshwater species flocks like cichlids, whitefish, and stickleback (Walker 1997; Hulsey et al. 2013; Præbel et al. 2013). While this transition is ubiquitous across Earth's oceans, in acanthuriforms it occurred primarily in the ancient Tethys Sea and Indo-Pacific regions, where the complex topography offers protection against predators and facilitates water flow in shallow reef environments that are highly productive and dynamic (Johansen et al. 2008; Siqueira et al. 2021). Although the number of transitions from non-planktivores to planktivores has exceeded reversals over the course of acanthuriforms' evolutionary history, it is noteworthy that planktivore lineages are generally younger than their non-planktivore counterparts. Consequently, younger lineages, such (e.g., *Naso*), have experienced reversals to the ancestral condition, leading to higher rates of transitions from planktivores to non-planktivores compared to the reverse direction. This occurred alongside similar diversification rates among both trophic guilds, contrary to our initial prediction based on an evolutionary dead-end perspective. Planktivory is considered the principal evolutionary destination among trophic transitions in reef fishes (Siqueira et al. 2020). However, unicornfishes, despite not further specializing their diet, can revert to their ancestral state and return to the benthic habitat, presenting an exception to this trend. The transition to a planktivore diet often coincides with a shift in habitat from the bottom to the mid-water column. It is conceivable that the reliance on a specific dietary resource, in conjunction with environmental changes or competitive pressures, may constrain the adaptive potential of planktivore species.

These repeated instances of convergence towards a planktivory diet have been associated with specialized morphological traits, including variations in reduction of the size of the jaws, teeth, and overall feeding apparatus, longer gill rakers, slender fusiform body shapes, and bifurcated tails. These changes have been observed in cichlids (Cooper et al. 2010), damselfishes (Cooper and Westneat 2009), surgeonfishes (Friedman et al. 2016), carangids (Duarte-Ribeiro et al. 2018), grunts (Tavera et al. 2018) and snappers and fusiliers (Rincon-Sandoval et al. 2020), among others. Multiple genes and environmental factors have been demonstrated to influence the genomic mechanisms underlying these complex traits (DeLorenzo et al. 2023). Our PhyloG2P analyses revealed positive selection on genes mainly associated with metabolic processes (e.g., atp-dependent activity) and anatomical structure morphogenesis (e.g., craniofacial development, dorso-ventral axis formation). While most PSGs are either found in one planktivore species or convergently evolved in more than two planktivore lineages, they are also found in non-planktivore branches. However, there are three PSGs that are found only in planktivores. One of them is the ergosterol biosynthesis 28 homolog (ERG28) gene, which is involved in cholesterol synthesis (Capell-Hattam et al. 2022) and is found only in the louvar, the earliest planktivore transition of the group. The evolutionary pressures, such as prey availability or predation risk, may be correlated with the selection of ERG28 gene, reflecting an adaptation to optimize energy metabolism and nutrient utilization. The second PSG is the solute carrier family 25 member 33 (SLC25A33), which is responsible for transport of sugars, amino acids, oligonucleotides, ions across mitochondrial membranes (Barat et al. 2019). This gene could be associated with the digestion of zooplankton and small crustaceans in planktivore species. Another PSG is the early growth response 2 (EGR2), which was identified in unicornfishes. EGR2 is implicated in cell growth and differentiation (Veremeyko et al. 2018). Rather than being linked to the planktivory lifestyle typical of most unicornfishes, it could instead underpin the distinctive skull morphology observed in these fishes. Finally, although adaptation to new environments has been shown to be accompanied by changes in repeat content (Casacuberta and González 2013; Schrader and Schmitz 2019; Bista et al. 2023; Marcionetti and Salamin 2023), we did not detect any significant differences between algal and zooplankton eaters.

While our study delved into the evolutionary dynamics of planktivory in surgeonfishes and their relatives using robust phylogenomic trees encompassing both fossil and extant species, it is important to acknowledge several caveats associated with our PhyloG2P analyses. Firstly, our analysis was constrained by the availability of data, particularly the use of approximately 3000 genes from the ray-finned fishes BUSCO database to identify genomic regions associated with planktivory transitions. The limited gene set may not capture the full spectrum of genetic variations linked to planktivory. We did not find any PSGs unique to planktivores that evolved convergently in at least two planktivore lineages. For future research, the annotated reference genome produced in this study will aid in identifying candidate genes, noncoding regions, regulatory elements, and a greater number of protein-coding genes beyond conserved BUSCO genes. By examining the gains and losses of these genomic regions, we may identify portions of the genome undergoing convergent evolution, particularly in planktivore species (e.g., genes associated with taste receptor activity; Hecker et al. 2019). To further explore the genomic basis of planktivory, skull shape obtained from CT scan images (Buser et al. 2018; Evans et al. 2021, 2023) across acanthuriforms may unveil ecomorphological correlations, which can in turn facilitate the identification of molecular coevolutionary relationships between different trophic-related traits. By elucidating the genetic basis underlying these trophic and morphological adaptations, we can gain deeper insights into the genomic basis driving the evolution of planktivory not only in surgeonfishes and their relatives, but also in other fish groups.

3.6 Conclusion

In conclusion, this study offers valuable insights into the evolutionary dynamics of trophic transitions within acanthuriform fishes, spanning from fossils to living species and encompassing examinations of both ecological and molecular factors. Our integrative phylogenetic analyses, suggest that origin of this group happened immediately after the K-Pg mass extinction event, leading to the emergence of most non-planktivore acanthuriform fossil species. Our findings reveal that the multiple independent transitions to a planktivore lifestyle in acanthuriforms do not signify an evolutionary dead end. Net diversification among planktivores in the water column is comparable to that of non-planktivore species, although extinction rates are slightly higher in non-planktivore lineages. Additionally, multiple planktivore lineages have reverted to the ancestral non-planktivore condition. Additionally, these transitions appear to be influenced by cool ocean temperatures, coral reef configuration, plankton and algae availability, and water movements. Our whole-genome analyses identified several genes showing convergent evolution, highlighting their roles in metabolic processes and body shape adaptations. Notably, three genes (ERG28, EGR2, and SLC25A33) were found under positive selection exclusively in planktivore lineages. Overall, these findings enhance our understanding of the evolutionary mechanisms driving trophic transitions in acanthuriform fishes and marine fishes more generally, shedding light on the ecological and genetic factors shaping their diversification and adaptation to this recurrent dietary shift.

3.7 References

Adams C.G., Gentry A.W., Whybrow P.J. 1983. Dating the terminal Tethyan event. Utr. Micropaleontol. Bull. 30:273–298.

Ahi E.P., Brunel M., Tsakoumis E., Chen J., Schmitz M. 2022. Appetite regulating genes in zebrafish gut; a gene expression study. PLoS One. 17:1–25.

Ahi E.P., Richter F., Sefc K.M. 2023. Gene expression patterns associated with caudal fin shape in the cichlid Lamprologus tigripictilis. Hydrobiologia. 850:2257–2273.

Alonge M., Lebeigle L., Kirsche M., Jenike K., Ou S., Aganezov S., Wang X., Lippman Z.B., Schatz M.C., Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 23:1–19.

Arcila D., Tyler J.C. 2017. Mass extinction in tetraodontiform fishes linked to the Palaeocene-Eocene thermal maximum. Proc. R. Soc. B Biol. Sci. 284.

Aristide L., Fernández R. 2023. Genomic Insights into Mollusk Terrestrialization: Parallel and Convergent Gene Family Expansions as Key Facilitators in Out-of-the-Sea Transitions. Genome Biol. Evol. 15:1–17.

Bao W., Kojima K.K., Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA. 6:4–9.

Baraf L.M., Pratchett M.S., Cowman P.F. 2019. Ancestral biogeography and ecology of marine angelfishes (F: Pomacanthidae). Mol. Phylogenet. Evol. 140:106596.

Barat A., Sahoo P.K., Kumar R., Goel C., Siva C., Ali S. 2019. Data on solute carrier transporter genes of a threatened Himalayan fish species – Schizothorax richardsonii. Data Br. 23:103712.

Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Bellwood D.R. 1996. The Eocene fishes of Monte Bolca: The earliest coral reef fish assemblage. Coral Reefs.

Bellwood D.R., Goatley C.H.R., Bellwood O. 2017. The evolution of fishes and corals on reefs: Form, function and interdependence. Biol. Rev.

Bellwood D.R., Goatley C.H.R., Brandl S.J., Bellwood O. 2014. Fifty million years of herbivory on coral reefs: Fossils, fish and functional innovations. Proc. R. Soc. B Biol. Sci. 281.

Benson G. 1999. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Betancur-R. R., Broughton R.E., Wiley E.O., Carpenter K., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton II J.C., Zhang F., Buser T., Campbell M., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Ortí G. 2013. The Tree of Life and a New Classification of Bony Fishes. PLoS Curr.:0–45.

Betancur-R. R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17.

Bista I., Wood J.M.D., Desvignes T., McCarthy S.A., Matschiner M., Ning Z., Tracey A., Torrance J., Sims Y., Chow W., Smith M., Oliver K., Haggerty L., Salzburger W., Postlethwait J.H., Howe K., Clark M.S., William Detrich H., Christina Cheng C.H., Miska E.A., Durbin R. 2023. Genomics of cold adaptations in the Antarctic notothenioid fish radiation. Nat. Commun. 14.

Broughton R.E., Betancur-R. R., Li C., Arratia G., Ortí G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr.

Burress E.D., Holcomb J.M., Tan M., Armbruster J.W. 2017. Ecological diversification associated with the benthic-to-pelagic transition by North American minnows. J. Evol. Biol. 30:549–560.

Buser T.J., Sidlauskas B.L., Summers A.P. 2018. 2D or Not 2D? Testing the Utility of 2D Vs. 3D Landmark Data in Geometric Morphometrics of the Sculpin Subfamily Oligocottinae (Pisces; Cottoidea). Anat. Rec. 301:806–818.

Capell-Hattam I.M., Fenton N.M., Coates H.W., Sharpe L.J., Brown A.J. 2022. The Non Catalytic Protein ERG28 has a Functional Role in Cholesterol Synthesis and is Coregulated Transcriptionally. J. Lipid Res. 63:100295.

Casacuberta E., González J. 2013. The impact of transposable elements in environmental adaptation. Mol. Ecol. 22:1503–1517.

Chen H.I., Turakhia Y., Bejerano G., Kingsley D.M. 2023. Whole-genome Comparisons Identify Repeated Regulatory Changes Underlying Convergent Appendage Evolution in Diverse Fish Lineages. Mol. Biol. Evol. 40:1–18.

Cheng H., Jarvis E.D., Fedrigo O., Koepfli K., Urban L., Gemmell N.J., Li H. 2022. Haplotype-resolved assembly of diploid genomes without parental data. Nat. Biotechnol.:1–13.

Clements K.D., Gray R.D., Howard Choat J. 2003. Rapid evolutionary divergences in reef fishes of the family Acanthuridae (Perciformes: Teleostei). Mol. Phylogenet. Evol.

Conway J.R., Lex A., Gehlenborg N. 2017. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics. 33:2938–2940.

Cooper W.J., Parsons K., McIntyre A., Kern B., McGee-Moore A., Albertson R.C. 2010. Bentho-pelagic divergence of cichlid feeding architecture was prodigious and consistent during multiple adaptive radiations within African Rift-Lakes. PLoS One. 5.

Cooper W.J., Westneat M.W. 2009. Form and function of damselfish skulls: Rapid and repeated evolution into a limited number of trophic niches. BMC Evol. Biol. 9:1–17.

Davis K.E., Hill J., Astrop T.I., Wills M.A. 2016. Global cooling as a driver of diversification in a major marine clade. Nat. Commun. 7:1–8.

De-Kayne R., Selz O.M., Marques D.A., Frei D., Seehausen O., Feulner P.G.D. 2022. Genomic architecture of adaptive radiation and hybridization in Alpine whitefish. Nat. Commun. 13:1–13.

DeLorenzo L., Mathews D., Brandon A.A., Joglekar M., Carmona Baez A., Moore E.C., Ciccotto P.J., Roberts N.B., Roberts R.B., Powder K.E. 2023. Genetic basis of ecologically relevant body shape variation among four genera of cichlid fishes. Mol. Ecol. 32:3975–3988.

DeSantis L.R.G. 2016. Dental microwear textures: Reconstructing diets of fossil mammals. Surf. Topogr. Metrol. Prop. 4.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol.

Duarte-Ribeiro E., Davis A.M., Rivero-Vega R.A., Ortí G., Betancur R. 2018. Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. Proc. R. Soc. B Biol. Sci. 285.

Dupin J., Matzke N.J., Särkinen T., Knapp S., Olmstead R.G., Bohs L., Smith S.D. 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeogr. 44:887–899.

Eastment R. V., Wong B.B.M., McGee M.D. 2024. Convergent genomic signatures associated with vertebrate viviparity. BMC Biol. 22:34.

Elmer K.R., Fan S., Kusche H., Luise Spreitzer M., Kautt A.F., Franchini P., Meyer A. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat. Commun. 5.

Emms D.M., Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 20:1–14.

Evans K.M., Larouche O., Gartner S.M., Faucher R.E., Dee S.G., Westneat M.W. 2023. Beaks promote rapid morphological diversification along distinct evolutionary trajectories in labrid fishes (Eupercaria: Labridae). Evolution (N. Y).:1–15.

Evans K.M., Larouche O., Watson S.J., Farina S., Habegger M.L., Friedman M. 2021. Integration drives rapid phenotypic evolution in flatfishes. Proc. Natl. Acad. Sci. U. S. A. 118:1–10.

Felsenstein J., Ackerly D.D., Mcpeek M.A. 2012. A comparative method for both discrete and continuous characters using the threshold model. Am. Nat. 179:145–156.

Flügel E., Kiessling W. 2002. Patterns of Phanerozoic Reef Crises. .

Flynn J.M., Hubley R., Goubert C., Rosen J., Clark A.G., Feschotte C., Smit A.F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117:9451–9457.

Foo S.A., Teague C.H., Asner G.P. 2022. Warming Alters the Relationship Between Benthic Cover and Herbivores on Hawaiian Reefs. Front. Mar. Sci. 9:1–12.

Friedman M., Carnevale G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. J. Geol. Soc. London.

Friedman S.T., Price S.A., Corn K.A., Larouche O., Martinez C.M., Wainwright P.C. 2020. Body shape diversification along the benthic–pelagic axis in marine fishes. Proc. R. Soc. B Biol. Sci. 287:20201053.

Friedman S.T., Price S.A., Hoey A.S., Wainwright P.C. 2016. Ecomorphological convergence in planktivorous surgeonfishes. J. Evol. Biol. 29:965–978.

Ghezelayagh A., Harrington R.C., Burress E.D., Campbell M.A., Buckner J.C., Chakrabarty P., Glass J.R., Mccraney W.T., Unmack P.J., Thacker C.E., Alfaro M.E., Friedman S.T., Ludt W.B., Cowman P.F., Friedman M., Price S.A., Dornburg A., Faircloth B.C., Wainwright P.C., Near T.J. 2022. Prolonged morphological expansion of spiny-rayed fishes following the end-Cretaceous. Nat. Ecol. Evol. 6:1211–1220.

Glazer A.M., Cleves P.A., Erickson P.A., Lam A.Y., Miller C.T. 2014. Parallel developmental genetic features underlie stickleback gill raker evolution. Evodevo. 5:1–16.

Goloboff P.A., Catalano S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. 32:221–238.

Gray J.S. 1997. Marine biodiversity: patterns, threats and conservation needs. Biodivers. Conserv. 6:153–175.

Harnik P.G., Lotze H.K., Anderson S.C., Finkel Z. V., Finnegan S., Lindberg D.R., Liow L.H., Lockwood R., McClain C.R., McGuire J.L., O'Dea A., Pandolfi J.M., Simpson C., Tittensor D.P. 2012. Extinctions in ancient and modern seas. Trends Ecol. Evol. 27:608–617.

Hecker N., Sharma V., Hiller M. 2019. Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. Proc. Natl. Acad. Sci. U. S. A. 116:3036–3041.

Hobson E.S. 1991. Trophic Relationships of Fishes Specialized to Feed on Zooplankters above Coral Reefs. Ecol. Fishes Coral Reefs (P.F. Sale, ed), Acad. Press. San Diego.:69–95.

Hodge J.R., Song Y., Wightman M.A., Milkey A., Tran B., Štajner A., Roberts A.S., Hemingson C.R., Wainwright P.C., Price S.A. 2021. Constraints on the Ecomorphological Convergence of Zooplanktivorous Butterflyfishes. Integr. Org. Biol. 3.

Holt C., Yandell M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12.

Huber M., Caballero R. 2011. The early Eocene equable climate problem revisited. Clim. Past. 7:603–633.

Hughes L.C., Ortí G., Huang Y., Sun Y., Baldwin C.C., Thompson A.W., Arcila D., Betancur-R. R., Li C., Becker L., Bellora N., Zhao X., Li X., Wang M., Fang C., Xie B., Zhou Z., Huang H., Chen S., Venkatesh B., Shi Q. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. 115:6249–6254.

Hughes L.C., Ortí G., Saad H., Li C., White W.T., Baldwin C.C., Crandall K.A., Arcila D., Betancur-R R. 2020. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. Mol. Ecol. Resour. 21:816–833.

Hulsey C.D., Roberts R.J., Loh Y.H.E., Rupp M.F., Streelman J.T. 2013. Lake Malawi cichlid evolution along a benthic/limnetic axis. Ecol. Evol. 3:2262–2272.

IUCN. 2021. The IUCN Red List of Threatened Species. Available from https://www.iucnredlist.org.

Johansen J.L., Bellwood D.R., Fulton C.J. 2008. Coral reef fishes exploit flow refuges in high-flow habitats. Mar. Ecol. Prog. Ser. 360:219–226.

Jordan G., Goldman N. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol. Biol. Evol. 29:1125–1139.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics.

Kiessling W., Simpson C. 2011. On the potential for ocean acidification to be a general cause of ancient reef crises. Glob. Chang. Biol. 17:56–67.

Klanten S.O., Van Herwerden L., Choat J.H., Blair D. 2004. Patterns of lineage diversification in the genus Naso (Acanthuridae). Mol. Phylogenet. Evol.

Kosakovsky Pond S.L., Poon A.F.Y., Velazquez R., Weaver S., Hepler N.L., Murrell B., Shank S.D., Magalis B.R., Bouvier D., Nekrutenko A., Wisotsky S., Spielman S.J., Frost S.D.W., Muse S. V. 2020. HyPhy 2.5 - A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. 37:295–299.

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

Kulbicki M., Parravicini V., Bellwood D.R., Arias-Gonzàlez E., Chabanet P., Floeter S.R., Friedlander A., McPherson J., Myers R.E., Vigliola L., Mouillot D. 2013. Global biogeography of reef fishes: A hierarchical quantitative delineation of regions. PLoS One. 8.

Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62:789–804.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Li C., Ortí G., Zhang G., Lu G. 2007. A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7:1–11.

Li D., Zhang J. 2014. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. Mol. Biol. Evol. 31:303–309.

Liu D., Wang X., Guo H., Zhang X., Zhang M., Tang W. 2021. Chromosome-level genome assembly of the endangered humphead wrasse Cheilinus undulatus: Insight into the expansion of opsin genes in fishes. Mol. Ecol. Resour. 21:2388–2406.

Losos J.B. 2011. Convergence, adaptation, and constraint. Evolution (N. Y). 65:1827–1840.

Lowery C.M., Bown P.R., Fraass A.J., Hull P.M. 2020. Ecological Response of Plankton to Environmental Change: Thresholds for Extinction. Annu. Rev. Earth Planet. Sci. 48:403–429.

Marcionetti A., Salamin N. 2023. Insights into the Genomics of Clownfish Adaptive Radiation: The Genomic Substrate of the Diversification. Genome Biol. Evol. 15:1–16.

Marramà G., Garbelli C., Carnevale G. 2016. A clade-level morphospace for the eocene fishes of Bolca: Patterns and relationships with modern tropical shallow marine assemblages. Boll. della Soc. Paleontol. Ital. 55:139–156.

Matzke N.J. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Packag. version 0.2.

Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

Matzke N.J. 2019. BioGeoBEARS - Run BioGeoBEARS on multiple trees. Available from https://github.com/nmatzke/BioGeoBEARS/blob/master/R/BioGeoBEARS_on_multiple_trees_v1.R.

McInerney F.A., Wing S.L. 2011. The paleocene-eocene thermal maximum: A perturbation of carbon cycle, climate, and biosphere with implications for the future. Annu. Rev. Earth Planet. Sci. 39:489–516.

Melendez-Vazquez F., Lucaci A., Selberg A., Clavel J., Rincon-Sandoval M., Santaquiteria A., Betancur-R R., Arcila D. From Cold to Warm: Evolution of Endothermy in Ray-Finned Fishes. .

Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.

Musilova Z., Cortesi F., Matschiner M., Davies W.I.L., Patel J.S., Stieb S.M., De Busserolles F., Malmstrøm M., Tørresen O.K., Brown C.J., Mountford J.K., Hanel R., Stenkamp D.L., Jakobsen K.S., Carleton K.L., Jentoft S., Marshall J., Salzburger W. 2019. Vision using multiple distinct rod opsins in deep-sea fishes. Science (80-.). 364:588–592.

O'Dea, Aaron, Lessios H.A., Coates A.G., Eytan R.I., Restrepo-Moreno S.A., Cione A.L., Collins L.S., de Queiroz A., Farris D.W., Norris R.D., Stallard R.F., Woodburne M.O., Aguilera O., Aubry M.-P., Berggren W.A., Budd A.F., Cozzuol M.A., Coppard S.E., Duque-Caro H., Finnegan S., Gasparini G.M., Grossman E.L., Johnson K.G., Keigwin L.D., Knowlton N., Leigh E.G., Leonard-Pingel J.S., Marko P.B., Pyenson N.D., Rachello-Dolmen P.G., Soibelzon E., Soibelzon L., Todd J.A., Vermeij G.J., Jackson J.B.C. 2016. Formation of the Isthmus of Panama. Sci. Adv. 2:1–12.

OBIS. 2021. Data from the Ocean Biogeographic Information System. Intergovernmental Oceanographic Commission of UNESCO. .

Parker J., Tsagkogeorga G., Cotton J.A., Liu Y., Provero P., Stupka E., Rossiter S.J. 2013. Genome-wide signatures of convergent evolution in echolocating mammals. Nature. 502:228–231.

Præbel K., Knudsen R., Siwertsson A., Karhunen M., Kahilainen K.K., Ovaskainen O., Østbye K., Peruzzi S., Fevolden S.E., Amundsen P.A. 2013. Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. Ecol. Evol. 3:4970–4986.

Purnell M., Seehausen O., Galis F. 2012. Quantitative three-dimensional microtextural analyses of tooth wear as a tool for dietary discrimination in fishes. J. R. Soc. Interface. 9:2225–2233.

Putnam N.H., Connell B.O., Stites J.C., Rice B.J., Hartley P.D., Sugnet C.W., Haussler D., Rokhsar D.S. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26:342–350.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol.

Ranwez V., Douzery E.J.P., Cambon C., Chantret N., Delsuc F. 2018. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol.

Raymo M.E., Ruddiman W.F. 1992. Tectonic Forcing of Late Cenozoic Climate. Nature. 359:117–122.

Reddin C.J., Kocsis Á.T., Kiessling W. 2019. Climate change and the latitudinal selectivity of ancient marine extinctions. Paleobiology. 45:70–84.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol.

Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero A., Walker Jr. H.J., Carpenter K.E., Sheaves M., Ortí G., Arcila D., Betancur-R. R. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. Proc. Natl. Acad. Sci. U. S. A.

Rögl F. 1999. Palaeogeographic considerations for Mediterranean and Paratethys seaways. Ann. Naturhist. Mus. Wien. 99A:279–310.

Ronquist F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol.

Roycroft E., Achmadi A., Callahan C.M., Esselstyn J.A., Good J.M., Moussalli A., Rowe K.C. 2021. Molecular Evolution of Ecological Specialisation: Genomic Insights from the Diversification of Murine Rodents. Genome Biol. Evol. 13:1–16.

Rundle H.D., Nagel L., Boughman J.W., Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science (80-.). 287:306–308.

Santaquiteria A., Siqueira A.C., Duarte-Ribeiro E., Carnevale G., White W.T., Pogonoski J.J., Baldwin C.C., Ortí G., Arcila D., Ricardo B.R. 2021. Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences. Syst. Biol. 70:1145–1162.

Schrader L., Schmitz J. 2019. The impact of transposable elements in adaptive evolution. Mol. Ecol. 28:1537–1549.

Scotese C.R., Song H., Mills B.J.W., van der Meer D.G. 2021. Phanerozoic paleotemperatures: The earth's changing climate during the last 540 million years. Earth-Science Rev. 215:103503.

Seppey M., Ioannidis P., Emerson B.C., Pitteloud C., Robinson-Rechavi M., Roux J., Escalona H.E., McKenna D.D., Misof B., Shin S., Zhou X., Waterhouse R.M., Alvarez N. 2019. Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. Genome Biol. 20:1–14.

Shapiro M.D., Summers B.R., Balabhadra S., Aldenhoven J.T., Miller A.L., Cunningham C.B., Bell M.A., Kingsley D.M. 2009. The Genetic Architecture of Skeletal Convergence and Sex Determination in Ninespine Sticklebacks. Curr. Biol. 19:1140–1145.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E. V., Zdobnov E.M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212.

Singh P., Irisarri I., Torres-Dowdall J., Thallinger G.G., Svardal H., Lemmon E.M., Lemmon A.R., Koblmüller S., Meyer A., Sturmbauer C. 2022. Phylogenomics of trophically diverse cichlids disentangles processes driving adaptive radiation and repeated trophic transitions. Ecol. Evol. 12:1–15.

Siqueira A.C., Bellwood D.R., Cowman P.F. 2019. Historical biogeography of herbivorous coral reef fishes: The formation of an Atlantic fauna. J. Biogeogr. 46:1611–1624.

Siqueira A.C., Morais R.A., Bellwood D.R., Cowman P.F. 2020. Trophic innovations fuel reef fish diversification. Nat. Commun. 11:1–11.

Siqueira A.C., Morais R.A., Bellwood D.R., Cowman P.F. 2021. Planktivores as trophic drivers of global coral reef fish diversity patterns. Proc. Natl. Acad. Sci. U. S. A. 118.

Siqueira A.C., Muruga P., Bellwood D.R. 2023. On the evolution of fish–coral interactions. Ecol. Lett. 26:1348–1358.

Smit A.F.A., Hubley R., Green P. RepeatMasker Open-4.0. http://www.repeatmasker.org. .

Smith M.D., Wertheim J.O., Weaver S., Murrell B., Scheffler K., Kosakovsky Pond S.L. 2015. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. 32:1342–1353.

Sorenson L., Santini F., Carnevale G., Alfaro M.E. 2013a. A multi-locus timetree of surgeonfishes (Acanthuridae, Percomorpha), with revised family taxonomy. Mol. Phylogenet. Evol. 68:150–160.

Sorenson L., Santini F., Carnevale G., Alfaro M.E. 2013b. A multi-locus timetree of surgeonfishes (Acanthuridae, Percomorpha), with revised family taxonomy. Mol. Phylogenet. Evol.

Spalding M.D., Fox H.E., Allen G.R., Davidson N., Ferdaña Z.A., Finlayson M., Halpern B.S., Jorge M.A., Lombana A., Lourie S.A., Martin K.D., McManus E., Molnar J., Recchia C.A., Robertson J. 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. Bioscience. 57:573–583.

Spond S. 2022. BUSTED-PH. .

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.

Stanke M., Keller O., Gunduz I., Hayes A., Waack S., Morgenstern B. 2006. AUGUSTUS: A b initio prediction of alternative transcripts. Nucleic Acids Res. 34:435–439.

Steininger F., Rögl F. 1979. The paratethys history. A contribution towards the Neogene geodynamics of the alpine orogene. Ann. Geol. des Pays Hell. 3:1153–1165.

Steininger F.F., Rögl F. 1984. Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. Geol. Soc. Spec. Publ. 17:659–668.

Steinthorsdottir M., Coxall H.K., de Boer A.M., Huber M., Barbolini N., Bradshaw C.D., Burls N.J., Feakins S.J., Gasson E., Henderiks J., Holbourn A.E., Kiel S., Kohn M.J., Knorr G., Kürschner W.M., Lear C.H., Liebrand D., Lunt D.J., Mörs T., Pearson P.N., Pound M.J., Stoll H., Strömberg C.A.E. 2021. The Miocene: The Future of the Past. Paleoceanogr. Paleoclimatology. 36.

Tavera J., Acero P. A., Wainwright P.C. 2018. Multilocus phylogeny, divergence times, and a major role for the benthic-to-pelagic axis in the diversification of grunts (Haemulidae). Mol. Phylogenet. Evol. 121:212–223.

Tea Y.-K., Zhou Y., Ewart K.M., Cheng G., Kawasaki K., Dibattista J.D., Ho S.Y.W., Lo N., Fan S. 2024. The spotted parrotfish genome provides insights into the evolution of a coral reef dietary specialist (Teleostei : Labridae : Scarini : Cetoscarus ocellatus) . :1–16.

Tebbett S.B., Siqueira A.C., Bellwood D.R. 2022. The functional roles of surgeonfishes on coral reefs : past , present and future. Springer International Publishing.

Thomas P.D., Ebert D., Muruganujan A., Mushayahama T., Albou L.P., Mi H. 2022. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 31:8–22.

Tyler J.C. 1970. Osteological Aspects of Interrelationships of Surgeon Fish Genera (Acanthuridae) Author (s): James C . Tyler Source : Proceedings of the Academy of Natural Sciences of Philadelphia , 1970 , Vol . 122 Published by : Academy of Natural Sciences Stable U. 122:87–124.

Van Valkenburgh B., Want X., Damuth J. 2010. Cope's Rule, hypercarnivory, and extinction in North American canids. Science (80-.). 306:101–104.

Veremeyko T., Yung A.W.Y., Anthony D.C., Strekalova T., Ponomarev E.D. 2018. Early growth response gene-2 is essential for M1 and M2 macrophage activation and plasticity by modulation of the transcription factor CEBPβ. Front. Immunol. 9:1–23.

Vizueta J., Macías-Hernández N., Arnedo M.A., Rozas J., Sánchez-Gracia A. 2019. Chance and predictability in evolution: The genomic basis of convergent dietary specializations in an adaptive radiation. Mol. Ecol. 28:4028–4045.

Walker J.A. 1997. Ecological morphology of lacustrine threespine stickleback Gasterosteus aculeatus L. (Gasterosteidae) body shape. Biol. J. Linn. Soc. 61:3–50.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer- Verlag, New York.

Winterbottom R. 1993. Myological Evidence for the Phylogeny of Recent Genera of Surgeonfishes (Percomorpha , Acanthuridae), with Comments on the Acanthuroidei. Am. Soc. Ichthyol. Herpetol. 1993:21–39.

Zhang C., Sayyari E., Mirarab S. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 10562 LNBI:53–75.

Zhang S., Song Y., Liu M., Yuan Z., Zhang M., Zhang H., Seim I., Fan G., Liu S., Liu X. 2023. Chromosome-level genome of butterflyfish unveils genomic features of unique colour patterns and morphological traits. DNA Res. 30:1–10.

Zimin A. V., Marçais G., Puiu D., Roberts M., Salzberg S.L., Yorke J.A. 2013. The MaSuRCA genome assembler. Bioinformatics. 29:2669–2677.

Appendix A

Supplementary Material for Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences

Data repository

All data and code are available in the Dryad digital repository: https://doi.org/10.5061/dryad.4xgxd2580.

Supplementary Materials and Methods

Data assembly, alignment, and quality control

We used the PHYLUCE package pipeline (Faircloth 2016) to process the raw sequence reads (http://phyluce.readthedocs.io/en/latest/tutorial-one.html). We first demultiplexed and trimmed fastq files to remove adapter contamination and low-quality bases using Trimmomatic v0.36 (Bolger et al. 2014), as implemented in illumiprocessor v2.0 (Faircloth 2016). We initially removed three species due to poor sequence quality and assembled retained reads into contigs (assemblo_trinity.py) using Trinity v1.5.0 (Grabherr et al. 2011). We then matched assembled contigs to a FASTA file with the bait set used for enrichment (fish-uce-1k-probes.fasta) using match_contigs_to_probes.py and extracted targeted UCE loci shared among taxa using get_fastas_from_match_counts.py. This step generated a 'monolithic' FASTA file with all UCEs for the 77 newly sequenced individuals (61 species in 9 families). After this step, we added to the 'monolithic' file UCE data for the 113 species (112 syngnatharians and the scombroid outgroup, *Taractichthys longipinnis*) sequenced by the previous study (Longo et al. 2017). We then 'exploded' this file to obtain individual UCE loci (explode_get_fastas_file.py) and aligned them using MAFFT (Katoh and Standley 2013) based on a maximum divergence of 0.2. We trimmed the resulting alignments using Gblocks v0.91b (get_gblocks_trimmed_alignments_from_untrimmed.py; Castresana 2000) to remove ambiguously-aligned flanking regions.

Initial assessments of phylogenetic relationships were conducted using FastTree-2 (Price et al. 2010) to identify possible cases of contamination and misidentification. After this step, we removed a total of six species from downstream phylogenomic analyses (see main text). For each specimen, we also mined the CO1 sequences from our raw data and from that of Longo et al. (2017). We then mapped the CO1 sequences against the Barcode of Life Database (BOLD) and National Center for Biotechnology Information (NCBI) databases for verification of species identifications. We detected one contaminated and five mis-identified specimens (two from our dataset and four from Longo et al. 2017) and updated their identity. These are: *Notopogon lilliei* CSIROGT7994 (the sample for *N. xenosoma* CSIROGT7588 became contaminated with this sample, resulting in two sets of sequences for *N. lilliei* CSIROGT7994; this was evident after comparing the original CO1 sequence produced by CSIRO and the CO1 sequence we extracted using the UCE data generated in this study), *Callionymus bairdi* UPRFL0500 (formerly identified as *Diplogrammus pauciradiatus*), *Fistularia petimba* CEO87 (formerly identified as *F. tabacaria*), *Parupeneus multifasciatus* CEO31 (formerly identified as *P. trifasciatus*), *Halicampus crinitus* CEO60 (formerly identified as *Cosmocampus elucens*), and *Hippocampus barbouri* CEO118 (formerly identified as *H. histrix*).

Fossil calibrations

We based lower bounds for each of the calibration schemes on minimum age constraints (i.e., youngest fossil ages; Parham et al. 2012); soft upper bounds used mostly fossils placed deeper in the tree. With the exception of crown Syngnatharia, for which we used a Cauchy distribution in MCMCTree, all other calibrations used uniform distributions (Table S2). Trees for some of the subsets (see below) had clades that were incongruent with the relationships resolved using the complete datasets (e.g., non-monophyletic Aulostomoidea in astral_S04 tree). We thus excluded those incongruent calibrations on a case-by-case basis.

(1) Root. MRCA: *Syngnathus louisianae*, *Taractichthys longipinnis*. To avoid artifactual root age estimations (i.e., excessively old or young ages), we applied a secondary calibration on the root that is based on ages obtained from independent, large-scale fish trees that applied multiple fossil percomorph calibrations, both within and outside Syngnatharia. Hard lower bound: 92 Ma; hard upper bound: 103.5 Ma (see details on Table S1). Prior setting MCMCTree: B(0.920,1.035,1e-300,1e-300) (crown calibration). While fossils cannot establish hard upper bounds, some of the oldest acanthomorph fossils provide evidence that the maximum age of the root (crown Syngnatharia + Pelagiaria) could be in this age range. These are the Albian/Cenomanian (~100 Ma) †Pseudomonocentrididae trachichthyoids, †*Handuichthys interopercularis* and †*Pseudomonocentris microspinosus*, from Muhi Quarry, El Doctor Formation, State of Hidalgo, Mexico (González-Rodriguez et al. 2013), and the plectocretacicoid, †*Plectocretacicus clarae*, from the Cenomanian (~96.9–95.0 Ma) of Lebanon (Forey et al. 2003).

(2) Syngnatharia. MRCA: *Syngnathus louisianae*, *Eurypegasus draconis*. Hard lower bound: †*Gasterorhamphosus zuppichinii* (Sorbini 1981). Diagnosis and phylogenetic placement: placement of †*Gasterorhamphosus zuppichini* in Syngnathiformes is supported by the following character states: absence of the anal-fin spine, enlarged dorsal-fin spine with serrated posterior margin, elongated tubular snout, ribs absent, cleithrum bears enlarged posterodorsal process, coracoid with rod-like anteroventral process and pectoral-fin rays simple (Pietsch 1978; Orr 1995). Stratigraphic horizon and locality: Upper Cretaceous "Calcari di Melissano", Porto Selvaggio, Lecce province, Italy (Sorbini 1981). Paleogeographic domain: western-central Tethys. Absolute age estimate: 83.6 Ma (lower Campanian; see comments below). Prior setting MCMCTree: L(0.836,0.01,0.0001,1e-300) (crown calibration). Comments: the age of †*G. zuppichinii* is somewhat problematic. Most time-calibrated fish trees previously estimated (Near et al. 2012; Betancur-R. et al. 2017; Alfaro et al. 2018; Hughes et al. 2018) have used the conservative age of 72.1 Ma, corresponding to the base of the Maastrichtian age. This age was suggested in a short note by L. Sorbini (1981) based on a single analysis of the calcareous nannoplankton content on a single sample from the Canale site of the Nardò locality. However, the †*G. zuppichinii* specimen was collected at the Porto Selvaggio site, and various sites of the Nardò locality clearly have different ages. In general, the age of the Calcari di Melissano formation in the Nardò area seems to extend up to the lower Campanian (about 83 Ma; Schlüter et al., 2008). Moreover, an unpublished †*Gasterorhamphosus* specimen collected by G. Carnevale from another site is in a position stratigraphically lower than that of Canale, for which preliminary biostratigraphic analyses seem to suggest a lower Campanian age. We thus use a lower Campanian age for †*G. zuppichinii*. The upper bound of the calibration was initially chosen based on secondary ages obtained by previous studies (e.g., 80 Ma; see Table S1). The new evidence for the age of †*Gasterorhamphosus* implies that all previous studies underestimated the crown age of Syngnatharia. We used a lower bound calibration density (Cauchy distribution) with the parameters indicated above, chosen so that the minimum age of the clade is closer in time to its oldest fossil (dos Reis and Yang 2017). Preliminary tests using other constraints (e.g., an upper bound of 92 Ma, reflecting the mean value of stem means from previous studies; Table S1; Fig. S1) resulted in an extremely short Syngnatharia stem (~5 Ma), which is incongruent with patterns observed in unconstrained phylograms and previous time calibrated studies (Alfaro et al. 2018; Hughes et al. 2018), where the relative length of the stem branch is substantially longer (>10 Ma; see Fig. S1). Other tests using wider calibration densities also failed to capture the relative length of the stem branch, also producing overestimated crown ages (e.g., 93.5 Ma).

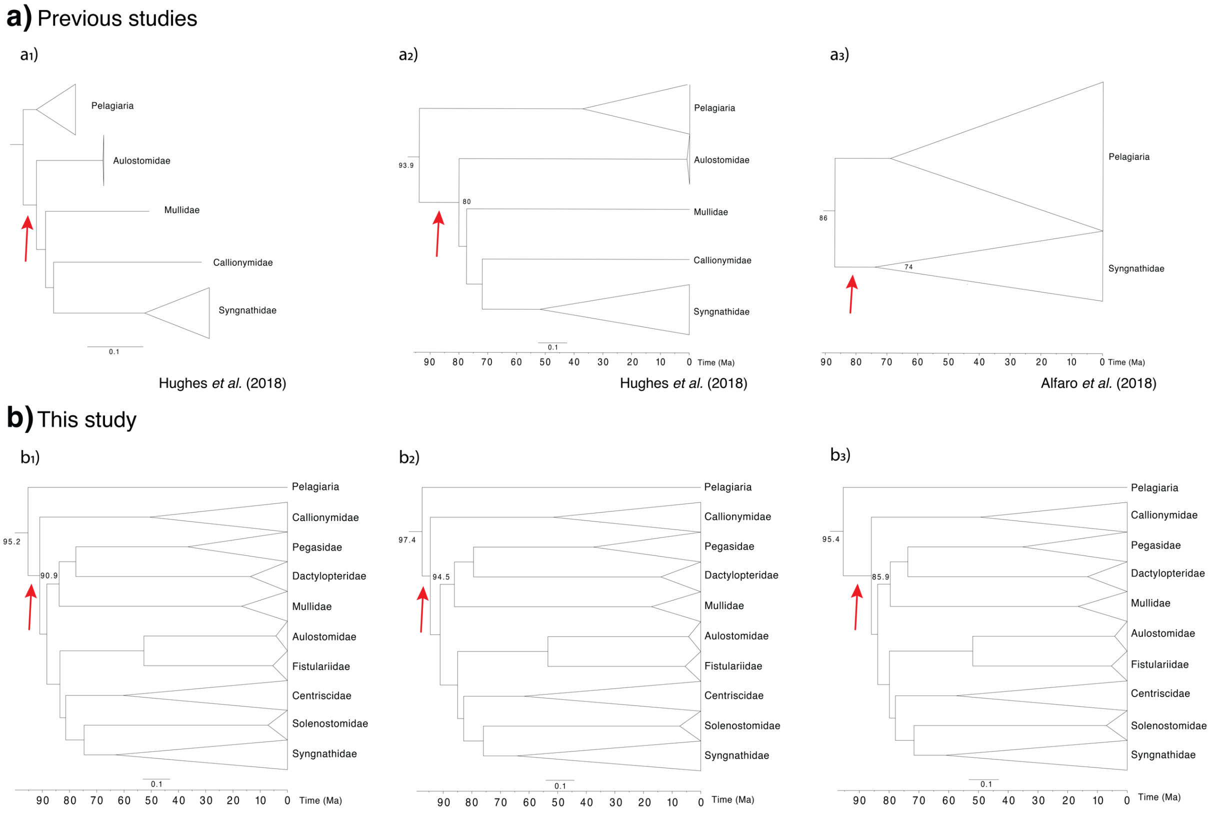


Figure S1. Justification for the calibration selected for crown Syngnatharia (see text above). **a)** Trees estimated by previous studies: a_1) maximum likelihood phylogram of (Hughes et al. 2018), a_2) time-calibrated phylogeny of Hughes et al. (2018), and a_3) time-calibrated phylogeny of (Alfaro et al. 2018). **b)** Time-calibrated phylogenies inferred in this study using a subset of ~90 genes, with alternative parameters used in MCMCTree that include a hard lower bound of 83.6 Ma and: b_1) a 92.4 Ma upper bound in a uniform distribution (B(0.836,0.924,1e-300,1e-300)), b_2) a wide density Cauchy distribution (L(0.836,0.1,1,1e-300)), and b_3) a concentrated density Cauchy distribution (L(0.836,0.01,0.0001,1e-300)) calibration selected. The red arrow indicates the stem of Syngnatharia, with b_1 and b_2 showing extremely short stem lengths.

(3a) Solenostomidae. MRCA: *Solenostomus cyanopterus*, *Syngnathus louisianae*. Hard lower bound: †*Solenorhynchus elegans* (Heckel 1854). Diagnosis and phylogenetic placement: †*Solenorhynchus elegans* exhibits some character states that place this fossil in the family

Solenostomidae: considerably elongated and slender body, body covered with stellate bony plates, head relatively small, two separate dorsal fins (each on a raised base), anal fin opposite to the soft dorsal fin, pelvic fins relatively large and elongate caudal fin (Bannikov and Carnevale 2017). Stratigraphic horizon and locality: Early Eocene, Upper Ypresian (Friedman and Carnevale 2018), Monte Bolca, Italy (Bannikov and Carnevale 2017). Paleogeographic domain: western Tethys. The †*Solenorhynchus elegans* fossil was found in the Monte Postale site of the Monte Bolca locality (Bannikov and Carnevale 2017), but whether it is found also from Pesciaria site is unclear. Stratigraphy and age of the Pesciaria and Monte Postale sites are detailed in Papazzoni et al. (2017). Absolute age estimate: 48.5 Ma (Friedman & Carnevale, 2018; see comment below). Soft upper bound: 83.6 Ma (based on †*G. zuppichinii*; see above). Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05). Comment: the age constraint of Monte Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 was used for this calibration. This calibration and calibration 3b below are redundant. While this should be treated as a stem calibration (with MRCA *Solenostomus cyanopterus*, *Solenostomus paradoxus*), it is instead applied as crown calibration one node below due to limitations of MCMCTree.

(3b) Syngnathidae. MRCA: *Syngnathus louisianae*, *Solenostomus cyanopterus*. Hard lower bound: †*Prosolenostomus lessinii* (Blot 1980). Diagnosis and phylogenetic placement: the placement of *†Prosolenostomus lessinii* in the family Syngnathidae is supported by a strong elongated body, completely covered by armored plates; short tube-shaped snout in a small head; and apparent absence or reduced dorsal, anal and pelvic fins (Orr 1995; Bannikov 2014; Carnevale et al. 2014; Bannikov and Carnevale 2017). Stratigraphic horizon and locality: Early Eocene, Upper Ypresian, Monte Bolca, Italy (Blot 1980). Paleogeographic domain: western Tethys. Absolute age estimate: 48.5 Ma (Friedman & Carnevale, 2018; see comment below). Soft upper bound: 83.6 Ma (based on †*G. zuppichinii*; see above). Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05). Comment: the age constraint of Monte Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 was used for this calibration. This calibration and calibration 3a above are redundant. While this should be treated as a stem calibration (with MRCA *Syngnathus louisianae*, *Entelurus aequoreus*), it is instead applied as crown calibration one node below due to limitations of MCMCTree.

(4) *Hippocampus*. MRCA: *Hippocampus abdominalis*, *Hippocampus kuda*. Hard lower bound: †*Hippocampus sarmaticus* sp. nov. (Žalohar et al., 2009). Diagnosis and phylogenetic placement: †*Hippocampus sarmaticus* is a fully developed seahorse species that does not differ considerably from other extant species of *Hippocampus* (Žalohar et al., 2009). Stratigraphic horizon and locality: Coprolitic horizon, Lower Sarmatian, Middle Miocene (Horvat 2003), Tunjice Hills (southwestern margins of the Central Paratethys), Slovenia (Žalohar et al., 2009). Paleogeographic domain: Pannonian Basin, Central Paratethys. Absolute age estimate: 11.6 Ma (Horvat 2003). Soft upper bound: 48.5 Ma (based on *†Prosolenostomus lessinii*; see above). Prior setting MCMCTree: B(0.116,0.485,1e-300,0.05) (crown calibration).

(5) Aulostomoidea. MRCA: *Aulostomus maculatus*, *Aeoliscus strigatus*. Hard lower bound: †*Eekaulostomus cuevasae* (Cantalice and Alvarado-Ortega 2016). Diagnosis and phylogenetic placement (Cantalice and Alvarado-Ortega 2016): the placement of †*Eekaulostomus cuevasae* as sister group of Aulostomidae and Fistulariidae is supported by a rigid stellate scutes covering the whole body and part of the snout, pelvic-fin insertion well advanced, just behind the postcleithrum, and a relative large number of principal caudal-fin rays and presence of two thin and elongated spines in the dorsal and anal fins (Nelson et al., 2016; Wheeler, 1955). Stratigraphic horizon and locality: Early Palaeocene, Danian (Cuevas-García and Alvarado-Ortega 2009), marine deposits of Belisario Domínguez, Chiapas, southeastern Mexico (Cantalice and Alvarado-Ortega 2016). Paleogeographic domain: Gulf of Mexico/Proto Caribbean Sea. Absolute age estimate: 61.5 Ma (63.0 ±1.5 Ma as estimated by Cuevas-García & Alvarado-Ortega, 2009). Soft upper bound: 83.6 Ma (based on †*G. zuppichinii*; see above). Prior setting MCMCTree: B(0.615,0.836,1e-300,0.05). Comments: while this should be treated as a stem calibration (with MRCA *Aulostomus maculatus*, *Fistularia petimba*), it is instead applied as crown calibration one node below due to limitations of MCMCTree. Another redundant calibration that can be applied to this node consists of †*Gerpegezhus paviai* (family Gerpegezhidae, the sister group of the extant Centriscidae) with hard lower bound of 55.8 Ma (Bannikov & Carnevale, 2012; Gavrilov et al., 2003). †*Gerpegezhus paviai* is, however, younger than †*Eekaulostomus cuevasae* and therefore †*Gerpegezhus* is excluded from consideration.

(6) Fistulariidae. MRCA: *Fistularia corneta*, *Aulostomus maculatus*. Hard lower bound: †*Urosphen dubius* (Blainville 1818). Diagnosis and phylogenetic placement (Orr 1995): the placement of this fossil as the putative sister group of the Fistulariidae is supported by a naked and cylindrical body and toothed mouth; four anterior vertebrae elongate and consolidated into a rigid structure; ribs absent; distinctive caudal fin similar to the condition seen in extant fistulariid species; and middle, dorsal and ventral rays more elongated than in extant species. Stratigraphic horizon and locality: Early Eocene, Upper Ypresian, Monte Bolca, Italy (Bannikov 2014). Paleogeographic domain: western Tethys. Absolute age estimate: 48.5 Ma (Friedman & Carnevale, 2018; see comment below). Soft upper bound: 61.5 Ma (based on †*Eekaulostomus cuevasae*; see above). Prior setting MCMCTree: B(0.485,0.615,1e-300,0.05). Comment: the age constraint of Monte Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 was used for this calibration. While this should be treated as a stem calibration (with MRCA *Fistularia corneta*, *Fistularia petimba*), it is instead applied as crown calibration one node below due to limitations of MCMCTree.

(7) Pegasidae. MRCA: *Pegasus volitans, Dactylopterus volitans*. Hard lower bound: †*Ramphosus rastrum* (Volta 1796). Diagnosis and phylogenetic placement: †*Ramphosus rastrum* (family †Ramphosidae) shares the following character states with smembers of Pegasidae: nasals elongate and fused along the midline and forming a rostrum bearing one to several series of toothlike denticles; olfactory capsule bordered by nasal and lateral ethmoid; three infraorbital bones, of which the second and third joined to the preopercle; mouth inferior; premaxilla without ascending process, linked to maxilla by a mobile, (possibly) neomorphic element (ligamental ossification); maxilla linked to vomer by enlarged, mobile articular cartilage (maxillovomerine cartilage); suspensorium strongly directed anteriorly; anteriormost vertebral centra elongate; soft dorsal and anal fins placed posteriorly on body, containing few rays, subequal in length; and head enclosed in bony plates (Pietsch 1978). Stratigraphic horizon and locality: Early Eocene, Upper Ypresian (Friedman and Carnevale 2018), Monte Bolca, Italy (Pietsch 1978). Paleogeographic domain: western Tethys. Absolute age estimate: 48.5 Ma (Friedman & Carnevale, 2018; see comment below). Soft upper bound: 83.6 Ma (based on †*G. zuppichinii*; see above). Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05). Comment: the age constraint of Monte Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 was used for this calibration. While this should be treated as a stem calibration (with MRCA *Eurypegasus draconis*, *Pegasus volitans*), it is instead applied as crown calibration one node below due to limitations of MCMCTree.

Geologic calibrations based on trans-isthmian geminate taxa

Several geminate species pairs in Syngnatharia–including terminal clades on either side of the Panama Isthmus (Jordan and Evermann 1898)–are used as geologic calibrations in our tree. The final closure of the Isthmus of Panama, separating the eastern Pacific (EP) and the Caribbean Sea basins, is an unresolved debate. Although age constrains of 2.8-3.5 Ma have been traditionally used to calibrate phylogenies with this formation (e.g., Coates & Obando, 1996), recent studies have challenged the timing of the final closure of the Panama Isthmus. More specifically, Montes et al. (2015) proposed that the final closure of the Central American Seaway occurred during the Middle Miocene, which would place it at 13-15 Ma. O'Dea et al. (2016), however, continue to maintain support for a younger estimate of 2.8 Ma (Pleistocene). Given these ongoing controversies, we set a lower hard bound of 2.8 Ma (with Cauchy distributions), which reflects an undisputed minimum geologic age for this event, without the implementation of upper bounds as priors in the calibrations. Prior setting MCMCTree: L(0.028,0.1,1,1e-300).

(8) Geminate *Aulostomus*. MRCA: *Aulostomus maculatus*, *Aulostomus chinensis*.

(9a) Geminate *Mulloidichthys*. MRCA: *Mulloidichthys martinicus*, *Mulloidichthys dentatus*. Comment: this calibration and calibration 9b below are used simultaneously to account for topological uncertainties regarding the sister species of *Mulloidichthys martinicus* in different trees.

(9b) Geminate *Mulloidichthys*. MRCA: *Mulloidichthys martinicus*, *Mulloidichthys vanicolensis*. Comment: this calibration and calibration 9a above are used simultaneously to account for topological uncertainties regarding the sister species of *Mulloidichthys martinicus* in different trees.

(10) Geminate *Fistularia*. MRCA: *Fistularia commersonii*, *Fistularia petimba*.

(11) Geminate *Hippocampus*. MRCA: *Hippocampus ingens*, *Hippocampus reidi*.

Table S1. Ages of Syngnatharia estimated by previous studies.

Study	Mean crown age (Ma)	95% HPD	Mean stem age (Ma)	95% HPD
Alfaro et al. (2018)	74	(65-82)	86	(76-96)
Betancur-R et al. (2013)	74	(71-80)	96	(78-114)
Betancur-R et al. (2017)	74	-	95	-
Chen et al. (2014)	-	-	90	(72-103)
Hughes et al. (2018)	80	-	93	(87-102)
Near et al. (2012)	76	(73-82)	92	(80-102)
Rabosky et al. (2018)*	103.7	-	108.6	-

*Age estimates from Rabosky et al. (2018) are outlier values and are thus excluded from the estimations. Ma: millions of years.

Table S2. Priors used to estimate divergence times in MCMCTree.

MRCA	Absolute Age (Ma)	Distribution	Calibration type	Parameters
Syngnathus louisianae, *Taractichthys longipinnis*	92-103.5	Uniform	Hard upper and hard lower bound	B(0.920,1.035,1e-300,1e-300)
Syngnathus louisianae, *Eurypegasus draconis*	83.6	Cauchy	Hard lower bound	L(0.836,0.01,0.0001,1e-300)
Solenostomus cyanopterus, *Syngnathus louisianae*	48.5	Uniform	Soft upper and hard lower bound	B(0.485,0.836,1e-300,0.05)
Hippocampus abdominalis, *Hippocampus kuda*	11.6	Uniform	Soft upper and hard lower bound	B(0.116,0.485,1e-300,0.05)
Aulostomus maculatus, *Aeoliscus strigatus*	61.5	Uniform	Soft upper and hard lower bound	B(0.615,0.836,1e-300,0.05)
Fistularia corneta, *Aulostomus maculatus*	48.5	Uniform	Soft upper and hard lower bound	B(0.485,0.615,1e-300,0.05)
Pegasus volitans, Dactylopterus volitans	48.5	Uniform	Soft upper and hard lower bound	B(0.485,0.836,1e-300,0.05)
Aulostomus maculatus, *Aulostomus chinensis*	2.8	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Mulloidichthys martinicus, *Mulloidichthys dentatus*	2.8	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Mulloidichthys martinicus, *Mulloidichthys vanicolensis*	2.8	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Fistularia commersonii, *Fistularia petimba*	2.8	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Hippocampus ingens, *Hippocampus reidi*	2.8	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)

Ma: millions of years.

Supplementary Results

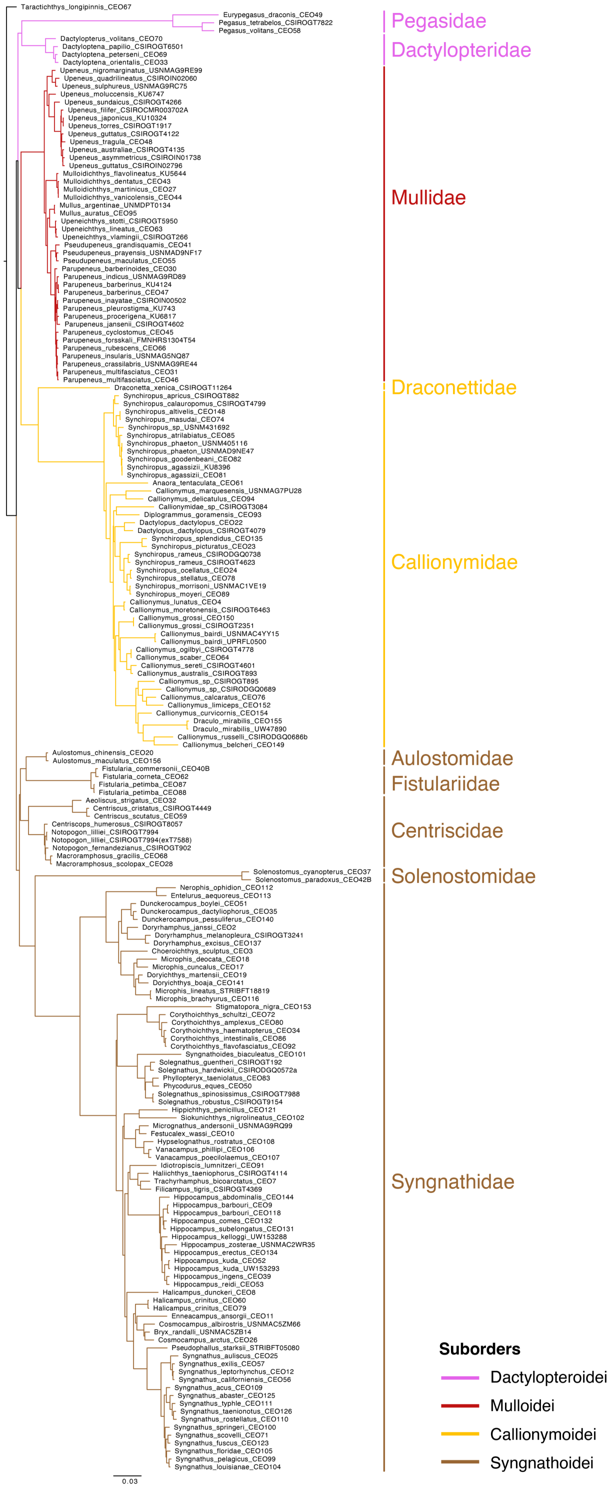


Figure S2. Phylogenetic tree inferred with RAxML (184 tips, 169 species) based on the 75% completeness matrix. Colors indicate suborders.

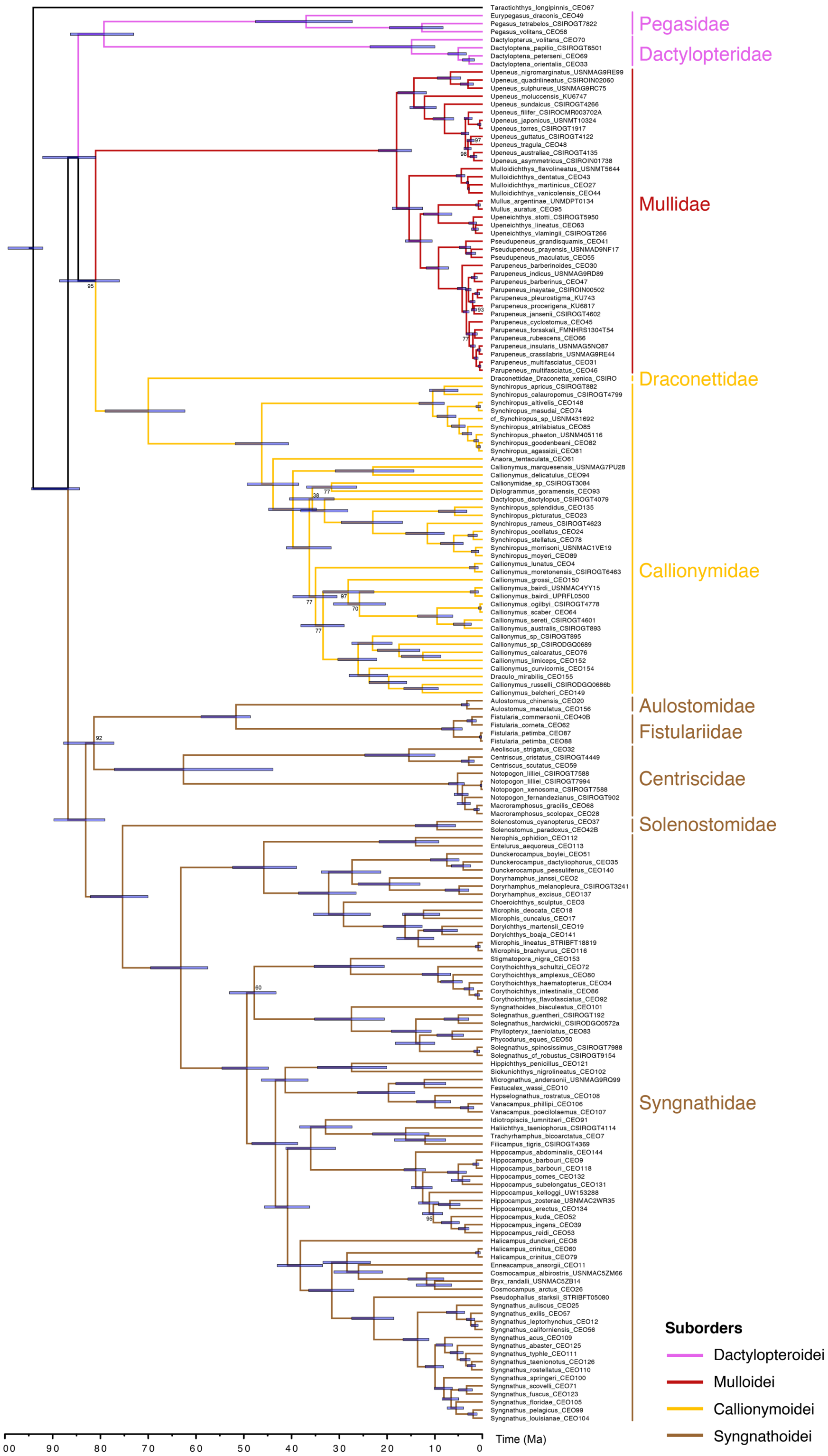


Figure S3. Phylogenetic tree inferred with RAxML for the 75% completeness matrix and time-calibrated in MCMCTree after excluding most duplicates (175 tips, 169 species; six duplicates were retained as they were re-identified after the analyses were completed). Colors indicate suborders. Nodal values indicate bootstrap support lower than 100. Ma: millions of years.

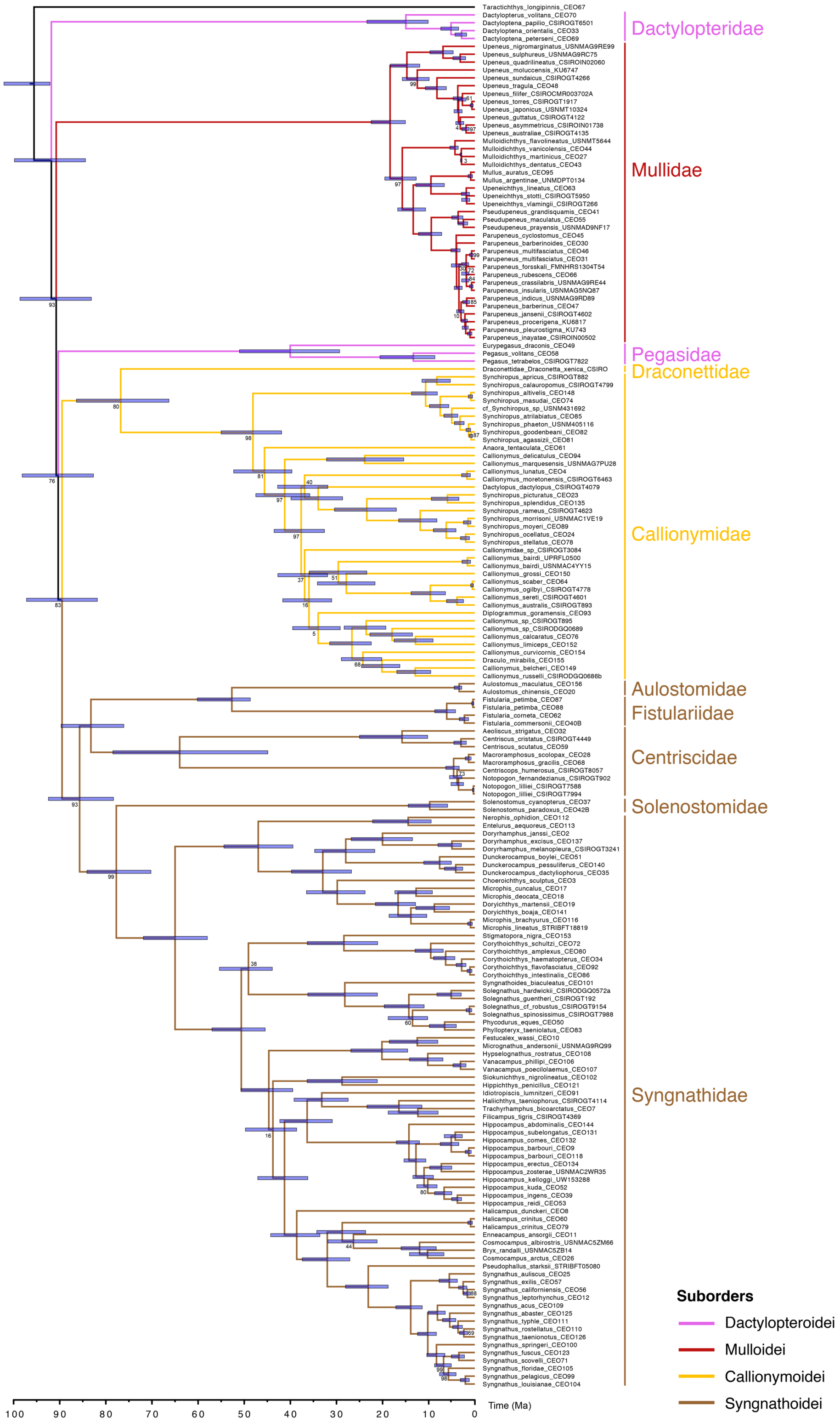


Figure S4. Phylogenetic tree inferred with ASTRAL-II (175 tips, 169 species) for the 75% completeness matrix and time-calibrated in MCMCTree. Colors indicate suborders. Nodal values indicate bootstrap support lower than 100. Ma: millions of years.

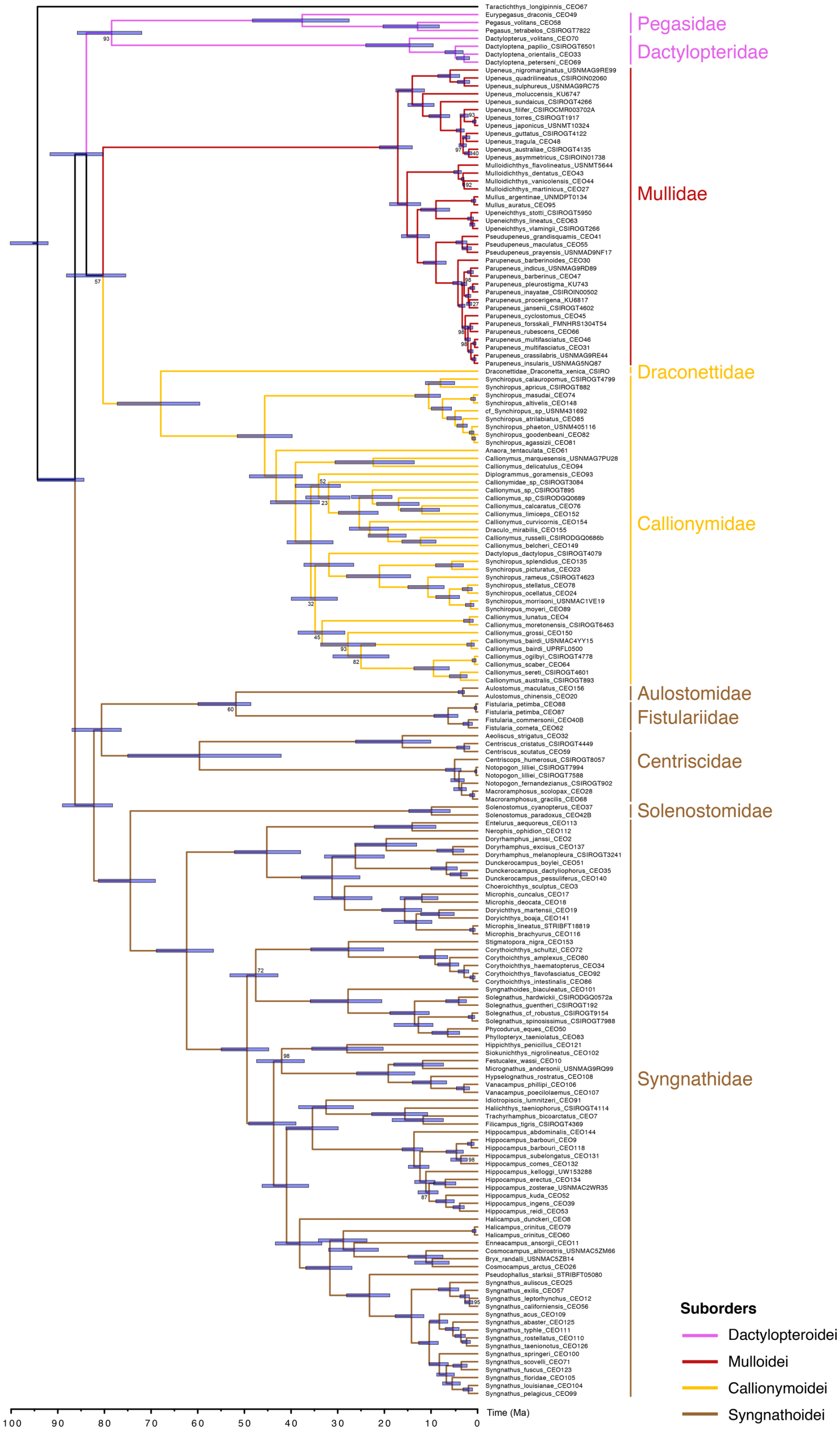


Figure S5. Phylogenetic tree inferred with RAxML (175 tips, 169 species) for the 90% completeness matrix and time-calibrated in MCMCTree. Colors indicate suborders. Nodal values indicate bootstrap support lower than 100. Ma: millions of years.

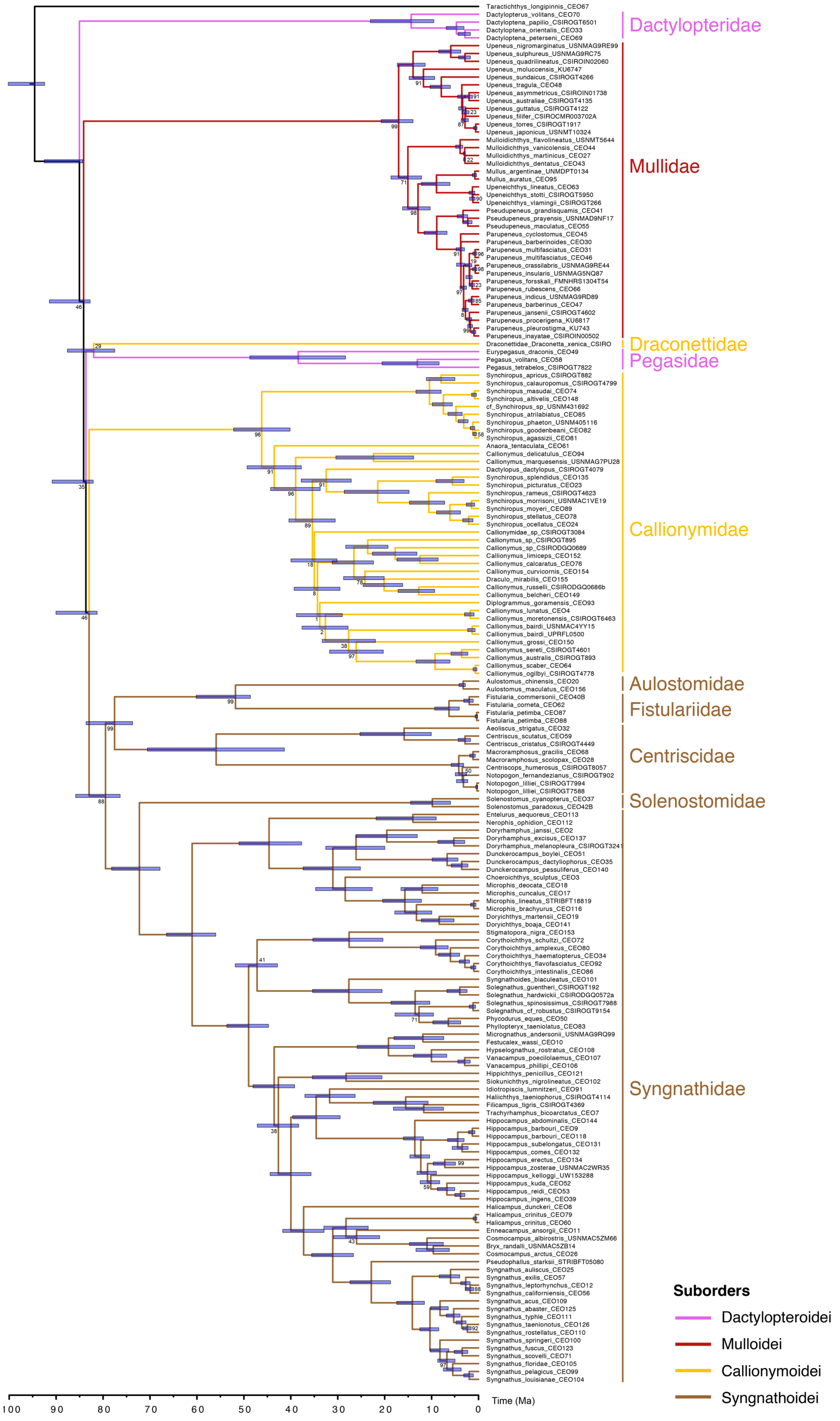


Figure S6. Phylogenetic tree inferred with ASTRAL-II (175 tips, 169 species) for the 90% completeness matrix and time-calibrated in MCMCTree. Colors indicate suborders. Nodal values indicate bootstrap support lower than 100. Ma: millions of years.

Figure S7. Time trees of Syngnatharia inferred with MCMCTree based on trees from two phylogenetic methods (RAxML and ASTRAL-II) applied to the 75% and 90% completeness matrices as well as the 12 subsets assembled from the 75% matrix. Ma: millions of years.

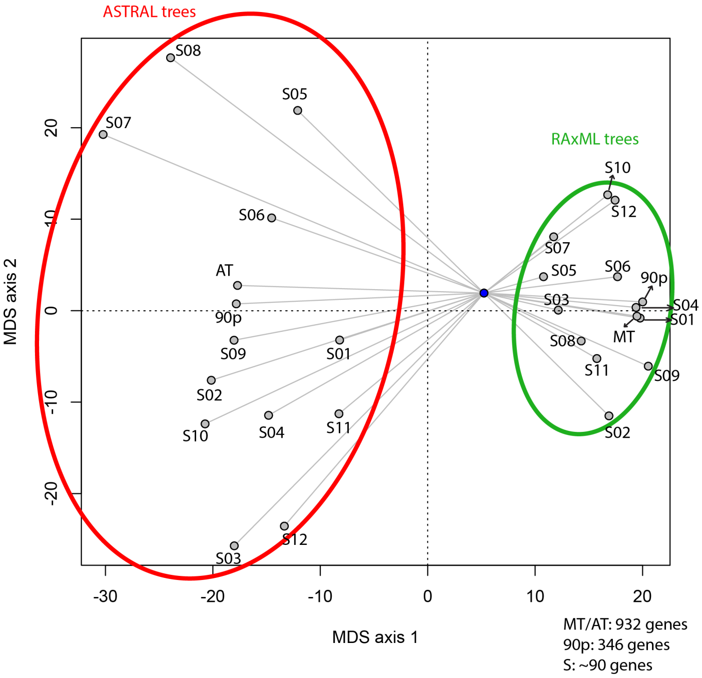


Figure S8. Tree space plot for the 28 trees estimated in this study. Note that while trees inferred with more than 300 loci (75% and 90 % completeness matrices) tend to be more tightly clustered in the tree space relative to subset trees, a weak association between RAxML and ASTRAL-II topologies obtained with the same genomic subset persists (e.g., 'MT' RAxML vs 'AT' ASTRAL-II, 'S01' RAxML vs. 'S01' ASTRAL-II, 'S07' RAxML vs. 'S07' ASTRAL-II). The average (centroid) tree in tree space is represented as a blue dot. MT: 'master tree', AT: alternative ASTRAL-II tree based on the 75% completeness matrix, 90p: trees based on 90 % completeness matrix, S: trees generated with each based subset.

Table S3. Divergence time estimates in millions of years (Ma) for major syngnatharian clades and families based on the 28 trees dated in MCMCTree.

	Number of million generations ran in MCMCTree		Syngnatharia Stem	Syngnatharia Crown	Syngnathoidei	Callionymoidei	Mulloidei	Dactylopteroidei	Syngnathidae	Solenostomidae	Centriscidae	Fistulariidae	Aulostomidae	Draconettidae	Callionymidae	Mullidae	Dactylopteridae	Pegasidae
MASTER_TREE_matrix75_RA	50	Age	94.06	86.79	83.09	69.99	18.02	79.27	63.19	9.55	62.64	6.1	3.31	69.99	46.24	18.02	14.87	36.96
		Min	92.05	84.38	79.07	62.31	14.9	73.04	57.53	5.68	43.88	4.27	2.81	62.31	40.64	14.9	9.98	27.3
		Max	99.35	94.39	89.75	79.03	21.81	86.31	69.51	14.15	77.08	8.59	4.46	79.03	51.82	21.81	23.59	47.53
MCMC_matrix75_ASTRAL	30	Age	95.57	91.82	85.69	76.78	18.38	NA	64.99	9.77	63.99	6.1	3.4	76.78	48.1	18.38	14.95	40.03
		Min	92.11	84.41	78.34	66.32	15.06	NA	57.99	5.9	44.88	4.19	2.82	66.32	41.9	15.06	10.12	29.3
		Max	102.1	99.82	92.45	86.37	22.48	NA	71.86	14.44	78.48	8.69	4.48	86.37	55.01	22.48	22.29	51.02
MCMC_matrix90_RAXML	10	Age	94.35	86.31	82.24	67.92	17.12	78.46	62.37	9.9	59.63	6.33	3.2	67.92	45.65	17.12	14.62	37.6
		Min	92.06	84.38	78.25	59.55	14	71.99	56.62	5.88	42.11	4.22	2.81	59.55	39.76	14	9.52	27.54
		Max	100.18	94.36	89.01	77.25	21.04	85.82	68.88	14.76	75	9.38	4.19	77.25	51.55	21.04	23.99	48.29
MCMC_matrix90_ASTRAL	20	Age	94.58	85.03	79.49	NA	17.02	NA	61.03	9.85	55.91	6.29	3.27	82.01	46.2	17.02	14.36	38.42
		Min	92.46	84.37	76.37	NA	13.96	NA	55.99	5.97	41.37	4.11	2.81	77.51	40.16	13.96	9.53	28.35
		Max	100.21	92.55	85.85	NA	20.76	NA	66.48	14.47	70.56	9.35	4.15	87.59	52.19	20.76	23.04	48.74
MCMC_S01_ASTRAL	8	Age	95.24	87.14	81.99	74.15	19.88	79.61	63.04	10.48	55.91	7.11	3.86	74.15	46.82	19.88	15.47	36.63
		Min	92.47	84.38	76.73	64.5	15.63	71.27	56.99	5.98	38.57	4.67	2.86	64.5	40.34	15.63	25.55	25.88
		Max	101.71	97.59	91.03	84.87	25.12	87.91	70.9	16.86	74.56	10.75	5.65	84.87	53.32	25.12	9.67	48.21
MCMC_S01_RAXML	8	Age	95.32	87.78	82.99	73.04	20.82	77.81	63.55	10.51	58.26	7.2	3.94	73.04	48.74	20.82	15.23	36.13
		Min	92.48	84.38	77.38	63.72	16.18	68.9	57.14	5.94	38.8	4.71	2.85	63.72	42.71	16.18	9.49	25.3
		Max	101.82	97.96	91.78	83.56	26.43	86.55	71.36	17.19	76.85	11.7	5.94	83.56	55.52	26.43	25.44	48.17
MCMC_S02_ASTRAL	3	Age	95.04	85.23	82.18	67.36	16.02	NA	63.06	10.42	55.93	5.78	4.37	67.36	44.2	16.02	19.6	35.86
		Min	92.47	84.36	78.97	58.9	12.79	NA	57.14	5.81	38.03	3.57	2.9	58.9	37.51	12.79	10.75	25.22
		Max	101.33	93.48	88.93	75.88	20.28	NA	69.72	16.77	74.15	9.09	7.25	75.88	50.62	20.28	33.98	47.21
MCMC_S02_RAXML	3	Age	95.11	84.42	81.54	70.87	16.1	76.88	61.61	10.23	55.91	5.68	4.45	70.87	48.47	16.1	18.97	36.09
		Min	91.48	84.36	78.12	62.66	12.72	69.64	55.55	5.46	36.55	3.57	2.91	62.66	42.76	12.72	10.27	24.83
		Max	101.5	91.11	86.23	77.93	20.62	82.4	67.6	22.16	75.69	8.87	7.38	77.93	54.61	20.62	34.08	47.94
MCMC_S03_ASTRAL	3	Age	94.5	84.94	80.73	70.53	22.27	NA	61.32	8.55	62.7	4.97	5.08	70.53	42.03	22.27	20.57	28.52
		Min	92.46	84.37	77.85	61.7	17.78	NA	56.3	4.44	41.49	3.3	3.05	61.7	35.8	17.78	12.17	28.1
		Max	99.99	91.5	85.25	77.58	27.69	NA	66.51	21.36	76.03	7.23	8.22	77.58	48.5	27.69	32.46	48.53
MCMC_S03_RAXML	3	Age	95.1	87.83	84.02	69.38	22.27	NA	63.38	8.59	61.42	5.21	5.6	69.38	44.2	22.27	20.64	37.89
		Min	92.47	84.39	78.68	59.72	17.42	NA	56.84	4.58	40.73	3.3	3.21	59.72	37.45	17.42	11.85	27.25
		Max	101.35	97.79	93.08	80.19	28.41	NA	71.33	13.89	77.93	7.65	8.29	80.19	50.67	28.41	33.25	48.62
MCMC_S04_ASTRAL	20	Age	98.4	94.21	87.56	78.33	20.42	NA	66.56	9.16	65.07	5.85	3.27	78.33	47.58	20.42	17.36	41.28
		Min	92.2	84.41	76.52	64.62	15.73	NA	56.52	4.55	45.24	3.91	2.81	64.62	38.47	15.73	10.66	28.08
		Max	103.35	102.65	96.36	90.68	26.27	NA	75.22	15.1	81.15	8.87	4.52	90.68	57.35	26.27	29.93	56.75
MCMC_S04_RAXML	3	Age	94.94	84.59	81.35	71.38	19.91	78.36	61.78	8.56	56.44	5.66	3.27	71.38	45.97	19.91	17.52	36.38
		Min	92.47	84.36	78.64	63.1	15.7	73.02	56.66	4.22	38.01	3.77	2.81	63.1	39.83	15.7	10.4	25.63
		Max	101.17	86.66	83.39	77.8	25.22	81.71	66.92	14.22	74.27	8.39	4.49	77.8	51.58	25.22	29.91	47.62
MCMC_S05_ASTRAL	3	Age	94.97	85.54	81.32	NA	19.33	NA	63.11	9.8	60.76	5.76	3.36	78.56	46.59	19.33	18.58	38.56
		Min	92.47	84.37	77.84	NA	15.45	NA	57.51	5.44	42.29	3.84	2.81	48.04	40.05	15.45	11.19	27.58
		Max	101.24	94.76	88.78	NA	24.09	NA	69.65	15.08	75.37	21.93	4.98	90.78	53.06	24.09	29.72	49.66
MCMC_S05_RAXML	3	Age	94.85	85.29	81.67	71.03	19.44	NA	62.82	9.64	62.8	5.63	3.36	71.03	46.14	19.44	18.76	36.15
		Min	92.46	84.37	77.88	62.85	15.4	NA	57.21	54.46	42.64	3.69	2.81	62.85	40.18	15.4	11.14	26.03
		Max	101	93.18	87.97	79.41	24.51	NA	68.89	14.9	76.32	8.21	4.85	79.41	51.77	24.51	30.02	47.31
MCMC_S06_ASTRAL	3	Age	95.09	84.96	79.78	NA	17.06	NA	61.69	9.02	53.83	6.37	3.53	82.92	46.08	17.06	13.2	38.5
		Min	92.48	84.36	76.35	NA	13.79	NA	56.45	2.9	38.37	4.13	2.82	80.82	39.41	13.79	8.28	28.01
		Max	101.36	91.99	85.17	NA	21.26	NA	67.16	14.04	71.41	9.94	5.15	89.37	52.23	21.26	22.42	49.05
MCMC_S06_RAXML	3	Age	95.16	84.67	81.98	73.78	17.44	75	62.3	9.16	59.59	6.49	3.59	73.78	44.24	17.44	13.49	37.5
		Min	92.48	84.36	79.51	64.88	13.96	68.6	57.13	49.7	39.04	4.21	2.82	64.88	37.87	13.96	8.27	27.17
		Max	101.59	89.11	85.45	80.03	22	80	67.59	14.34	76.23	10.2	5.38	80.03	50.34	22	23.47	48.12
MCMC_S07_ASTRAL	3	Age	94.62	84.64	81.28	NA	17.16	NA	63.02	11.02	55.11	6.19	3.12	79.73	45.49	17.16	18.68	36.83
		Min	92.46	84.36	78.73	NA	13.84	NA	57.72	6.26	39.91	3.94	2.81	71.16	38.77	13.84	11.11	26.23
		Max	100.36	87.88	84.07	NA	21.52	NA	68.35	16.92	71.67	9.33	3.99	83.41	55.26	21.52	31.87	48.16
MCMC_S07_RAXML	8	Age	95.52	88.74	85.55	70.57	17.19	NA	65.42	11.11	56.49	6.31	3.11	70.57	45.21	17.19	19.02	35.78
		Min	92.48	84.39	79.65	60.5	13.7	NA	58.2	6.18	39.97	3.99	2.81	60.5	37.82	13.7	11.1	25.24
		Max	102.09	99.33	95.64	81.44	21.65	NA	74.06	17.31	74.38	9.54	4.01	81.44	52.82	21.65	32.79	47.76
MCMC_S08_ASTRAL	3	Age	95.14	84.87	80.29	NA	16.3	78.12	61.77	8.51	50.83	5.75	3.36	84.04	39.67	16.3	15.69	36.89
		Min	92.47	84.36	77.07	NA	12.96	70.69	56.41	4.17	36.44	3.63	2.81	82.51	32.11	12.96	9.69	26.27
		Max	101.58	91.74	85.73	NA	20.88	83.78	67.33	14.22	69.44	8.79	4.62	90.7	47.93	20.88	26.96	48.12
MCMC_S08_RAXML	3	Age	95.24	85.03	81.36	68.65	17.12	NA	62.14	8.5	52.18	5.89	3.37	68.65	41.2	17.12	15.42	36.22
		Min	92.47	84.36	78.11	58.59	13.67	NA	56.56	4.24	36.41	3.71	2.81	58.59	34.24	13.67	9.57	25.55
		Max	101.66	92.53	87.48	77.27	21.73	NA	68.2	13.34	71.41	8.99	4.81	77.27	48.55	21.73	26.56	47.77
MCMC_S09_ASTRAL	8	Age	94.98	85.54	80.64	68.14	16.72	NA	61.78	7.36	55.17	6.96	3.54	68.14	46.18	16.72	11.31	36.56
		Min	92.47	84.37	76.68	59.46	13.3	NA	56.2	3.79	39.99	4.49	2.82	59.46	40.02	13.3	7.24	25.81
		Max	101.16	94.14	88.01	76.54	21.21	NA	68.17	12.14	71.2	10.24	5.12	76.54	52.19	21.21	17.67	48.11
MCMC_S09_RAXML	3	Age	94.99	84.93	81.45	67.61	17.09	77.48	61.46	7.51	59.34	7.01	3.62	67.61	46.13	17.09	12.43	30.06
		Min	92.46	84.36	78.3	59.13	13.58	70.27	56.08	3.94	41.27	4.58	2.82	59.13	40.07	13.58	7.9	25.36
		Max	101.17	91.68	86.58	75.26	21.69	83.43	67.08	12.58	75.15	10.31	5.35	75.26	51.9	21.69	20.17	47.62
MCMC_S10_ASTRAL	3	Age	94.81	84.62	80.99	69.19	18.72	NA	61.87	8.84	58.56	6.13	3.16	69.19	49.07	18.72	10.89	39.33
		Min	92.46	84.36	77.88	61.72	14.51	NA	56.55	4.73	39.72	3.97	2.81	61.72	43.78	14.51	6.52	28.43
		Max	100.96	87.78	83.4	75.72	23.9	NA	67.17	14.13	74.23	9.24	4.36	75.72	54.84	23.9	17.28	49.91
MCMC_S10_RAXML	3	Age	94.93	84.6	80	69.69	19.43	79.85	61.87	8.94	57.34	6.16	3.2	69.69	49.73	19.43	11.13	39.2
		Min	92.47	84.36	77.59	62.03	15.07	74.83	56.44	4.63	38.58	4	2.81	62.03	44.46	15.07	6.5	28.12
		Max	101.21	87.31	83.19	76.17	25.1	83.2	67.15	14.5	73.83	9.93	4.43	76.17	55.88	25.1	18.22	49.99
MCMC_S11_ASTRAL	10	Age	96.23	89.94	NA	70.46	26.5	NA	64.97	7.74	53.03	6.4	3.75	70.46	48.88	26.5	12.93	40.28
		Min	92.52	84.39	NA	61.1	20.52	NA	57.69	3.84	28.64	4.55	2.83	61.1	42.6	20.52	7.97	28.84
		Max	102.66	100.25	NA	80.43	33.9	NA	73.2	13.25	72.85	10.04	5.45	80.43	55.56	33.9	22.07	51.58
MCMC_S11_RAXML	3	Age	95.14	85.1	82.4	69.11	26.47	NA	63.21	7.66	56.94	6.64	3.65	69.11	48.76	26.47	12.77	38.44
		Min	92.47	84.37	79.34	61.21	20.62	NA	57.32	3.72	38.21	4.51	2.82	61.21	43.14	20.62	7.78	27.54
		Max	101.51	93.22	89.62	77.15	33.91	NA	69.78	13.02	74.58	10.32	5.28	77.15	54.73	33.91	22.14	49.32
MCMC_S12_ASTRAL	8	Age	94.95	86.28	NA	66.98	21.99	NA	65.68	9.43	53.53	5.81	3.19	66.98	46.84	21.99	12.8	37.38
		Min	92.47	84.38	NA	59.09	17.19	NA	59.45	4.84	39.59	3.94	2.81	59.09	40.62	17.19	8.02	26.75
		Max	101.1	95.85	NA	75.26	27.9	NA	73.04	15.08	70.77	8.74	4.31	75.26	53.07	27.9	21.18	48.63
MCMC_S12_RAXML	3	Age	94.8	84.94	81.94	68.47	22.21	80.67	63.27	9.11	60.7	5.88	3.21	68.47	47.4	22.21	12.56	38.06
		Min	92.46	84.37	78.88	60.53	17.34	75.75	57.66	4.75	41.16	3.88	2.81	60.53	41.6	17.34	7.87	27.06
		Max	100.88	91.86	87.92	76.19	28.14	86.06	69.3	14.34	75.38	8.88	4.24	76.19	53.28	28.14	21.39	49.13

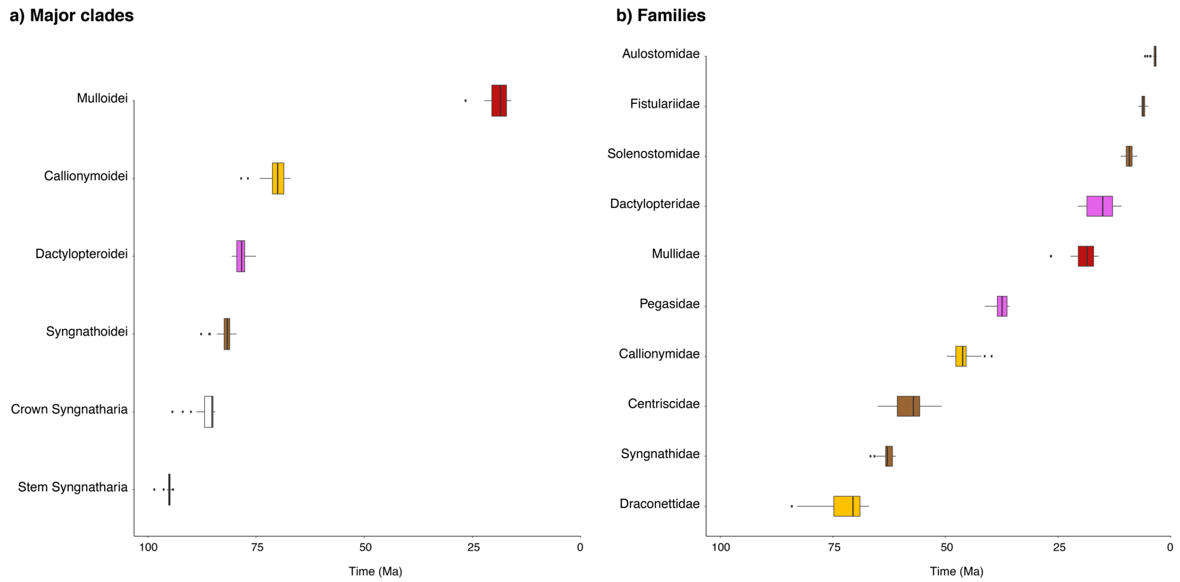


Figure S9. Divergence time estimates for major syngnatharian **(a)** clades and **(b)** families based on the 28 trees dated in MCMCTree. Boxplots are colored by suborder. Ma: millions of years.

Table S4. Summary statistics of the 12 biogeographic models implemented in BioGeoBEARS for the three alternative area schemes. The best-fitting model for each scheme is indicated in bold.

Schemes	**LnL**	**Number of parameters**	**d**	**e**	**j**	**w**	**AICc**	**AICc weight**
Six-area scheme								
DEC	-580.6	2	0.035	0.018	0	1	1165	2.00E-33
DEC+*j*	-579.7	3	0.033	0.016	0.013	1	1166	1.80E-33
DEC+*w*	-555.4	3	0.023	0.01	0	0.23	1117	6.10E-23
DEC+*j*+*w*	-612.9	4	0.0094	0.023	0.0008	0.013	1234	2.40E-48
DIVALIKE	-597	2	0.04	0.017	0	1	1198	1.50E-40
DIVALIKE+*j*	-596	3	0.037	0.014	0.014	1	1198	1.50E-40
DIVALIKE+*w*	-589.1	3	0.024	0.009	0	0.23	1184	1.40E-37
DIVALIKE+*j*+*w*	-645.2	4	0.011	0.024	0.0056	0.016	1299	2.10E-62
BAYAREALIKE	-539	2	0.035	0.052	0	1	1082	2.40E-15
BAYAREALIKE+*j*	**-504.3**	**3**	**0.018**	**0.022**	**0.088**	**1**	**1015**	**1**
BAYAREALIKE+*w*	-542.3	3	0.021	0.04	0	0.27	1091	3.20E-17
BAYAREALIKE+*j*+*w*	-545.5	4	0.0056	0.041	0.009	0.0095	1099	4.40E-19

Seven-area scheme								
DEC	-687.1	2	0.075	0.051	0	1	1378	2.60E-44
DEC+*j*	-686.7	3	0.071	0.048	0.018	1	1380	1.30E-44
DEC+*w*	-674.1	3	0.039	0.027	0	0.34	1354	4.00E-39
DEC+*j*+*w*	-671.8	4	0.03	0.017	0.037	0.3	1352	1.40E-38
DIVALIKE	-694.9	2	0.092	0.059	0	1	1394	1.10E-47
DIVALIKE+*j*	-694.9	3	0.092	0.059	1.00E-05	1	1396	4.00E-48
DIVALIKE+*w*	-732.7	3	0.021	0.011	0	0.15	1472	1.50E-64
DIVALIKE+*j*+*w*	-699.1	4	0.064	0.041	0.074	0.45	1406	2.00E-50
BAYAREALIKE	-632.8	2	0.079	0.098	0	1	1270	1.00E-20
BAYAREALIKE+*j*	-605.2	3	0.025	0.03	0.14	1	1217	3.40E-09
BAYAREALIKE+*w*	-632.8	3	0.075	0.095	0	0.89	1272	3.40E-21
BAYAREALIKE+*j*+*w*	**-584.7**	**4**	**0.012**	**0.015**	**0.087**	**0.26**	**1178**	**1**
Eight-area scheme								
DEC	-709.1	2	0.076	0.049	0	1	1422	4.20E-37
DEC+*j*	-707.7	3	0.064	0.044	0.024	1	1422	6.30E-37
DEC+*w*	-709.9	3	0.046	0.039	0	0.29	1426	7.30E-38
DEC+*j*+*w*	-695.2	4	0.033	0.019	0.017	0.39	1399	6.10E-32
DIVALIKE	-716.3	2	0.084	0.055	0	1	1437	3.10E-40
DIVALIKE+*j*	-716.3	3	0.084	0.055	1.00E-05	1	1439	1.10E-40
DIVALIKE+*w*	-749.9	3	0.025	0.009	0	0.21	1506	3.10E-55
DIVALIKE+*j*+*w*	-718.2	4	0.067	0.044	8.80E-05	0.63	1445	6.40E-42
BAYAREALIKE	-654.5	2	0.068	0.091	0	1	1313	2.30E-13
BAYAREALIKE+*j*	**-624.3**	**3**	**0.024**	**0.029**	**0.15**	**1**	**1255**	**1**
BAYAREALIKE+*w*	-654.7	3	0.067	0.093	0	0.99	1316	6.70E-14
BAYAREALIKE+*j*+*w*	-768.1	4	0.0088	0.066	0.051	0.0022	1544	1.30E-63

LnL: LogLikelihood; d: dispersal; e: extinction; j: founder-speciation; w: dispersal matrix power exponential; AICc: Corrected Akaike Information Criterion.

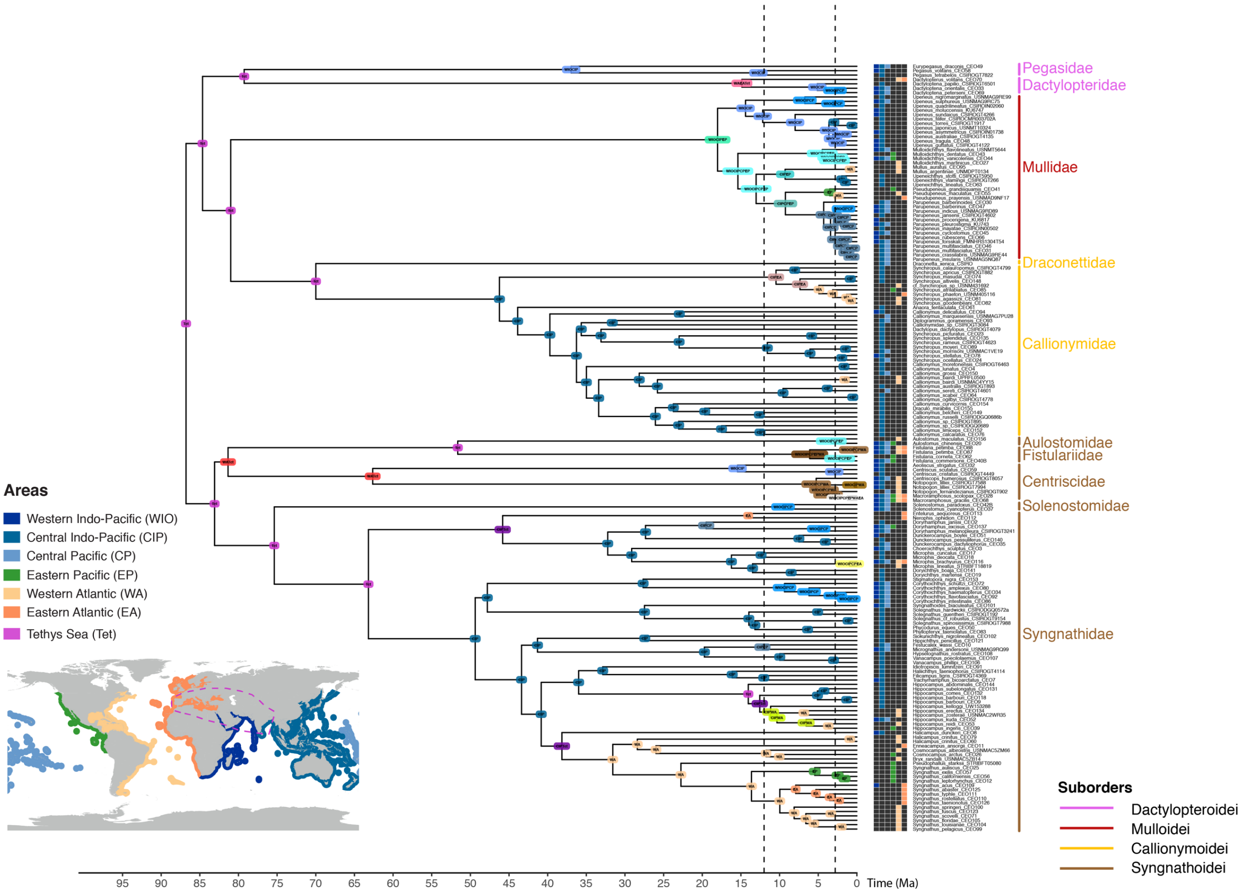


Figure S10. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model based on the six-area scheme (BayAREA+*j*) applied to the 'master tree' in BioGeoBEARS. Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the final Tethys Seaway closure (dated at 12 Ma; 12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Ma: millions of years.

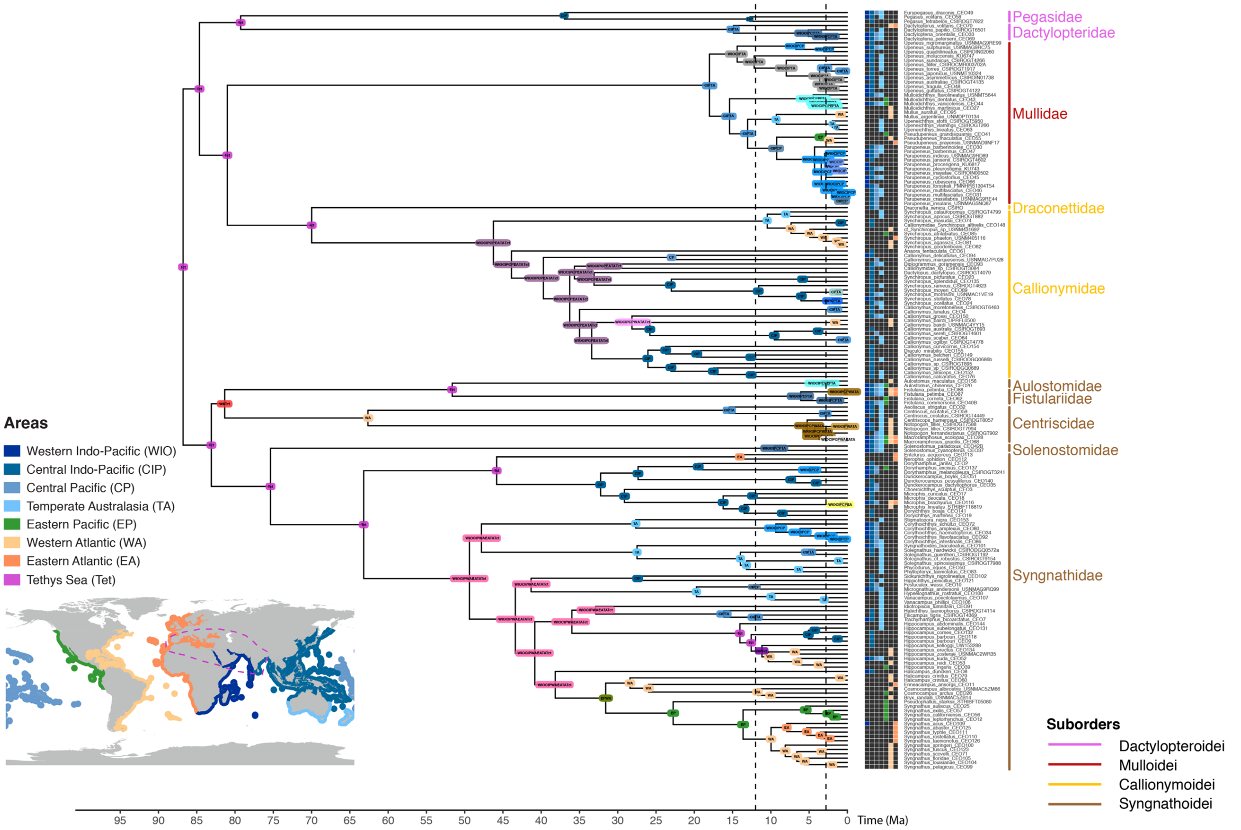


Figure S11. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model based on the seven-area scheme (BayAREA+*j*+*w*) applied to the 'master tree' in BioGeoBEARS. Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the final Tethys Seaway closure (dated at 12 Ma; 12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Ma: millions of years.

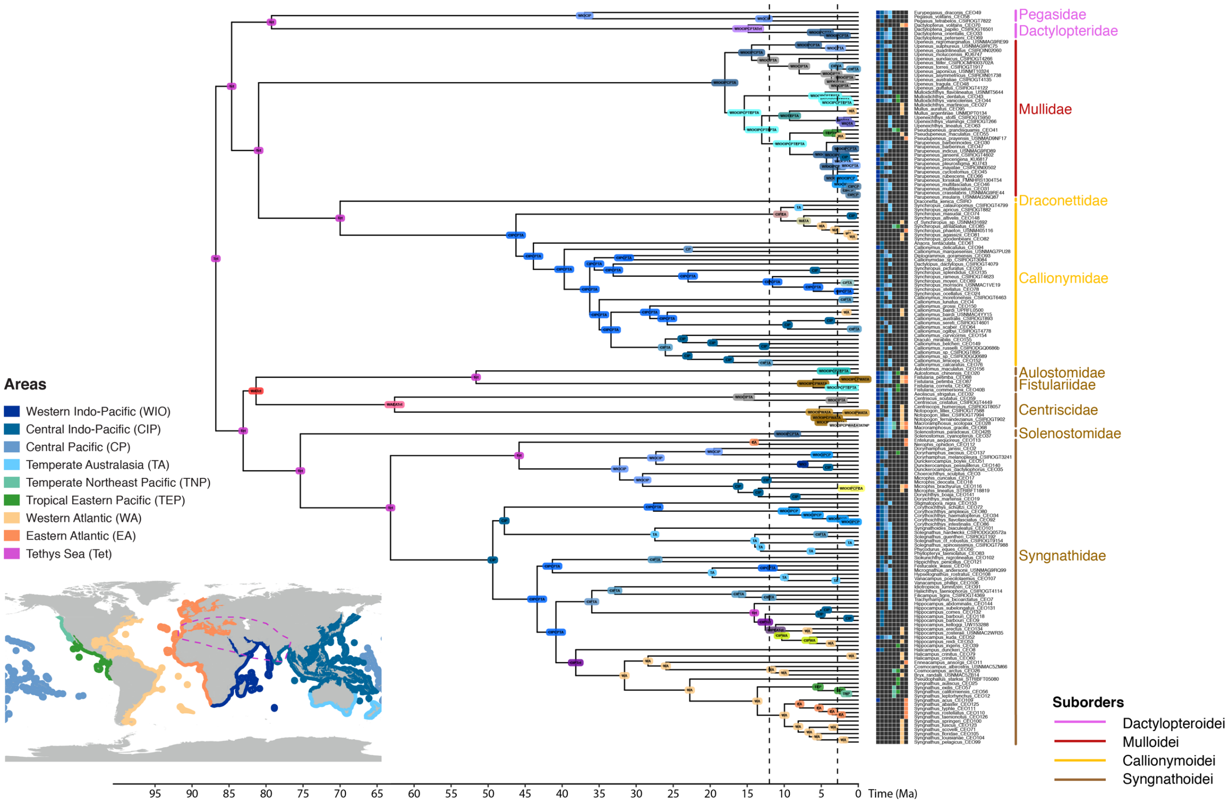


Figure S12. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model based on the eight-area scheme (BayAREA+*j*) applied to the 'master tree' in BioGeoBEARS. Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the final Tethys Seaway closure (dated at 12 Ma; 12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Ma: millions of years.

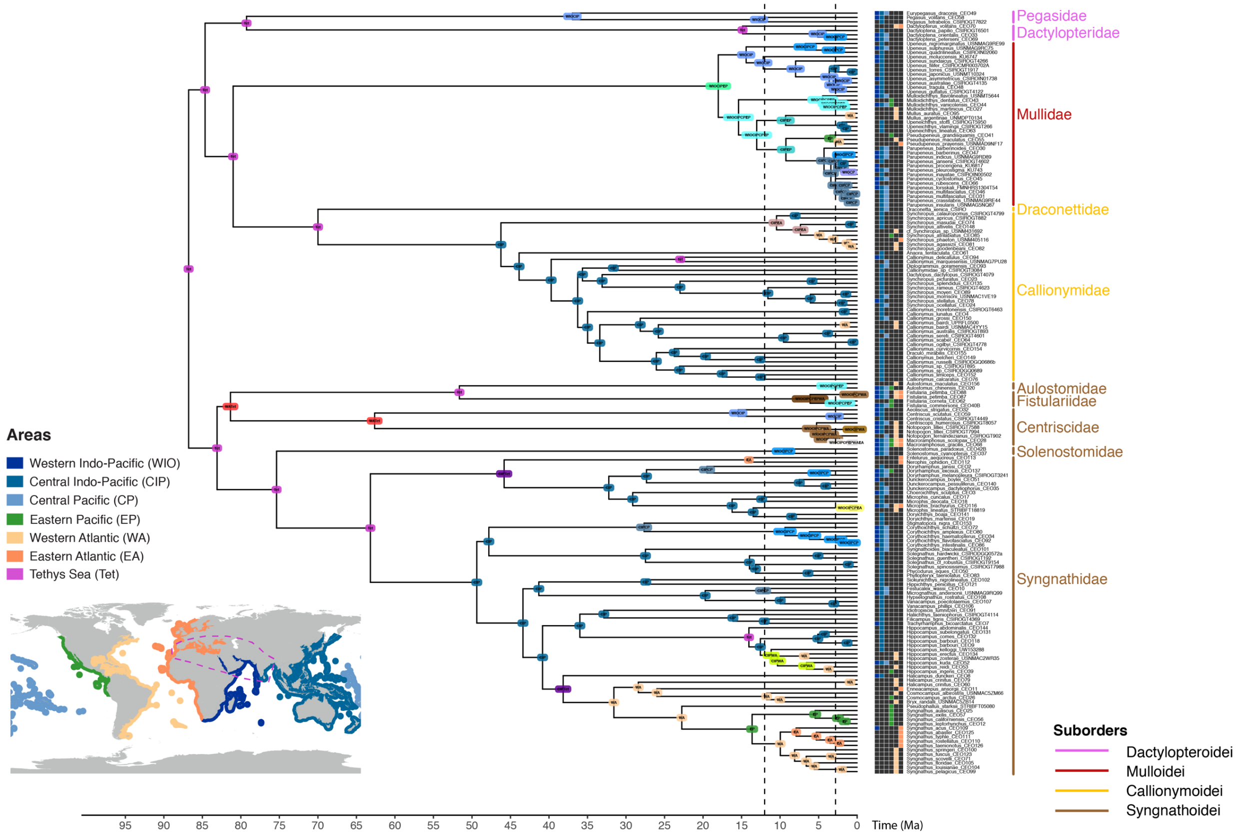


Figure S13. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model based on the six-area scheme (BayAREA+*j*) applied on the 28 trees using the ‘master tree’ as a fix topology in BioGeoBEARS. Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the final Tethys Seaway closure (dated at 12 Ma; 12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Ma: millions of years.

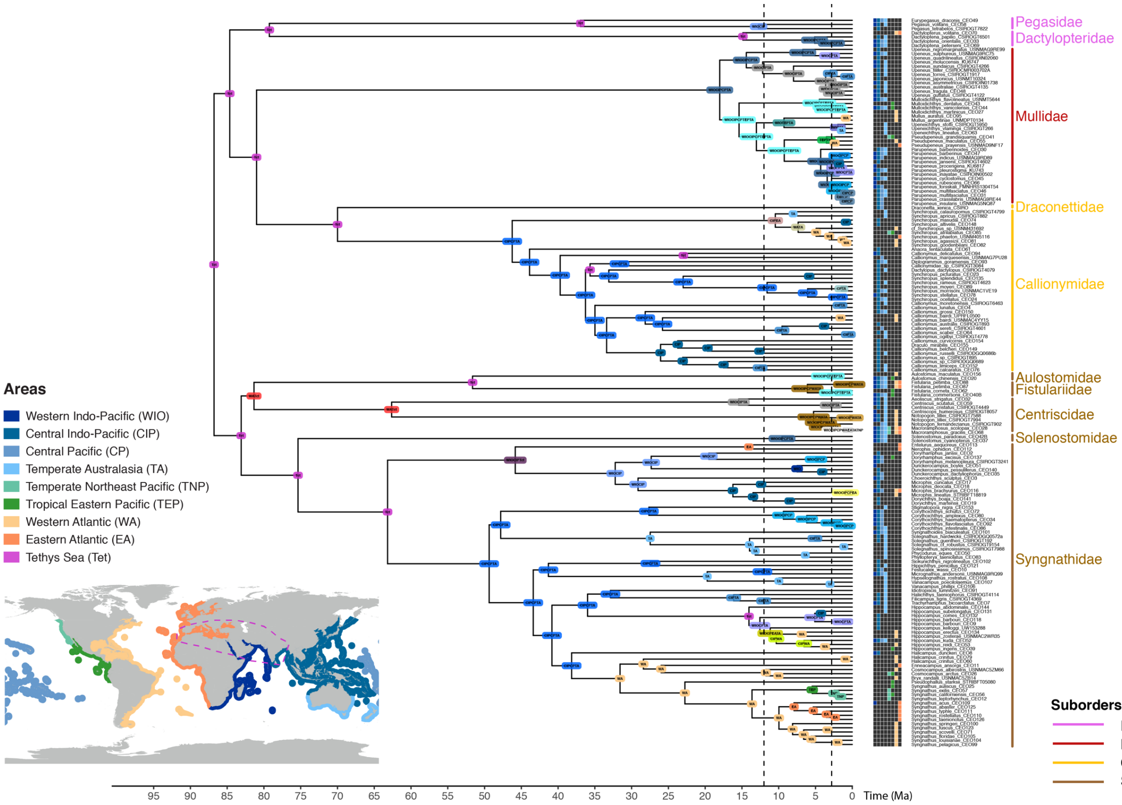


Figure S14. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model based on the eight-area scheme (BayAREA+*j*) applied on the 28 trees using the 'master tree' as a fix topology in BioGeoBEARS. Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the final Tethys Seaway closure (dated at 12 Ma; 12–18 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Ma: millions of years.

Table S5. Comparisons of ancestral range estimations in BioGeoBEARS between the three area schemes according to the best-fit biogeographic model estimated for the 'master tree' alone and the 28-tree averaging approach, including and excluding the jump dispersal parameter (*j*).

	6 AREAS + Tethys				**7 AREAS + Tethys**				**8 AREAS + Tethys**			
	BAYAREALIKE		**BAYAREALIKE+*j***		**BAYAREALIKE**		**BAYAREALIKE+*j*+*w***		**BAYAREALIKE**		**BAYAREALIKE+*j***	
	'Master tree' alone	28-tree averaging	'Master tree' alone	28-tree averaging	'Master tree' alone	28-tree averaging	'Master tree' alone	28-tree averaging	'Master tree' alone	28-tree averaging	'Master tree' alone	28-tree averaging
Dactylopteroidei												
Dactylopterus1	CIPEA	Tet	WAEATet	Tet	WIOCIPWAEATet	Tet	CIPTA	Tet	WIOCIPEATet	Tet	WIOCIPCPTATet	Tet
Mulloidei												
Pseudupeneus1	CIPWA	CIPWA	CIPCPEP	CIPEP	CIPWA	WIOCIPWA	CIPCP	WIOCIPCPTA	CIPWA	WIOCIPWA	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Pseudupeneus2	EPWA	EPWA	EP	EP	EPWA	EPWA	EP	EP	TEPWA	TEPWA	TEP	TEPTNP
Pseudupeneus3	WA	WA	WA	WA	WA	WA	WA	WA	WA	WAEA	WA	WA
Mullus1	CIPWA	CIPWA	CIPEP	CIPEP	WATA	WATA	TA	WIOCIPTA	WATA	WATA	WIOTEPTA	WIOTEPTA
Mullus2	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA
Mulloidichthys1	CIPEPWA	CIPEPWA	WIOCIPCPEP	WIOCIPCPEP	CIPEPWA	CIPEPWA	WIOCIPCPTA	WIOCIPCPTA	CIPTEPWA	CIPTEPWA	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Mulloidichthys2	WIOCIPCPWA	CIPEPWA	WIOCIPCPEP	WIOCIPCPEP	CIPEPWA	CIPEPWA	WIOCIPCPTA	WIOCIPCPTA	CIPTEPWA	CIPTEPWA	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Callionymoidei												
Synchiropus1	CIPWA	CIPWA	CIPEA	CIPEA	CIPWA	CIPWA	WA	CIP	CIPWA	CIPWA	WATA	WATA
Synchiropus1	WA	WA	WA	WA	WA	EPWA->WA	WA	WA	WA	WA	WA	WA
Synchiropus2	WA->EP	EPWA->EP	WA->EP	WA->EP	EPWA->EP	EPWA->EP	WA->EP	WA->EP	TEPWA->TEPTNP	TEPWA->TEPTNP	WA->TEPTNP	WA->TEPTNP
Synchiropus3	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA
Callionymus1	CIPCP	CIPCP	CIP	CIP	CIPCPTA	CIPCPTA	CIP	CIP	CIPCPTA	CIPCPTA	CIPCPTA	CIPCPTA
Syngnathoidei												
Aulostomus1	WIOCIPEPWA	WIOCIPEPWA	WIOCIPCPEP	WIOCIPCPEP	WIOCIPEPWA	Tet	WIOCIPCPEPTA	WIOCIPCPEPTA	WIOCIPTEPWA	Tet	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Entelurus+Nerophis1	WIOCIPTet	WIOCIPTet	CIPTet	CIPTet	WIOCIPCPTATet	WIOCIPCPWAEATATet	Tet	CIP	WIOCIPTet	WIOCIPCPWAEATATet	Tet	WIOCIPTet
Hippocampus	CIPTet	CIPWA	CIPTet	CIP	CIPTet	CIPWATet	Tet	CIPTA	CIPTet	CIPWA	CIPTet	WIOCPTA
Hippocampus1a	CIPWA->WA	CIPWA->WA	CIPWA->WA	CIPWA->WA	CIPWA->WA	CIPWA->WA	CIPTet->WA	CIPTA->WA	CIPWA->WA	CIPWA->WA	CIPWA->WA	CIPWA->WA
Hippocampus1b	CIPWA->WA	CIPWA->EPWA	CIPWA->WA	CIPWA->WA	CIPWA->EPWA	CIPWA->EPWA	WA->WIOCIPCPTA	CIPCIPCPTA->EP	CIPWA->TEPWA	CIPWA->TEPWA	CIPWA->WA	CIPWA->WA
Hippocampus2	WA->EP	EPWA->WA and EP	WA->EP	WA->EP	EPWA->WA and EP	EPWA->WA and EP	WA->EP	EP->WA and EP	TEPWA->WA and TEP	TEPWA->WA and TEP	WA->TEP	WA->TEP
Halicampus+Syngnathus1	CIPTet	WIOCIPTet	CIPTet	CIPTet	WAEATet	WIOCIPCPWAEATATet	WIOCIPCPWAEATATet	CIPTA	WIOCIPTet	WIOCIPCPWAEATATet	CIPTet	CIPCPTA
Halicampus+Syngnathus1	WA	WA	WA	WA	EPWA	EPWAEA	EPWA	WA	WAEA	WAEATet	WA	WA
Halicampus+Enneacampus1	WA	WA	WA	WA	EPWA	EPWAEA	WA	WA	WAEA	WAEA	WA	WA
Halicampus+Cosmocampus1	WA	EPWA	WA	WA	EPWA	EPWA	WA	WA	TEPWA	TEPWA	WA	WA
Halicampus+Pseudophallus1	EPWA	EPWA	WA	WA	EPWA	EPWA	EP	WA	TEPWA	TEPWA	WA	WA
Syngnathus1	EPWA->EP	EPWA->EP	WA->EP	WA->EP	EPWA->EP	EPWA->EP	EP->EP	EP->EP	TEPWA->TEP/TNP	TEPWA->TEP/TNP	WA->TEP/TNP	WA->TEP/TNP
Syngnathus2	EPWA->WA	EPWA->WAEA	WA	EP->WA	EPWA->WAEA	EPWA->WAEA	EP->WA	EP->WA	TEPWA->WAEA	TEPWA->WAEA	WA	WA
Syngnathus3	WA->EA	WAEA->EA and WA	WA->EA	WA->EA	WAEA->EA and WA	WAEA->EA and WA	WA->EA	WA->EA	WAEA->EA and WA	WAEA->EA	WA->EA	WA->EA

WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; TNP: Temperate Northeast Pacific; TEP: Tropical Eastern Pacific; WA: Western Atlantic; EA: Eastern Atlantic; Tet: Tethys Sea.

Table S6. Comparisons of ancestral range estimations in BioGeoBEARS between the three area schemes according to the best-fit biogeographic model estimated for the 'alternative tree' and the 28-tree averaging approach, including and excluding the jump dispersal parameter (*j*). Differences in ancestral range estimates compared to those in Table S5 are shown in bold.

	6 AREAS + Tethys				**7 AREAS + Tethys**				**8 AREAS + Tethys**			
	BAYAREALIKE		**BAYAREALIKE+*j***		**BAYAREALIKE**		**BAYAREALIKE+*j***		**BAYAREALIKE**		**BAYAREALIKE+*j***	
	'Alternative tree' alone	28-tree averaging	'Alternative tree' alone	28-tree averaging	'Alternative tree' alone	28-tree averaging	'Alternative tree' alone	28-tree averaging	'Alternative tree' alone	28-tree averaging	'Alternative tree' alone	28-tree averaging
Dactylopteroidei												
Dactylopterus1	CIPEA	Tet	WAEATet	**WAEATet**	WIOCIPWAEATet	Tet	**WIOCIPCPTATet**	**WIOCIPEATet**	WIOCIPEATet	Tet	WIOCIPCPTATet	**WIOCIPCPTATet**
Mulloidei												
Pseudupeneus1	CIPWA	CIPWA	CIPCPEP	**CIPCPEP**	CIPWA	**CIPWA**	**WIOCIPCPEPTA**	**CIPWA**	CIPWA	**CIPWA**	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Pseudupeneus2	**WA**	EPWA	EP	EP	EPWA	EPWA	EP	**EPWA**	TEPWA	TEPWA	**TEPTNP**	TEPTNP
Pseudupeneus3	WA	WA	WA	WA	WA	WA	WA	WA	WA	**WA**	WA	WA
Mullus1	CIPWA	CIPWA	CIPEP	CIPEP	WATA	WATA	**EPTA**	**WATA**	WATA	WATA	**TEPTA**	**TEPTA**
Mullus2	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA	WA
Mulloidichthys1	CIPEPWA	CIPEPWA	WIOCIPCPEP	WIOCIPCPEP	CIPEPWA	CIPEPWA	**WIOCIPCPEPTA**	**CIPEPWA**	CIPTEPWA	CIPTEPWA	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Mulloidichthys2	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**	**WA**
Callionymoidei												
Synchiropus1	CIPWA	CIPWA	CIPEA	CIPEA	CIPWA	CIPWA	**WATA**	**CIPWA**	CIPWA	CIPWA	WATA	**EATATet**
Synchiropus1	WA	WA	WA	WA	WA	EPWA->WA	WA	WA	WA	WA	WA	WA
Synchiropus2	WA->EP	**WA->EP**	WA->EP	WA->EP	EPWA->EP	**WA->EP**	WA->EP	**EPWA->EP**	TEPWA->TEPTNP	**WA->TEPTNP**	WA->TEPTNP	WA->TEPTNP
Synchiropus3	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA	WA->EA
Callionymus1	CIPCP	CIPCP	CIP	CIP	CIPCPTA	CIPCPTA	CIP	**CIPCPTA**	CIPCPTA	CIPCPTA	**CIP**	CIPCPTA
Syngnathoidei												
Aulostomus1	WIOCIPEPWA	**Tet**	WIOCIPCPEP	WIOCIPCPEP	WIOCIPEPWA	Tet	WIOCIPCPEPTA	**WIOCIPEPTA**	WIOCIPTEPWA	Tet	WIOCIPCPTEPTA	WIOCIPCPTEPTA
Entelurus+Nerophis1	WIOCIPTet	WIOCIPTet	CIPTet	CIPTet	**WIOCIPTet**	WIOCIPCPWAEATATet	Tet	**WIOCIPTet**	WIOCIPTet	**WIOCIPCPTATet**	Tet	**Tet**
Hippocampus	CIPTet	CIPWA	CIPTet	CIP	CIPTet	**CIPTet**	**CIPTet**	**CIPTet**	CIPTet	**CIPTet**	CIPTet	**CPTA**
Hippocampus1a	CIPWA->WA	CIPWA->WA	CIPWA->WA	**CIP->WA**	CIPWA->WA	CIPWA->WA	**CIPWA->WA**	**CIPWA->WA**	CIPWA->WA	CIPWA->WA	CIPWA->WA	**CIPEA->WA**
Hippocampus1b	CIPWA->WA	CIPWA->EPWA	CIPWA->WA	CIPWA->WA	CIPWA->EPWA	CIPWA->EPWA	**CIPWA->WA**	**CIPWA->EPWA**	CIPWA->TEPWA	CIPWA->TEPWA	CIPWA->WA	CIPWA->WA
Hippocampus2	WA->EP	EPWA->WA and EP	WA->EP	WA->EP	EPWA->WA and EP	EPWA->WA and EP	WA->EP	**EPWA->WA and EP**	TEPWA->WA and TEP	TEPWA->WA and TEP	WA->TEP	WA->TEP
Halicampus+Syngnathus1	CIPTet	WIOCIPTet	CIPTet	**CIP**	**WIOCIPTet**	WIOCIPCPWAEATATet	**CIPTet**	**WIOCIPTet**	WIOCIPTet	WIOCIPCPWAEATATet	CIPTet	CIPCPTA
Halicampus+Syngnathus1	WA	WA	WA	**CIPTet**	EPWA	**EPWA**	**WA**	**EPWA**	WAEA	WAEATet	WA	**Tet**
Halicampus+Enneacampus1	WA	WA	WA	WA	EPWA	EPWAEA	WA	**EPWA**	WAEA	WAEA	WA	WA
Halicampus+Cosmocampus1	WA	EPWA	WA	WA	EPWA	EPWA	WA	**EPWA**	TEPWA	TEPWA	WA	WA
Halicampus+Pseudophallus1	EPWA	EPWA	WA	WA	EPWA	EPWA	**WA**	**EPWA**	TEPWA	TEPWA	WA	WA
Syngnathus1	EPWA->EP	EPWA->EP	WA->EP	**EP**	EPWA->EP	EPWA->EP	EP	**EPWA->EP**	TEPWA->TEP/TNP	TEPWA->TEP/TNP	WA->TEP/TNP	WA->TEP/TNP
Syngnathus2	EPWA->WA	**EPWA->WA**	WA	EP->WA	EPWA->WAEA	EPWA->WAEA	EP->WA	**EPWA->EAWA**	TEPWA->WAEA	TEPWA->WAEA	WA	WA
Syngnathus3	WA->EA	**WA->EA**	WA->EA	WA->EA	WAEA->EA and WA	WAEA->EA and WA	WA->EA	**EAWA->EA and WA**	WAEA->EA and WA	WAEA->EA and WA	WA->EA	WA->EA

WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; TNP: Temperate Northeast Pacific; TEP: Tropical Eastern Pacific; WA: Western Atlantic; EA: Eastern Atlantic; Tet: Tethys Sea.

Table S7. Ages (mean values) for the genus *Hippocampus* estimated by previous studies.

Study	**Crown group *Hippocampus* (Ma)**	**First Atlantic colonization (Ma)**	**Second Atlantic colonization (Ma)**
Teske et al. 2007	20 (~13–28)	14.2–15.12	3.1–4.6
Li et al. 2021	23.6 (~20–27)	13.6–15.6	3.6–4.9
This study	14 (~11–16)	6.8–10.33	3.7–6.6

Ma: millions of years.

Supplementary References

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Bannikov A.F. 2014. The systematic composition of the Eocene actinopterygian fish fauna from Monte Bolca, northern Italy, as known to date. Misc. Paleontol. 12:22–34.

Bannikov A.F., Carnevale G. 2012. A long-bodied centriscoid fish from the basal Eocene of Kabardino-balkaria, northern Caucasus, Russia. Naturwissenschaften.

Bannikov A.F., Carnevale G. 2017. Eocene ghost pipefishes (Teleostei, Solenostomidae) from Monte Bolca, Italy. Boll. della Soc. Paleontol. Ital. 56:319–331.

Betancur-R. R., Broughton R.E., Wiley E.O., Carpenter K., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton II J.C., Zhang F., Buser T., Campbell M., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Ortí G. 2013. The Tree of Life and a New Classification of Bony Fishes. PLoS Curr.:0–45.

Betancur-R. R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17.

Blainville H. 1818. Des ichthyolites de Monte Bolca, ou Vestena-Nuova dans le Veronais, in Nouveau. Dict. d'Histoire Appl. aux Arts, a 1' Agric. a l'Economie Rural. Domest. a la Med. 27:334–361.

Blot J. 1980. The ichthyofauna of the deposits of Monte Bolca (Province de Vérone, Italie). Cat. systématique présentat l'état actuel des Rech. Concern. cette faune. Bull Mus nation d'Hist nat, Paris, 4e Ser. sec. C. 2:339–396.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics.

Cantalice K., Alvarado-Ortega J. 2016. Eekaulostomus cuevasae gen. and sp. nov., an ancient armored trumpetfish (Aulostomoidea) from Danian (Paleocene) marine deposits of Belisario Domínguez, Chiapas, southeastern Mexico. Palaeontol. Electron.:1–24.

Carnevale G., Bannikov A.F., Marramà G., Tyler J.C., Zorzin R. 2014. The Pesciara-Monte Postale Fossil-Lagerstätte: 2. Fishes and other vertebrates. Rend. della Soc. Paleontol. Ital. 4:37–63.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol.

Chen W.-J., Santini F., Carnevale G., Chen J.-N., Liu S.-H., LavouÃ© S., Mayden R.L. 2014. New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). Front. Mar. Sci. 1:1–17.

Coates A.G., Obando J.A. 1996. The geologic evolution of the Central American Isthmus. Evolution and Environment in Tropical America. .

Cuevas-García M., Alvarado-Ortega J. 2009. Estudio arqueológico y paleontológico de los fósiles marinos que proceden del sitio de Palenque, Chiapas. Inf. la Prim. Tempor. campo 2008. Inst. Nac. Antropol. e Hist.:54.

Faircloth B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. 32:786–788.

Forey P.L., Yi L., Patterson C., Davies C.E. 2003. Fossil fishes from the cenomanian (upper cretaceous) of Namoura, Lebanon. J. Syst. Palaeontol. 1:227–330.

Friedman M., Carnevale G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. J. Geol. Soc. London.

Gavrilov Y., Shcherbinina E., Oberhänsli H. 2003. Paleocene- Eocene boundary events in the northeastern Peri-Tethys. Geol. Soc. Am. Spec. Pap. 369:147–168.

González-Rodriguez K., Schultze H.-P., Arratia G. 2013. Mesozoic Fishes 5—Global Diversity and Evolution. Arratia G., Schultze H.-P. Wilson M.V.H. (eds.), Mesozoic Fishes 5 – Glob. Divers. Evol. null:457-487. Verlag Dr. Friedrich Pfeil, München.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., Di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol.

Heckel J. 1854. Bericht über die vom Herrn Cavaliere Achille de Zigno hier angelangte Sammlung fossiler Fische. Sitzungsberichte der Kais. Akad. der Wissenschaften, Math. Cl. 11:122–138.

Horvat A. 2003. Paleontology, biostratigraphy and paleoecology of the Miocene diatoms (Bacillariophyta) from Slovenia. .

Hughes L.C., Ortí G., Huang Y., Sun Y., Baldwin C.C., Thompson A.W., Arcila D., Betancur-R. R., Li C., Becker L., Bellora N., Zhao X., Li X., Wang M., Fang C., Xie B., Zhou Z., Huang H., Chen S., Venkatesh B., Shi Q. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. 115:6249–6254.

Jordan D.S., Evermann B.W. 1898. Fishes NORTH AND MIDDLE AMERICA. Bull. United States Natl. Museum.:1241–2183.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol.

Li C., Olave M., Hou Y., Qin G., Schneider R.F., Gao Z., Tu X., Wang X., Qi F., Nater A., Kautt A.F., Wan S., Zhang Y., Liu Y., Zhang H., Zhang B., Zhang H., Qu M., Liu S., Chen Z., Zhong J., Zhang H., Meng L., Wang K., Yin J., Huang L., Venkatesh B., Meyer A., Lu X. 2021. Genome sequences reveal global dispersal routes and convergent developmental mechanism in seahorse evolution. Nat. Commun.

Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48.

Montes C., Cardona A., Jaramillo C., Pardo A., Silva C., Valencia V., Ayala C., Pérez-Angel L., Rodriguez-Parra L., Ramirez V., Niño H. 2015. Middle Miocene closure of the Central American Seaway. Science (80-.). 348:226–229.

Near T.J., Eytan R.I., Dornburg A., Kuhn K.L., Moore J.A., Davis M.P., Wainwright P.C., Friedman M., Smith W.L. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci.

Nelson J.S., Grande T.C., Wilson M.V.H. 2016. Fishes of the World: Fifth Edition. .

O'Dea, Aaron, Lessios H.A., Coates A.G., Eytan R.I., Restrepo-Moreno S.A., Cione A.L., Collins L.S., de Queiroz A., Farris D.W., Norris R.D., Stallard R.F., Woodburne M.O., Aguilera O., Aubry M.-P., Berggren W.A., Budd A.F., Cozzuol M.A., Coppard S.E., Duque-Caro H., Finnegan S., Gasparini G.M., Grossman E.L., Johnson K.G., Keigwin L.D., Knowlton N., Leigh E.G., Leonard-Pingel J.S., Marko P.B., Pyenson N.D., Rachello-Dolmen P.G., Soibelzon E., Soibelzon L., Todd J.A., Vermeij G.J., Jackson J.B.C. 2016. Formation of the Isthmus of Panama. Sci. Adv. 2:1–12.

Orr J.W. 1995. Phylogenetic relationships of gasterosteiform fishes (Teleostei: Acanthomorpha).

Papazzoni C.A., Fornaciari E., Giusberti L., Vescogni A., Fornaciari B. 2017. Integrating shallow benthic and calcareous nannofossil zones: the lower eocene of the Monte Bolca Postale section (northern Italy). Palaios.

Parham J.F., Donoghue P.C.J., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver J.E., Van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L., Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J. 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61:346–359.

Pietsch T. 1978. Evolutionary Relationships of the Sea Moths (Teleostei : Pegasidae) with a Classification of Gasterosteiform Families Author (s): Theodore W . Pietsch Published by : American Society of Ichthyologists and Herpetologists (ASIH) Stable URL : https://w. Copeia.:517–529.

Price M.N., Dehal P.S., Arkin A.P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One.

Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.

dos Reis M., Yang Z. 2017. MCMCTree tutorials. :1–12.

Schlüter M., Steuber T., Parente M. 2008. Chronostratigraphy of Campanian-Maastrichtian platform carbonates and rudist associations of Salento (Apulia, Italy). Cretac. Res. 29:100–114.

Sorbini L. 1981. The Cretaceous fishes of Nardò. I°. Order Gasterosteiformes (Pisces). Boll Mus. Civ. di Stor. Nat. di Verona. 8:1–27.

Teske P.R., Hamilton H., Matthee C.A., Barker N.P. 2007. Signatures of seaway closures and founder dispersal in the phylogeny of a circumglobally distributed seahorse lineage. BMC Evol. Biol. 7.

Volta G. 1796. Ittiolit. Veronese del Mus. Bozziano. 2.

Wheeler A. 1955. Preliminary revision of the fishes of the genus Aulostomus. Ann. Mag. Nat. Hist. 12:613–623.

Žalohar J., Hitij T., Križnar M. 2009. Two new species of seahorses (Syngnathidae, Hippocampus) from the Middle Miocene (Sarmatian) Coprolitic Horizon in Tunjice Hills, Slovenia: The oldest fossil record of seahorses. Ann. Paleontol.

Appendix B

Supplementary Material for Colonization dynamics explain diversity patterns of syngnatharian fishes across marine realms while morphological similarities persist among them

Data repository

All data and code are available in the Dryad digital repository:

https://doi.org/10.5061/dryad.2bvq83bxw.

Supplementary Materials and Methods:

Updates on quality control, phylogenetic inference and divergence times from the previous study Santaquiteria et al. (2021)

Santaquiteria *et al.* (2021) inferred a set of time-calibrated trees for 169 syngnatharian species (plus *Taractichthys longipinnis*, a pelagiarian species used as outgroup) using largely independent gene datasets of UCEs to assess the evolutionary and biogeographic history of this group (see Appendix 1). Building upon the UCE dataset first generated for 112 spp. of syngnatharians by Longo *et al.* (2017), Santaquiteria *et al.* (2021) expanded their taxonomic sampling to include a total of 169 species, assembled 14 phylogenomic matrices and estimated diversification times using MCMCTree (dos Reis and Yang 2019) based on 28 tree input topologies. More specifically, four trees were based on the 75% (932 UCEs) and 90% (346 UCEs) matrices, and 24 trees on genomic subsets of ~90 loci each. All trees were inferred using RAxML v8.2.4 (Stamatakis 2014) and ASTRAL-II (Mirarab and Warnow 2015). After performing additional quality control on the genetic data generated by the previous studies, we removed two individuals from the matrices: *Aulostomus chinensis* (sequenced by Longo *et al.* 2017) and misidentified; actual identity *A. maculatus*) and *Solegnathus cf. robustus* (ambiguous identification).

For this study, we first re-estimated the 14 multispecies-coalescent trees using ASTRAL-III v.5.6.3 (Zhang et al. 2018) after collapsing gene tree branches with low bootstrap support (BS) values (<33%), which overall produced topologies more concordant with those estimated in RAxML (note that Santaquiteria *et al.* (2021) used ASTRAL-II). Unlike the previous study using ASTRAL-II, the ASTRAL-III analyses conducted here resolved the monophyly of all suborders in the large datasets (75 and 90% matrix) as well as in subsets S01 and S09. To assess topological disparity, we estimated tree space plots for the new 28 trees (14 trees inferred in RAxML and 14 in ASTRAL-III) using a multidimensional scaling (MDS) visualization implemented in the R package *treespace* (Jombart et al. 2017).

To account for uncertainty in divergence times, we re-calibrated the 28 trees using MCMCTree and RelTime (Tamura et al. 2012), which can handle genome-scale datasets. We included four additional fossil calibrations, for a total of 14 calibration points (10 fossils, 3 geological, and one secondary root calibration point, see justifications below), and pruned a total of four undetermined species from the trees (*Synchiropus* sp. USNM431692, Callionymidae sp. CSIROGT3084, *Callionymus* sp1 CSIROGT895, and *Callionymus* sp2 CSIRODGQ0689). This resulted in a total of 56 backbone time trees (28 MCMCTree trees and 28 in RelTime trees) with 163 syngnatharian species and one outgroup. MCMCTree, which employs a Bayesian framework (dos Reis and Yang 2019), is implemented in PAML v4.9 (Yang 2007), while RelTime, which utilizes a maximum likelihood framework, is implemented in MEGAX (Stecher et al. 2020). Our main motivation for implementing these alternative dating approaches is that MCMCTree requires a constant birth-death prior, which assumes that the rates of speciation and extinction remain constant over time and may fail to capture the complex dynamics of diversification such as bursts of speciation in different lineages. By contrast, RelTime does not rely on a specific tree prior, thereby allowing for speciation rate variation across branches in a phylogeny (Tamura et al. 2012, 2018; Mello et al. 2021). For MCMCTree, we ran the 75% and 90% matrices for 20–50 million generations and the genomic subsets for 4–12 million generations until convergence was reached based on effective sampling size (ESS) values >200. We used the approximate likelihood method and the HKY85 model. Prior parameters for the MCMCTree runs were as follows: independent rate relaxed-clock model, BDparas: 1, 1, 0.25; kappa_gamma: 6, 2; alpha_gamma: 1, 1; rgene_gamma: 2, 200, 1; sigma2_gamma: 2, 5, 1. We conducted two independent runs for each data set. To check for convergence, we used Tracer v1.7.1 (Rambaut et al. 2018) to examine trace plots and ESS values for each parameter, after a 10% burn-in. To calibrate the trees inferred using RAxML and ASTRAL-III, we used two distinct approaches in RelTime. We used *RelTime-Branch Lengths* with a Max Relative Rate Ratio of 20 for the RAxML trees, using their topology as input. On the other hand, for ASTRAL-III we used *RelTime-ML* with the GTR+I model and the default setting to also optimize branch lengths. We input the ASTRAL-III topologies along with a concatenated DNA alignment comprising the genes used to infer the corresponding trees in each case.

To expand the taxonomic sampling of the Santaquiteria *et al.* (2021) dataset, we downloaded all the Cytochrome Oxidase I (COI) sequences available from National Center for Biotechnology Information (NCBI) and the Barcode of Life Data System (BOLD) repositories (8137 individuals from 321 species). We conducted a series of quality control steps to curate these downloaded sequences. We parsed the sequences through BOLD using the python script *bold_identification* (Yang et al. 2020) to identify possible cases of contamination and misidentification. After this step, we removed a total of 1256 sequences. We then conducted an initial assessment of phylogenetic relationships using FastTree-2 (Price et al. 2010) to select one individual per monophyletic species (318 species). We mined COI sequences from the UCE raw data and aligned them with additional COI sequences retrieved from NCBI and BOLD using MACSE v2.03 (Ranwez et al. 2018). Finally, we inferred a maximum likelihood tree in RAxML using the tree generated with the full UCE dataset (75% matrix, 932 UCEs) and constructed in RAxML as backbone constraint, and removed 60 misplaced species and 98 duplicates with the UCE data, yielding 160 (62 NCBI and 98 BOLD) additional species (not examined in Santaquiteria *et al.* (2021). See Appendix 1 for details on the molecular data used in this study.

Using the 56 backbone trees as reference, we then inferred the placement of the newly added COI sequences in all trees via backbone constraint maximum likelihood (ML) searches in RAxML as described above. Next, we time-calibrated the expanded trees with 323 species using treePL (Smith and O'Meara 2012) via *congruification* (Eastman et al. 2013), an approach that uses fixed secondary calibration points obtained from all shared nodes in the reference backbone phylogenies. We used the "congruify" function implemented in the R package *geiger* (Harmon et al. 2008). We treated the expanded RAxML tree based on the 75% UCE matrix and dated in RelTime as the "master tree" hereafter, with the MCMCTree counterpart called the "alternative tree" hereafter. We also compared the topologies and age estimates of the trees inferred with the two dating methods (MCMCTree and RelTime) using density trees and boxplots as implemented in *DensiTree* (Bouckaert 2010) and *ggplot2* (Wickham 2008), respectively.

Subsequent to completing all downstream analyses, which included extensive run times (up to six months) for GeoHiSSE, BioGeoBears, and BAMM, a new syngnatharian phylogeny was published by Stiller *et al.* (2022), focusing on the family Syngnathidae. In their study, they sequenced a total of 238 species using UCE data, which also built upon the Longo *et al.* (2017) dataset comprising 113 taxa. Among the 238 species, 188 were found to be shared with our study (116 with UCE data and 72 with COI data). Within our 72 COI sequences, 26 species exhibited conflicting phylogenetic placements when compared to the Stiller *et al.* (2022) tree. These inconsistencies ranged from modest discrepancies in 17 species to more pronounced incongruences in 9 species (see details in Table S1). However, it is worth noting that the 26 species with incongruent placements were not included in our morphological matrix, and thus any potential effects resulting from these discrepancies would only impact biogeographic and diversification analyses. To assess the sensitivity of these analyses to the placement of the 26 species, we performed additional tests. Specifically, we pruned out these species from our "master tree" and then re-ran the ancestral range reconstructions and diversification analyses, using DR statistics. In all cases, the analyses conducted with these pruned trees produced similar results, indicating that there were no meaningful effects driven by these topology differences (Figs. S37 and S38).

Table S1. Taxa with conflicting phylogenetic placement between this study and Stiller *et al.* (2022).

Taxa with modestly incongruent placements	Taxa with highly incongruent placements
Syngnathidae_Microphis_jagorii_BOLD_ANGBF46344_19	Syngnathidae_Hippichthys_cyanospilos_BOLD_SBF396_11
Syngnathidae_Apterygocampus_epinnulatus_BOLD_GBMIN96762_17	Syngnathidae_Leptoichthys_fistularius_BOLD_FMVIC693_08
Syngnathidae_Bhanotia_fasciolata_BOLD_ANGBF46233_19	Syngnathidae_Stipecampus_cristatus_NCBI_KY066151.1
Syngnathidae_Nannocampus_pictus_BOLD_GBMIN131828_17	Syngnathidae_Halicampus_macrorhynchus_BOLD_ANGBF46318_19
Syngnathidae_Festucalex_cinctus_BOLD_ANGBF46315_19	Syngnathidae_Micrognathus_natans_NCBI_KY066131.1
Syngnathidae_Festucalex_scalaris_NCBI_KY066096.1	Syngnathidae_Micrognathus_andersonii_USNMAG9RQ99
Syngnathidae_Penetopteryx_nanus_BOLD_GBMIN126168_17	Syngnathidae_Halicampus_grayi_BOLD_FOAL742_10
Syngnathidae_Syngnathus_caribbaeus_BOLD_BZLWD277_07	Syngnathidae_Urocampus_carinirostris_BOLD_ANGBF46556_19
Syngnathidae_Lissocampus_caudalis_BOLD_ANGBF46331_19	Syngnathidae_Hippocampus_camelopardalis_BOLD_ABRMF015_06
Syngnathidae_Lissocampus_runa_BOLD_ANGBF46332_19	
Syngnathidae_Hippocampus_mohnikei_BOLD_GBMIN123422_17	
Syngnathidae_Hippocampus_guttulatus_BOLD_ABRMF034_06	
Syngnathidae_Hippocampus_pontohi_BOLD_ANGBF46117_19	
Syngnathidae_Hippocampus_bargibanti_BOLD_ANGBF45947_19	
Syngnathidae_Hippocampus_denise_BOLD_ANGBF45993_19	
Syngnathidae_Hippocampus_spinosissimus_BOLD_ANGBF55091_19	
Syngnathidae_Hippocampus_capensis_BOLD_ABRMF078_06	

Calibration points

Based on recommendations by Parham *et al.* (2012) we based lower bounds for each of the calibration schemes on minimum age constraints (i.e., youngest fossil ages). To estimate divergence times in MCMCTree, most calibrations used uniform distributions. The only exception is crown Syngnatharia, for which we used a Cauchy distribution. Soft upper bounds used mostly fossils placed deeper in the tree (Table S2). To estimate ages using RelTime we used distribution densities based on the algorithm proposed by Hedman (Hedman 2010), which is implemented in R (Lloyd et al. 2016). This approach, which has been applied in multiple previous studies (e.g., Harrington et al. 2016; Alfaro et al. 2018; Rabosky et al. 2018), uses a list of fossil outgroup age records based on the oldest minima to produce a probable distribution of the origin of a given clade. Here, we used the outgroup age sequence of fossils that were older or the same age of the oldest fossil placed on the focal clade following Harrington *et al.* (2016), Alfaro *et al.* (2018), and Friedman *et al.* (2019): 247.1 Ma, Holostei, †*Watsonulus eugnathoides*; 236.0 Ma, †*Prohalecites porroi*; 221.0 Ma, †Pholidophoridae, †*Knerichthys bronni*; 193.81 Ma, †*Dorsetichthys bechei*; 181.7 Ma, †*Leptolepis coryphaenoides*; 166.1 Ma,†Ichthyodectiformes, †*Occithrissops willsoni*; 151.2 Ma, Elopomorpha, †*Anaethalion zapporum*; 150.94 Ma, Otocephala, †*Tischlingerichthys viholi*; 150.94 Ma, non-eurypterygian Euteleostei †*Leptolepides haerteisi*; 125.0 Ma, Aulopiformes, †*Atolvorator longipectoralis*; 98.0 Ma, Lampridiformes, †*Aipichthys minor*; 98.0 Ma, Holocentroidei, †*Stichocentrus liratus*. The Hedman (Hedman 2010) method requires a maximum hard bound, therefore we used the †*Discoserra* fossil (322.8 Ma, stem neopterygian). We then extracted the 95% confidence interval from each fossil node and calculated the mean and standard deviation in order to implement log-normal distributions for divergence time analysis. For both divergence estimation methods (RelTime and MCMCTree), we excluded incongruent calibrations from input topologies on a case-by-case basis. As previously mentioned, following the completion of all our analyses, the study conducted by Stiller *et al.* (2022) was published. In their research, the authors introduced five additional fossil calibrations that were not utilized here: †*Maroubrichthys serratus* (Stem *Microphini*+*Doryramphini*+*Maroubra*+*Heraldia*), †*Doryrhamphus* sp. (Stem *Microphini*+*Doryrhamphini*), †*Hipposyngnathus neriticus* (Stem *Nerophis*), †*Hippotropiscis frenki* (Stem pygmy pipehorses), †*Hippohaliichthys edis* (Stem *Haliichthys*+*Halicampus grayi*+*Trachyrhamphus*+*Filicampus*). Nevertheless, ages estimated using both Santaquiteria *et al.* (2021) and this study were largely congruent with those inferred by Stiller *et al.* (2022) using BEAST (Fig. S4).

(1) Root. MRCA: *Syngnathus louisianae*, *Taractichthys longipinnis*. See Santaquiteria *et al.* (2021) for a justification of the calibration. Prior setting MCMCTree: B(0.920,1.035,1e-300,1e-300) (crown calibration). A drawback of Reltime is that it does not allow the use of root calibrations, therefore this calibration point was excluded from those analyses.

(2) Syngnatharia. MRCA: *Syngnathus louisianae*, *Eurypegasus draconis*. See Santaquiteria *et al.* (2021) for a justification of the calibration. Prior setting MCMCTree: L(0.836,0.01,0.0001,1e-300) (crown calibration). Due to the unfeasibility of setting up a root calibration in Reltime, we substituted the "Syngnatharia" calibration point with a uniform distribution. To establish the *mintime* and *maxtime* parameters, we determined the 95% CI for each subset using MCMCTree.

(3a) Solenostomidae. MRCA: *Solenostomus cyanopterus*, *Syngnathus louisianae*. Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=48.5, mean=2.244, stddev=0.891. See Santaquiteria *et al.* (2021) for a justification of the calibration. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 48.5.

(3b) Syngnathidae. MRCA: *Syngnathus louisianae*, *Solenostomus cyanopterus*. Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=48.5, mean=2.244, stddev=0.891. See Santaquiteria *et al.* (2021) for a justification of the calibration. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 48.5.

(4) *Hippocampus*. MRCA: *Hippocampus abdominalis*, *Hippocampus kuda*. Prior setting MCMCTree: B(0.116,0.485,1e-300,0.05); RelTime: log-normal distribution, offset=11.6, mean=2.452, stddev=0.932 (crown calibration). See Santaquiteria *et al.* (2021) for a justification of the calibration. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 48.5, 11.6. Comment: For this calibration point we used the †*H. sarmanticus* fossil instead of *H. slovenicus* because our taxonomic sampling lacks some pygmy seahorse species. However, these two fossil species have the same age, thus the results were not affected.

(5) Aulostomoidea. MRCA: *Aulostomus maculatus*, *Aeoliscus strigatus*. Prior setting MCMCTree: B(0.615,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=61.5, mean=2.010, stddev=0.888. See Santaquiteria *et al.* (2021) for a justification of the calibration. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 61.5.

(6) Fistulariidae. MRCA: *Fistularia corneta*, *Aulostomus maculatus*. Prior setting MCMCTree: B(0.485,0.615,1e-300,0.05); RelTime: log-normal distribution, offset=48.5, mean=1.906, stddev=0.924. See Santaquiteria *et al.* (2021) for a justification of the calibration. Comment: it should be noted that this calibration point used †*Urosphenopsis dubius* from Bolca instead of the older †*Urosphenopsis sagitta* of the Danatinsk Formation of Turkmenistan, which was recently used by Stiller *et al.* (2022). The age of this fossil (54.17 Ma) is, however, younger than our estimates of the age of total group of Fistulariidae obtained from different analyses (mean 71.5–58.04 Ma RelTime trees and mean 58.84–55.27 Ma MCMCTree trees; except for two trees MCMC_ASTRAL_S11 [52.81 Ma] and RelTime_ASTRAL_S03 [54.01 Ma]). Therefore, this omission produced no major age conflicts. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 61.5, 48.5.

(7) Pegasidae (updated calibration). MRCA: *Pegasus volitans, Dactylopterus volitans*. The age of this fossil was updated from Santaquiteria *et al.* (2021), previously dated at 48.5 Ma. Hard lower bound: †*Ramphosus rosenkrantzi* (Volta 1796; Nielsen 1960). Diagnosis and phylogenetic placement: †*Ramphosus rosenkrantzi* (family †Ramphosidae) shares the following character states with members of Pegasidae: head encased by thick bony plates; nasals fused along the midline forming a rostrum bearing one to several series of tubercles and spines; preopercle enormously expanded medially; hypurals plus parhypural and uroneural consolidated into a bony plate; presence of a large rayless pterygiophore in the first dorsal fin; second dorsal and anal fins opposite and almost equal in length (Pietsch 1978; Calzoni et al. 2023). Stratigraphic horizon and locality: Early Eocene, Upper Ypresian, Fur Formation, Denmark (Nielsen 1960). Paleogeographic domain: Boreal realm and Tethys. Absolute age estimate: 54 Ma ((Schmitz et al. 2004; Storey et al. 2007); see comment below). Soft upper bound: 83.6 Ma (based on †*G. zuppichinii*; see above). Prior setting MCMCTree: B(0.540,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=54, mean=2.096, stddev=0.918. Comment: the age constraint of the Fur Formation is provided by the Palaeocene/Eocene boundary in the underlying Ølst Formation and radiometric dating of two ash layers within the formation that originated 55 and 54 Ma, respectively (Schmitz et al. 2004; Storey et al. 2007); the minimum age of 54 Ma is used for this calibration. While this should be treated as a stem calibration (with MRCA *Eurypegasus draconis*, *Pegasus volitans*), it is instead applied as crown calibration one node below due to limitations of RelTime and MCMCTree. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 54.

(8) Callionymoidei (new calibration). MRCA: *Draconetta xenica*, *Callionymus scaber*. Hard lower bound: †*Gilmourella minuta* (Carnevale and Bannikov 2019). Diagnosis and phylogenetic placement: the placement of this fossil (family-level *incertae sedis*) within the suborder Callionymoidei, sister to the family Callionymidae, is sustained by a series of morphological characters (Carnevale and Bannikov 2019). Features shared with Callionymidae and Draconettidae: infraorbital series reduced to the lachrymal; palatine rigidly attached to the ectopterygoid; metapterygoid absent; symplectic strongly developed; seven abdominal vertebrae; penultimate vertebra bearing expanded and plate-like neural and haemal spines; ribs absent; supraneural absent; basipterygia transversely oriented; and pelvic-fin base well in advance of pectoral fin. Features shared with only Callionymidae: upper jaw strongly protractile; endopterygoid absent; hypural and parhypural consolidated into a single plate; haemal spine of the penultimate vertebra fused to the centrum; and anal-fin rays mostly unbranched. Unique features that differ with those observed in extant callionymids and draconettids: short body; large and elongated head; well-developed snout; thin opercular bones; preopercle with a posterior blunt spine; interopercle ribbon-like; opercle subtriangular with fimbriated posterior margin; subopercle elongate and distally pointed; anteroposteriorly compact vertebrae; 14 short rays in the caudal fin; spinous dorsal-fin absent; five unbranched rays in dorsal and anal fins; and short pelvic fins. Stratigraphic horizon and locality: Late Early Eocene, Upper Ypresian, Middle Cuisian, Pesciara site, Bolca Lagerstätte, northeastern Italy (Carnevale and Bannikov 2019). Paleogeographic domain: western Tethys. Absolute age estimate: 48.5 Ma ((Papazzoni et al. 2014; Friedman and Carnevale 2018); see comment below). Prior setting MCMCTree: B(0.485,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=48.5, mean=2.244, stddev=0.891. Comment: the age constraint of Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 is used for this calibration. While this should be treated as a total group calibration (with MRCA *Synchiropus calauropomus*, *Callionymus scaber*), it is instead applied as crown calibration one node below due to limitations of RelTime and MCMCTree. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 48.5.

(9) Centriscidae (new calibration). MRCA: *Aeoliscus strigatus*, *Macroramphosus gracilis*. Hard lower bound: †*Aeoliscoides longirostris* (Blot 1980). Diagnosis and phylogenetic placement:

based on the short diagnosis given by Blot (Blot 1980) on a single specimen, this fossil seems to a relative of *†Aeoliscus heinrichi* (see below for morphological characters supporting this relationship; Parin and Micklich 1996). The assignment of *†Aeoliscoides longirostris* as total group Centriscidae (*sensu* (Pietsch 1978); *Centriscus*+*Aeoliscus*) is supported by its overall outline of the body, characterized by a long and tubular snout and a thick and almost horizontally oriented dorsal-fin spine; and a body armor separated into dorsal and ventral portions (Parin and Micklich 1996). Stratigraphic horizon and locality: Late Early Eocene, Upper Ypresian, Middle Cuisian, Pesciara site, Bolca Lagerstätte, northeastern Italy (Carnevale and Bannikov 2019). Paleogeographic domain: western Tethys. Absolute age estimate: 48.5 Ma ((Papazzoni et al. 2014; Friedman and Carnevale 2018); see comment below). Prior setting MCMCTree: B(0.485,0.615,1e-300,0.05); RelTime: log-normal distribution, offset=48.5, mean=1.906, stddev=0.924. Comment: the age constraint of Bolca is assigned to the interval NP14 and SBZ11, dated at 50.5-48.5 Ma (Friedman and Carnevale 2018); the minimum age of 48.5 is used for this calibration. While this should be treated as a total group calibration (with MRCA *Centriscus*, *Aeoliscus*), it is instead applied as crown calibration one node below at for Centriscidae *sensu lato* due to limitations of RelTime and MCMCTree. It should be noted that we did not include the older †*Gerpegezhus paviai* (family Gerpegezhidae) fossil for this calibration point as we consider it to be the sister group of extant Centriscidae (Gavrilov et al. 2003; Bannikov and Carnevale 2012) and not an stem of *Aeoliscus*+*Centriscus* as suggested by Stiller *et al.* (2022). Although there is some discrepancy about its placement, the age of this fossil (55.8 Ma) is still younger than our estimates of the age of crown Centriscidae obtained from different analyses (mean 67.52–54.8 Ma RelTime trees and mean 63.8–51.46 Ma MCMCTree trees). Therefore, this omission produced no major age conflicts. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 61.5, 48.5.

(10) *Mullus* (new calibration). MRCA: *Upeneichthys stotti*, *Mullus auratus*. Hard lower bound: *†Mullus* sp. (Carnevale et al. 2006). Diagnosis and phylogenetic placement: the placement of this fossil as total group *Mullus* is supported by a convex and obliquely directed anterior profile of the head; a short and blunt snout, similar to that of *M. surmuletus*; and a toothless upper jaw, typical of *Mullus* adults (Caldwell 1962). The shape of frontal, supraoccipital, mesethmoid, lateral ethmoid, nasal, and parasphenoid bones is also similar to that seen in *Mullus*. Stratigraphic horizon and locality: Lower Sarmatian, Volhynian, Middle Miocene, Tsurevsky Formation, western North Caucasus, Russia (Beluzhenko 2002). Paleogeographic domain: eastern Paratethys. Absolute age estimate: 13 Ma (Carnevale et al. 2006). Prior setting MCMCTree: B(0.130,0.836,1e-300,0.05); RelTime: log-normal distribution, offset=13, mean=2.587, stddev=0.924. While this should be treated as a total group calibration (with MRCA *Mullus argentinae*, *Mullus auratus*), it is instead applied as crown calibration one node below due to limitations of RelTime and MCMCTree. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 13.

(11) *Nerophis* (new calibration). MRCA: *Nerophis ophidion*, *Entelurus aequoreus*. Hard lower bound: *†Nerophis zapfei* (Bachmayer 1980). Diagnosis and phylogenetic placement: the placement of this middle Miocene species within the genus *Nerophis* is supported by the following characters (see (Bachmayer 1980): a remarkably elongate body with straight ventral margin; head short, representing more than one-fifteenth of standard length; opercular crest absent; dorsomedial crest of the snout absent; pectoral crests scarcely developed; spines on the scutes absent. Material from the Sarmatian of Moldova referred to this species also exhibits a single dorsal fin containing 43 rays (Popov 2017). Stratigraphic horizon and locality: laminated marls of the Leitha Limestone, St. Margarethen, Eisenstadt-Sopron Basin, Burgenland, Austria. Paleogeographic domain: central and eastern Paratethys. Absolute age estimate: 13.5 Ma. Prior setting MCMCTree: B(0.135,0.485,1e-300,0.05); RelTime: log-normal distribution, offset=13.5, mean=2.438, stddev=0.928. Comment: the age of the laminated marls of the Leitha Limestone of St. Margarethen has been established based on calcareous nannoplankton content, which indicates zone NN5b, around the Langhian-Serravallian boundary, approximately between 14.0 and 13.5 Ma, corresponding to the late Badenian of the Paratethys stratigraphy (Schmid et al. 2001). The minimum age of 13.5 Ma is used for this calibration. While this should be treated as a total group calibration for *Nerophis*, it is instead applied as crown calibration one node below due to limitations of RelTime and MCMCTree. Outgroup sequence ages: 247.1, 236.0, 221.0, 193.81, 181.7, 166.1, 151.2, 150.94, 150.94, 125, 98.0, 98.0, 83.6, 48.5,13.5.

Geologic calibrations based on trans-isthmian geminate taxa

Several geminate species pairs in Syngnatharia–including terminal clades on either side of the Panama Isthmus (Jordan and Evermann 1898)–are used as geologic calibrations in our tree. The final closure of the Isthmus of Panama, separating the eastern Pacific (EP) and the Caribbean Sea basins, is an unresolved debate. Although age constrains of 3.5–2.8 Ma have been traditionally used to calibrate phylogenies with this formation (e.g., Coates and Obando 1996), recent studies have challenged the timing of the final closure of the Panama Isthmus. More specifically, Montes *et al.* (2015) proposed that the final closure of the Central American Seaway occurred during the Middle Miocene, which would place it at 15–13 Ma. O'Dea *et al.* (2016), however, continue to maintain support for a younger estimate of 2.8 Ma (Pleistocene). Given these ongoing controversies, we set a lower hard bound of 2.8 Ma (with Cauchy distributions), which reflects an undisputed minimum geologic age for this event, without the implementation of upper bounds as priors in the calibrations. Prior setting MCMCTree: L(0.028,0.1,1,1e-300); RelTime: MinTime=2.8.

(12) Geminate *Aulostomus*. MRCA: *Aulostomus maculatus*, *Aulostomus chinensis*. Comment: this calibration point was only used for *congruification* analyses in TreePL (see main text).

(13a) Geminate *Mulloidichthys*. MRCA: *Mulloidichthys martinicus*, *Mulloidichthys dentatus*. Comment: this calibration and calibration 13b below are used simultaneously to account for topological uncertainties regarding the sister species of *Mulloidichthys martinicus* in different trees.

(13b) Geminate *Mulloidichthys*. MRCA: *Mulloidichthys martinicus*, *Mulloidichthys vanicolensis*. Comment: this calibration and calibration 13a above are used simultaneously to account for topological uncertainties regarding the sister species of *Mulloidichthys martinicus* in different trees.

(14) Geminate *Hippocampus*. MRCA: *Hippocampus ingens*, *Hippocampus reidi*.

Table S2. Calibration priors used to estimate divergence times in MCMCTree and RelTime.

MRCA	Absolute Age (Ma)	MCMCTree Distribution and Parameters	RelTime Distribution and Parameters
(1) *Syngnathus louisianae*, *Taractichthys longipinnis*	92-103.5	Uniform (Hard upper and hard lower bound): B(0.920,1.035,1e-300,1e-300)	N/A
(2) *Syngnathus louisianae*, *Eurypegasus draconis*	83.6	Cauchy (Hard lower bound): L(0.836,0.01,0.0001,1e-300)	Uniform
(3) *Solenostomus cyanopterus*, *Syngnathus louisianae*	48.5	Uniform (Soft upper and hard lower bound): B(0.485,0.836,1e-300,0.05)	Log-normal: offset=48.5, mean=2.244, stddev=0.891
(4) *Hippocampus abdominalis*, *Hippocampus kuda*	11.6	Uniform (Soft upper and hard lower bound): B(0.116,0.485,1e-300,0.05)	Log-normal: offset=11.6, mean=2.452, stddev=0.932
(5) *Aulostomus maculatus*, *Aeoliscus strigatus*	61.5	Uniform (Soft upper and hard lower bound): B(0.615,0.836,1e-300,0.05)	Log-normal: offset=61.5, mean=2.010, stddev=0.888
(6) *Fistularia corneta*, *Aulostomus maculatus*	48.5	Uniform (Soft upper and hard lower bound): B(0.485,0.615,1e-300,0.05)	Log-normal: offset=48.5, mean=1.906, stddev=0.924
(7) *Pegasus volitans*, *Dactylopterus volitans*	54	Uniform (Soft upper and hard lower bound): B(0.54,0.836,1e-300,0.05)	Log-normal: offset=54, mean=2.096, stddev=0.918
(8) *Draconetta xenica*, *Callionymus scaber*	48.5	Uniform (Soft upper and hard lower bound): B(0.485,0.836,1e-300,0.05)	Log-normal: offset=48.5, mean=2.244, stddev=0.891
(9) *Aeoliscus strigatus*, *Macroramphosus gracilis*	48.5	Uniform (Soft upper and hard lower bound): B(0.485,0.615,1e-300,0.05)	Log-normal: offset=48.5, mean=1.906, stddev=0.924
(10) *Upeneichthys stotti*, *Mullus auratus*	13	Uniform (Soft upper and hard lower bound): B(0.13,0.836,1e-300,0.05)	Log-normal: offset=13, mean=2.587, stddev=0.924
(11) *Nerophis ophidion*, *Entelurus aequoreus*	13.5	Uniform (Soft upper and hard lower bound): B(0.135,0.485,1e-300,0.05)	Log-normal: offset=13.5, mean=2.438, stddev=0.928

(12) *Aulostomus maculatus*, *Aulostomus chinensis*	2.8	Cauchy (Hard lower bound): L(0.028,0.1,1,1e-300)	Hard lower bound: MinTime=2.8
(13a) *Mulloidichthys martinicus*, *Mulloidichthys dentatus*	2.8	Cauchy (Hard lower bound): L(0.028,0.1,1,1e-300)	Hard lower bound: MinTime=2.8
(13b) *Mulloidichthys martinicus*, *Mulloidichthys vanicolensis*	2.8	Cauchy (Hard lower bound): L(0.028,0.1,1,1e-300)	Hard lower bound: MinTime=2.8
(14) *Hippocampus ingens*, *Hippocampus reidi*	2.8	Cauchy (Hard lower bound): L(0.028,0.1,1,1e-300)	Hard lower bound: MinTime=2.8

Ma: millions of years.

Table S3. List of randomly pruned tips from 12 species pairs with a terminal branch length <0.5 million years old, which may indicate taxonomic over-splitting (i.e., 'T-like' terminal nodes).

Species pruned
Callionymidae_Callionymus_scaber_CEO64
Callionymidae_Callionymus_sokonumeri_NCBI_KY371215.1
Callionymidae_Callionymus_valenciennei_NCBI_JF952835.1
Centriscidae_Notopogon_xenosoma_BOLD_FMVIC157_08
Mullidae_Parupeneus_macronemus_NCBI_MF123976.1
Mullidae_Upeneus_luzonius_NCBI_KY675471.1
Mullidae_Upeneus_suahelicus_NCBI_KP293705.1
Syngnathidae_Corythoichthys_conspicillatus_BOLD_MBFA990_07
Syngnathidae_Microphis_manadensis_BOLD_ANGBF46347_19
Syngnathidae_Pseudophallus_mindii_BOLD_BSFFA678_07
Syngnathidae_Syngnathus_euchrous_BOLD_ANGBF54497_19
Syngnathidae_Vanacampus_vercoi_BOLD_GBMIN126172_17

Biogeographic history and timing of regional colonization

We estimated ancestral ranges for the new syngnatharian phylogeny set with 323 species using the R package *BioGeoBEARS* (Matzke 2013) following the approach outlined in Santaquiteria *et al.* (2021). We implemented a seven-region biogeographic scheme (based on Spalding et al. 2007; Kulbicki et al. 2013) plus the now extinct Tethys Sea (Tet): Western Indian Ocean (WIO), Central Indo-Pacific (CIP), Central Pacific (CP), Temperate Australasia (TA), Tropical Eastern Pacific (TEP), Western Atlantic (WA), and Eastern Atlantic (EA). We built a presence/absence matrix by coding each extant species according to their geographic ranges primarily based on the International Union for Conservation of Nature (IUCN) Red List (IUCN 2021) and Ocean Biogeographic Information System (OBIS 2021) databases. We also used paleogeographic domain information as biogeographic constraints based on data obtained from the 10 fossils used to calibrate our trees. We evaluated 12 biogeographic models combining DEC (Ree and Smith 2008), DIVA (Ronquist 1997), and BAYAREA (Landis et al. 2013), with and without the jump-dispersal or founder-speciation event (*j*) (Matzke 2014) and the dispersal matrix power exponential (*w*) parameters (Dupin et al. 2017). We analyzed each model using three time slices based on two major geological events: (i) prior to the closure of the Tethys Seaway (92–12 Ma; Steininger and Rögl 1979; Adams et al. 1983; Rögl 1999), (ii) after the closure of the Tethys Seaway and prior to the last rising of the Panama Isthmus (12–2.8 Ma; O'Dea et al. 2016), and (iii) after the last rising of the Panama Isthmus (2.8–0 Ma). We also accounted for connectivity between areas by implementing three different dispersal probability categories: 1 (high connectivity), 0.5 (intermediate separation), and 0.0001 (wide separation/no connectivity). As the summary phylogeny, we used the "master tree". We then selected the best-fit biogeographic model based on the Akaike Information Criterion scores corrected for small sample size (AICc). Based on the sensitivity of biogeographic inferences previously identified for the group (see Santaquiteria *et al.* (2021), we accounted for both topological uncertainty and the inclusion/exclusion of the *j* parameter in biogeographic models (Ree and Sanmartín 2018; Klaus and Matzke 2020; Matzke 2022). We summarized ancestral range estimates from all 28 RelTime trees by overlying average probabilities across compatible nodes on the "master tree" (Matzke 2019) using the best-fit model with and without the *j* parameter.

We assessed the center of accumulation and time-for-speciation hypotheses by estimating the frequency and timing of colonization events between the seven areas. To accomplish this, we conducted biogeographic stochastic mapping (BSM) analyses by simulating 100 stochastic histories on the "master tree" based on the best-fit biogeographic models (Dupin et al. 2017). From each map, we extracted all states at every node and tip and identified all individual colonizations, their descendants, and colonization timing for each given region. For each region, we then calculated the number of cumulative lineages (due to a combination of colonization and speciation), number of independent colonization events, immigration and emigration rates, speciation rates, and extirpation rates across time by averaging over 100 histories. This approach follows the biogeographic methodology developed by Xing and Ree (Xing and Ree 2017) and implemented previously (e.g., Ding et al. 2020; Miller et al. 2022). We repeated all biogeographic analyses using the "alternative tree" (expanded tree inferred using the 932-UCE backbone tree and dated in MCMCTree) and all 28 MCMCTree trees.

Diversification rates among regions

To assess the influence of geographic distribution on lineage diversification dynamics (testing the *in situ* diversification rates hypothesis), we estimated diversification rates (i) between the three major oceanic realms (IP, EP, and Atl.), and (ii) within the Indo-Pacific by splitting it into four subareas (Western Indian Ocean, WIO; Central Indo-Pacific, CIP; Central Pacific, CP; Temperate Australasia, TA; Spalding et al. 2007). We used the geographic state-dependent diversification models implemented in the *hisse* package (Beaulieu and O'Meara 2016; Caetano et al. 2018), as well as two approaches that estimate lineage-specific diversification rates: the model-based Bayesian Analysis of Macroevolutionary Mixtures (BAMM; Rabosky 2014) and the non-parametric DR statistic (Jetz et al. 2012). To account for topological and divergence time uncertainty, we conducted these analyses in an MCMC framework using independently the set of 28 calibrated phylogenies in RelTime and MCMCTree. We also assessed the sensitivity of diversification rate analyses to 12 terminal nodes with shallow divergences (i.e., ‘T-like’ terminal nodes with branch lengths <0.5;) which may indicate taxonomic over-splitting and can “force” models to fit extremely fast rates (see also Rabosky 2016). We randomly pruned a tip from each ‘T-like’ clade (Table S3) and ran all analyses twice: using non-pruned trees with all 323 tips and pruned trees with 311 tips.

For geographic-dependent analyses, we fitted 24 different area-independent and area dependent models, including GeoSSE (no hidden states; Goldberg et al. (2011) and GeoHiSSE Caetano et al. (2018) both with and without the *j* parameter (Table S4). These models vary on whether or not diversification is linked to the focal area, if the area contractions are separated from lineage extinction or extirpation, and if the hidden states or unobserved characters are considered. A drawback of GeoHiSSE is that it only allows comparisons of two regions at a time. To overcome this limitation, we then conducted multiple pairwise comparisons by comparing the focal area versus the remaining areas (i.e., IP vs. [EP+Atl.], EP vs. [IP+Atl.], and Atl. vs. [IP+EP]). We built a matrix for each of the comparisons by coding each species as 1 if it is present in the focal area, 2 if it is present in any of the remaining areas, and 0 if it is widespread. We used Global Biodiversity Information Facility (GBIF 2022) records for each species and estimated species richness within each region based on the Marine Ecoregions of the World system (MEOW; Spalding et al. 2007) and plotted in QGIS (QGIS Development Team 2009). We then calculated the sampling fractions for each region across all comparisons (IP: 47.85%, EP: 65.52%, and Atl.: 65.96%). We calculated the AIC values for each of the models and averaged the best three models (equivalent to the 90–95% accumulative weight) using Akaike weights. We quantified the model-averaged tip-rated (speciation, extinction, and diversification) for each species and tree.

Table S4. Area independent and area dependent models used in this study implemented in GeoSSE and GeoHiSSE (Caetano et al. 2018). We tested a total of 24 models, the 12 models show below with and without the *j* parameter.

Model	**Description**	**Free parameters**
1	CID—original GeoSSE	4
2	Original GeoSSE, full model	7
3	CID—GeoHiSSE, three hidden rate classes, null model	9
4	GeoHiSSE, two rate classes, full model	15
5	CID—GeoHiSSE, five hidden rate classes, null model	13
6	CID—GeoHiSSE, two hidden rate classes	7
7	CID—GeoSSE + extirpation	6
8	GeoSSE + extirpation, full model	9
9	CID—GeoHiSSE + extirpation, three hidden rate classes, null model	11
10	GeoHiSSE + extirpation, two hidden rate classes, full model	19
11	CID—GeoHiSSE + extirpation, five hidden rate classes, null model	15
12	CID—GeoHiSSE + extirpation, two hidden rate classes	9

CID: Character-independent diversification; Null model: diversification and dispersion parameters are constrained to be equal among areas in the same hidden state category; Full model: all parameters of the model are free; Extirpation: separate rates of range reduction from the extinction of endemic lineages.

BAMM allows setting speciation and extinction priors and can also account for incomplete taxonomic sampling. For each tree, we estimated prior parameters for time-variable speciation and extinction models using the R package *BAMMTools* (Rabosky et al. 2014). We set the global sampling fraction to 0.5 (see above) and the expected number of regime shifts to 1. We ran four Markov chain Monte Carlo (MCMC) chains, each for 20 million generations with a sampling frequency of 1,000. After discarding the first 10% generations as burn-in, we assessed convergence of runs using the R package *coda* (Plummer et al. 2006). After running BAMM independently for each of our trees, we combined all results by calculating the mean diversification rate for each tip. Finally, we also estimated the tip rates using "DR statistics" function from (Jetz et al. 2012) for each tree.

For each analysis, we calculated the mean diversification rate across trees for each species and grouped all species according to their geographic distribution. We then statistically compared tip-associated lineage diversification rates between (i) all major realms, and (ii) all major subareas within the Indo-Pacific, both using trees that include all syngnatharian species as well within four separate suborder-level clades. Phylogenetic ANOVA is a method used to statistically compare groups through the residuals from phylogenetic regressions (Adams and Collyer 2018). However, after performing phylogenetic regressions between geographical areas and diversification rates, we found that the residuals did not follow a normal distribution (even after log transformation), a key assumption of linear regression models. As a result, we chose not to use ANOVA. Instead, to assess statistical significance of rate differences among groups, we conducted a phylogenetically-corrected, non-parametric Kruskal-Wallis test using custom code. Briefly, we placed the values of diversification rates into a vector and then multiplied this vector by the inverse of the square root of the covariance matrix (i.e., if this matrix is denoted as **P,** then $\mathbf{P} = \mathbf{Q}\mathbf{\Lambda}^{-1/2}\mathbf{Q}^{\top}$, where $\mathbf{Q}$ and $\mathbf{\Lambda}$ are the matrices of eigenvectors and eigenvalues of the covariance matrix, respectively; Garland and Ives 2000). By doing so, the diversification rates were weighted in this manner by the phylogeny, and we applied the Kruskal-Wallis test to these values. If the *p*-values were significant, we ran a pairwise comparison using the "kwAllPairsConoverTest" function with Bonferroni correction implemented in the R package *PMCMRplus* (Pohlert 2021).

Phylogenetically-corrected non-parametric Kruskal-Wallis test

For our comparative analyses (see main text), we implemented a phylogenetically-corrected, non-parametric Kruskal-Wallis test using custom code. This implementation follows Felsenstein (Felsenstein 1985) and Garland and Ives (Garland and Ives 2000) for deriving the phylogenetic linear model and provide mathematical justification for the use of the **P** matrix for data transformation. The **P** matrix is a function of the covariance matrix and typically includes phylogenetic information. By transforming the data with the **P** matrix, we were able to perform hypothesis testing using the non-parametric Kruskal-Wallis test. Below we explain how the **P** matrix naturally arises in the context of constructing a phylogenetic linear regression model.

Let the prediction of a linear model be affected by a random noise $\varepsilon \in \mathbb{R}^{n\times 1}$ that is not normally distributed:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon \,,$$

(1)

where $\mathbf{Y} \in \mathbb{R}^{n\times 1}$ is the response variable, $\mathbf{X} \in \mathbb{R}^{n\times p}$ is the matrix with instances, and $\beta \in \mathbb{R}^{p\times 1}$ is the coefficients vector.

Assume there exist an invertible matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$ such that its multiplication into both sides of equation 1 generates a new random noise $\mathbf{E} \in \mathbb{R}^{n\times 1}$ that follows a multivariable normal distribution with mean vector $\mathbf{0} \in \mathbb{R}^{n\times 1}$ and covariance matrix $\sigma^2\mathbf{I} \in \mathbb{R}^{n\times n}$:

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{P}\varepsilon = \boldsymbol{P}(\boldsymbol{Y} - \boldsymbol{X}\beta) \,, \\ \boldsymbol{E} &\sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I}) \,. \end{aligned}$$

We can re-write the above equation as $\mathbf{Y} = \mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E}$.

Finding distribution parameters of **Y**

We can then estimate the new covariance matrix and mean vector of $\mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E}$ from the distribution of $\mathbf{E}$. For the $-\mathbf{P}^{-1}\mathbf{E}$ term, the mean vector and covariance matrix is:

$$\begin{aligned} \mathrm{E}[-\mathbf{P}^{-1}\mathbf{E}] &= -\mathbf{P}^{-1}\mathrm{E}[\mathbf{E}] = \mathbf{0}\ , \\ \mathrm{Cov}[-\mathbf{P}^{-1}\mathbf{E}] &= \mathrm{E}[(-\mathbf{P}^{-1}\mathbf{E} - \mathrm{E}[-\mathbf{P}^{-1}\mathbf{E}])(-\mathbf{P}^{-1}\mathbf{E} - \mathrm{E}[-\mathbf{P}^{-1}\mathbf{E}])^{\top}] \\ &= \mathrm{E}[(-\mathbf{P}^{-1})(\mathbf{E} - \mathrm{E}[\mathbf{E}])(\mathbf{E} - \mathrm{E}[\mathbf{E}])^{\top}(-\mathbf{P}^{-1})^{\top}] \\ &= \mathbf{P}^{-1}\mathrm{Cov}[\mathbf{E}](\mathbf{P}^{-1})^{\top} \\ &= \mathbf{P}^{-1}\sigma^2\mathbf{I}(\mathbf{P}^{-1})^{\top} \\ &= \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top}\ , \end{aligned}$$

Without loss of generality, notice that we can define $\mathrm{Cov}[\varepsilon]$ as:

$$\begin{aligned} \mathrm{Cov}[\varepsilon] &= \mathrm{Cov}[\mathbf{P}^{-1}\mathbf{P}\varepsilon] \\ &= \mathrm{Cov}[\mathbf{P}^{-1}\mathbf{E}] \\ &= \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top}\ . \end{aligned}$$

(2)

Likewise, we can define $\mathrm{E}[\varepsilon]$ as:

$$\begin{aligned} \mathrm{E}[\varepsilon] &= \mathrm{E}[\mathbf{P}^{-1}\mathbf{P}\varepsilon] \\ &= \mathrm{E}[\mathbf{P}^{-1}\mathbf{E}] \\ &= \mathbf{P}^{-1}\mathrm{E}[\mathbf{E}] = \mathbf{0}\ . \end{aligned}$$

(3)

For the complete equation $\mathbf{Y} = \mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E}$, the mean vector and covariance matrix would be:

$$\begin{aligned} \mathrm{E}[\mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E}] &= \mathbf{X}\beta + \mathrm{E}[-\mathbf{P}^{-1}\mathbf{E}] \\ &= \mathbf{X}\beta\ , \\ \mathrm{Cov}[\mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E}] &= \mathrm{Cov}[-\mathbf{P}^{-1}\mathbf{E}] \\ &= \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top} \end{aligned}$$

And its distribution is given by:

$$\mathbf{Y} = \mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top})\ .$$

Let the covariance matrix of **X** be $\mathbf{\Omega} \in \mathbb{R}^{n\times n}$, where $\mathbf{\Omega}$ is a symmetric matrix and, thus, a positive semi-definite matrix. Let $\mathbf{P} = \mathbf{Q}\mathbf{\Lambda}^{-1/2}\mathbf{Q}^{\top}$, where $\mathbf{Q}$ and $\mathbf{\Lambda}$ are the eigenvectors and eigenvalues matrices of $\mathbf{\Omega}$, respectively, such that $\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top} = \mathbf{\Omega}$. Then, the covariance of $\mathbf{Y}$ can be expressed in function of $\mathbf{\Omega}$ as:

$$\mathbf{Y} = \mathbf{X}\beta - \mathbf{P}^{-1}\mathbf{E} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2\mathbf{\Omega}) \; .$$

(4)

Finding optimal β

Assume $\mathbf{Y}$ is independent and identically distributed (i.e., "i.i.d." assumption). Then, its density in equation 4 is given by:

$$\begin{aligned}
p(\mathbf{Y} \mid \mathbf{X}, \beta) \quad &= \frac{1}{(2\pi)^{n/2}} \frac{1}{|\sigma^2\mathbf{\Omega}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)^{\top}(\sigma^2\mathbf{\Omega})^{-1}(\mathbf{Y} - \mathbf{X}\beta)\right\} \\
\ln p(\mathbf{Y} \mid \mathbf{X}, \beta) \quad &= \ln\left\{\frac{1}{(2\pi)^{n/2}}\right\} + \ln\left\{\frac{1}{|\sigma^2\mathbf{\Omega}|^{1/2}}\right\} - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\beta)^{\top}\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\beta) \\
\ln p(\mathbf{Y} \mid \mathbf{X}, \beta) \quad &= c_1 + c_2 - c_3(\mathbf{Y} - \mathbf{X}\beta)^{\top}\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\beta)
\end{aligned}$$

Where c_1, c_2, and c_3 are constants from the model. From above equation we can clearly see that the maximization of $\ln p(\mathbf{Y} \mid \mathbf{X}, \beta)$ depends on the minimization of the right-most term:

$$\operatorname{argmax}_{\beta} \ln p(\mathbf{Y} \mid \mathbf{X}, \beta) \quad = \operatorname{argmin}_{\beta} (\mathbf{Y} - \mathbf{X}\beta)^{\top}\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\beta) \; .$$

From above relationship we can define our objective function as:

$$J(\beta) \quad = (\mathbf{Y} - \mathbf{X}\beta)^{\top}\mathbf{\Omega}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$$

(5)

Expanding the above equation, differentiating it with respect to β, and setting it to zero, we can obtain our optimal β:

$$\begin{aligned}
J(\beta) \quad &= \mathbf{Y}^{\top}\mathbf{\Omega}^{-1}\mathbf{Y} - 2\beta^{\top}\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{Y} + \beta^{\top}\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X}\beta \\
\frac{\partial J(\beta)}{\partial \beta} \quad &= -2\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{Y} + (\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X} + (\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X})^{\top})\beta = 0 \\
\beta \quad &= (\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}^{\top}\mathbf{\Omega}^{-1}\mathbf{Y})
\end{aligned}$$

Finding optimal $\boldsymbol{\beta}$ from transformed data

We can obtain the same optimal β by transforming data with the $\mathbf{P}$ matrix. Since we can re-write $\mathbf{\Omega}^{-1}$ as $\mathbf{P}^{\top}\mathbf{P} = \mathbf{\Omega}^{-1}$, then the objective function (i.e., equation 5) can also take this form:

$$\begin{aligned} J(\beta) &= (\mathbf{Y}-\mathbf{X}\beta)^{\top}\mathbf{P}^{\top}\mathbf{P}(\mathbf{Y}-\mathbf{X}\beta) \\ &= (\mathbf{PY}-\mathbf{PX}\beta)^{\top}(\mathbf{PY}-\mathbf{PX}\beta) \end{aligned}$$

Let $\mathbf{Y}^* = \mathbf{PY}$ and $\mathbf{X}^* = \mathbf{PX}$, such that we can re-write above equation as:

$$J(\beta) = (\mathbf{Y}^* - \mathbf{X}^*\beta)^{\top}(\mathbf{Y}^* - \mathbf{X}^*\beta)\ ,$$

whose solution generates same optimal β as the previous section. Furthermore, from equation 1 we can see the error can also get transformed by **P**:

$$\mathbf{E}^* = \mathbf{Y}^* - \mathbf{X}^*\beta = \mathbf{P}\varepsilon\ ,$$

And, from equation 2, the covariance of $\mathbf{E}^*$ is the identity matrix times a constant:

$$\begin{aligned} \mathrm{Cov}[\mathbf{E}^*] &= \mathrm{Cov}[\mathbf{P}\varepsilon] \\ &= \mathbf{P}\mathrm{Cov}[\varepsilon]\mathbf{P}^{\top} \\ &= \mathbf{P}\sigma^2\mathbf{P}^{-1}(\mathbf{P}^{-1})^{\top}\mathbf{P}^{\top} \\ &= \sigma^2\mathbf{I}\ . \end{aligned}$$

Likewise, from equation 3, its mean vector is $\mathbf{0}$:

$$\mathrm{E}[\mathbf{E}^*] = \mathrm{E}[\mathbf{P}\varepsilon] = \mathbf{P}\mathrm{E}[\varepsilon] = \mathbf{0}\ .$$

We theoretically show that transforming the data with the **P** matrix, which contains the phylogenetic information, can achieve a normal distribution of the residual errors in a phylogenetic linear regression. However, in our study, when we applied the transformation to our data (i.e., diversification and morphological rates) it did not result in a normal distribution of the residual errors. Consequently, we used the non-parametric Kruskal-Wallis test using the transformed data via **P** matrix as input.

Morphological disparity and rates by region

To assess whether phenotypic disparity varies across biogeographic areas, we examined the morphospace occupation and morphological evolutionary rates in Syngnatharia between the three major realms and within subareas of the Indo-Pacific using 2D geometric-morphometric analyses. We performed all morphological analyses considering all syngnatharian species and also within each suborder, as the results obtained with the entire clade may be obscured by the outstanding morphological disparity across the group. We placed a total of 12 landmarks and 2 semi-landmarks using photographs from 474 specimens, representing 171 species, obtained from museum collections (e.g., Smithsonian) and online repositories (e.g., Bray and Gomon 2021; Froese and Pauly 2021). In order to accommodate seahorse, pigmy pipehorse, and sea dragon specimens with bent body structures, we created two alternative schemes: a head-only set (including these specimens) and a full-body set (excluding them; Fig. 2.4 and Fig. S1). We examined 2 to 5 specimens per species, whenever possible, to account for intraspecific variation. To summarize variation in syngnatharian morphology within each morphometric scheme, we used the R package *geomorph* (Adams et al. 2021) to conduct both a Procrustes superimposition analysis, which corrects for size, scale, and rotation while calculating species-average coordinates, and subsequently a principal component analysis (PCA). We also calculated phylogenetically-corrected PCA (pPCA) scores based on all RelTime and MCMCTree phylogenies using *phytools* (Revell 2012). We selected PCs and pPCs explaining 95% of the variation in data (1–4 axes for head-only and 1–6 for full-body shape; see Extended Results) for downstream morphological analyses.

To examine contemporary trait disparity in syngnatharians, we used the R package *dispRity* (Guillerme 2018). Using "dispRity.per.group" function, we calculated the sum of variances for each biogeographical region based on both schemes using PC and pPC scores as input. We also analyzed disparity-through-time in syngnatharians using pPC scores obtained from our "master tree" and "alternative tree", the sum of variance metric, and the "dtt.dispRity" function in *dispRity.* We also evaluated the fit of different morphological evolutionary models using the R package *mvMORPH* (Clavel et al. 2015). We fitted a total of eight models using pPC and PC scores for each scheme across all trees (RelTime and MCMCTree): (i) a single rate Brownian motion (BM), ii) a single regime Orstein-Uhlenbeck (OU), (iii) an early burst (EB), (iv) multi regime BM, (v) multi regime OU, (vi) EB to independent rates OU shift, (vii) BM to independent rates OU shift, and (viii) EB to independent rates BM shift. For the models with multiple regimes and shifts, we applied a trait change at 78 Ma across the tree, as we observed by a steep decline of the morphological disparity at this time which also coincides with the origination of suborders (see Extended Results). To quantify the occupancy of syngnatharians in each of the major biogeographic regions, we calculated the overlap (Jaccard and Sørenson statistics) between each region using the R package *hypervolume* (Blonder et al. 2018). We used pPC scores obtained from the "master tree" and the "alternative tree" as input. We created four-dimensional hypervolumes for the head-only scheme and six-dimensional hypervolumes for the full-body scheme (see Extended Results). Finally, to assess morphological evolutionary rates within areas, we estimated rates for each syngnatharian lineage using BAMM (Rabosky 2014), with the caveat that each PC/pPC needed to be analyzed separately (Uyeda et al. 2015). We ran BAMM independently for the 28 RelTime trees and for the 28 MCMCTree trees. We then combined the MCMC results from all selected PCs and pPCs and calculated mean rates for each lineage across all 28 trees in each case. Lastly, we compared the statistical significance of morphological rates between realms and Indo-Pacific subareas using the modified Kruskal-Wallis test as explained for diversification rates (see above).

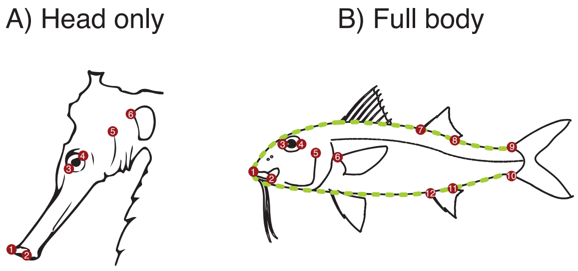


Figure S1. Geometric morphometrics digitization schemes to summarize the (A) head and (B) body shape variation in Syngnatharia: (1) rostral tip of premaxilla, (2) caudal end of maxilla, (3) anterior margin of eye, (4) posterior margin of eye, (5) dorsal end of opercule, (6) upper insertion of pectoral fin, (7) anterior insertion of second dorsal fin, (8) posterior insertion of second dorsal fin, (9) dorsal insertion of caudal fin, (10) ventral insertion of caudal fin, (11) posterior insertion of anal fin, and (12) anterior insertion of anal fin. Green points outline the sliding semi-landmark curves used for body shape.

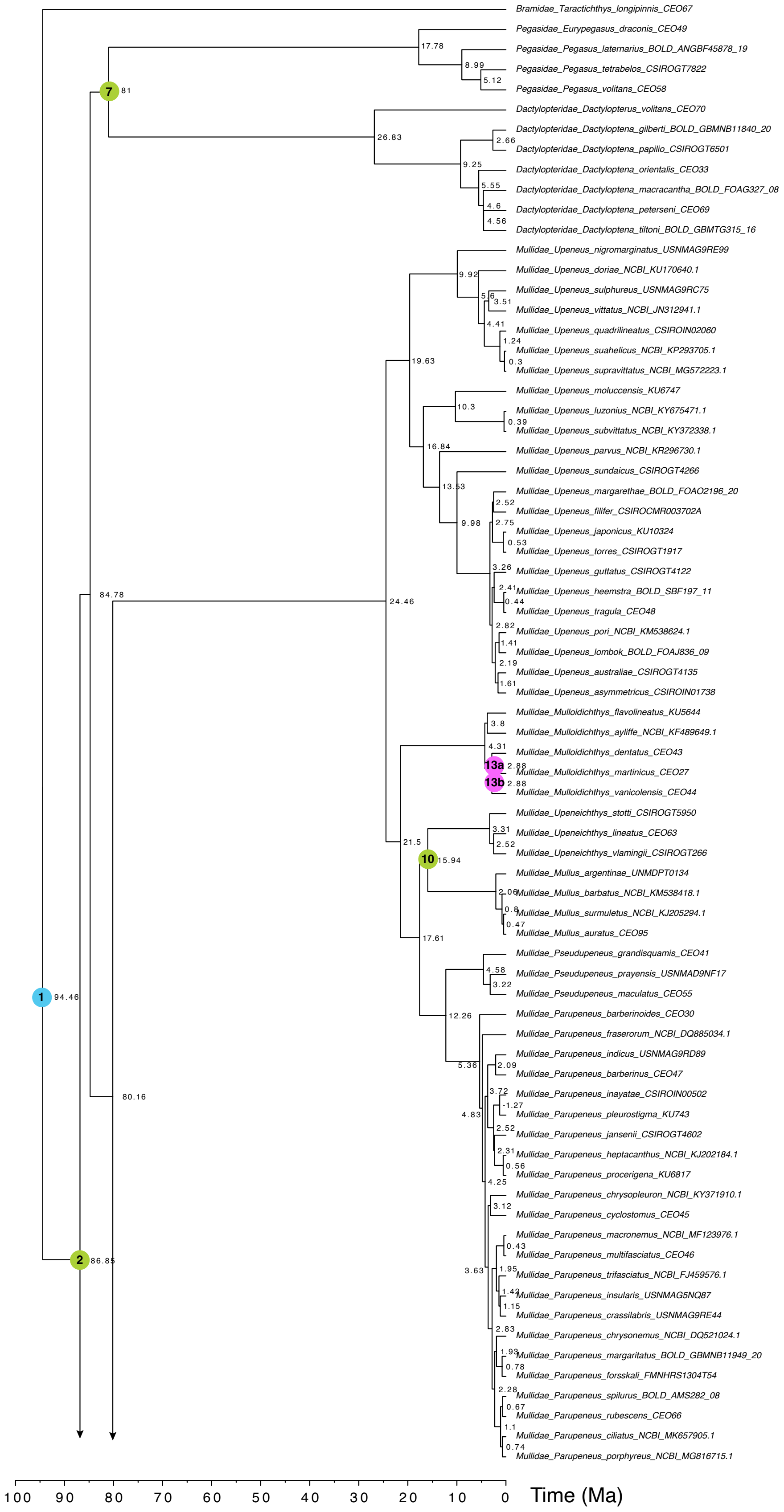
Supplementary Results

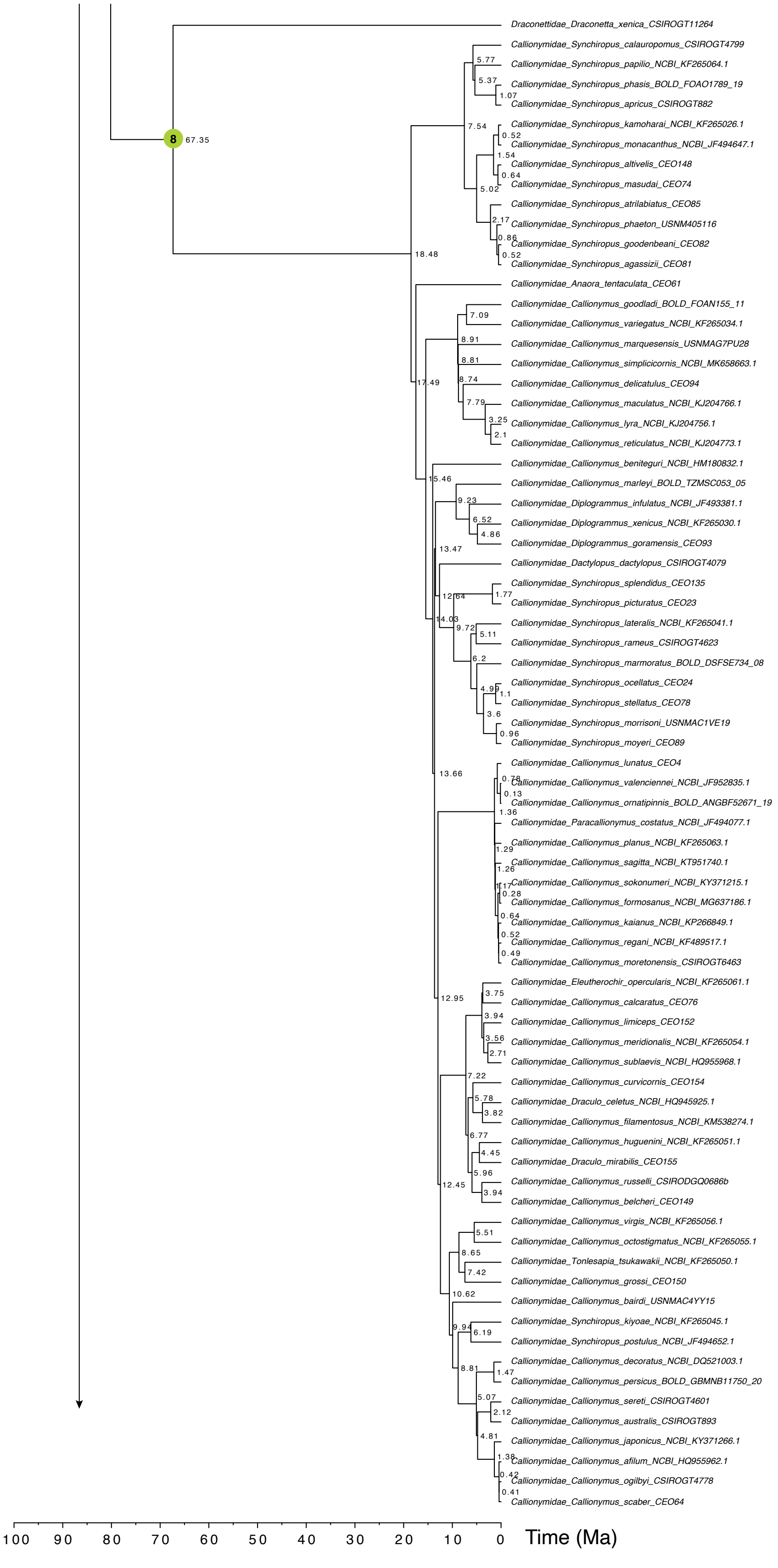
Expanded tree of Syngnatharia and tree uncertainty for downstream analyses

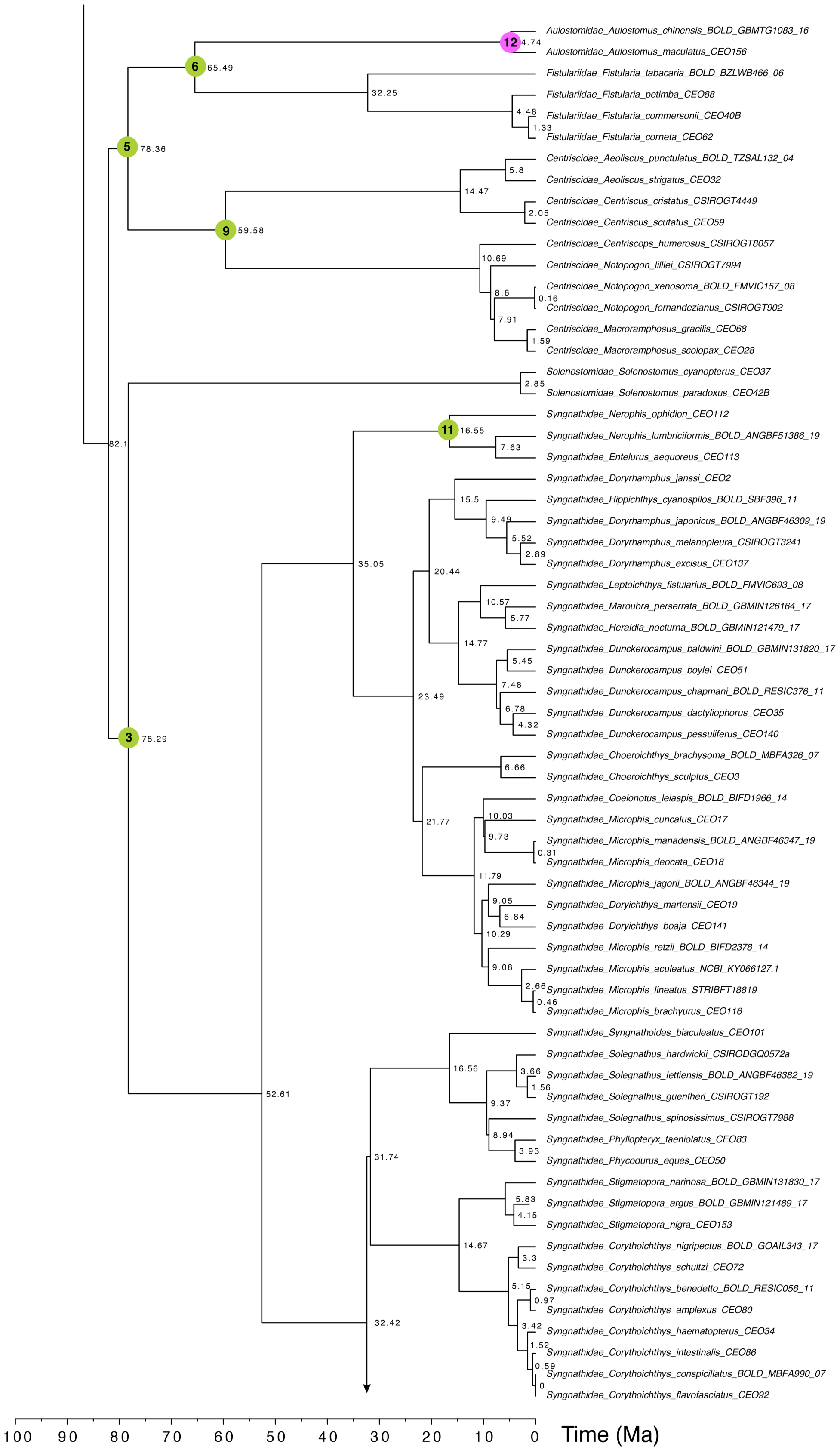
In agreement with previous studies (Longo et al. 2017; Santaquiteria et al. 2021; Stiller et al. 2022), our expanded trees estimated with 323 taxa provide congruent results with respect to the monophyly of each suborder and family. As previously shown (Rincon-Sandoval et al. 2020; Santaquiteria et al. 2021; Peterson et al. 2022), however, trees estimated by RAxML and ASTRAL-III cluster into two distinct groups in different regions of the tree space, with ASTRAL-III producing the largest disparity in tree space (Fig. S3). These results suggest that the majority of observed tree differences are due to the choice of phylogenetic inference method rather than due to gene set selection. With some exceptions noted below, divergence times for the suborders and some family-level clades of Syngnatharia are generally consistent between RelTime and MCMCTree, as well as with previous studies (Santaquiteria et al. 2021; Stiller et al. 2022; see Fig. S4).

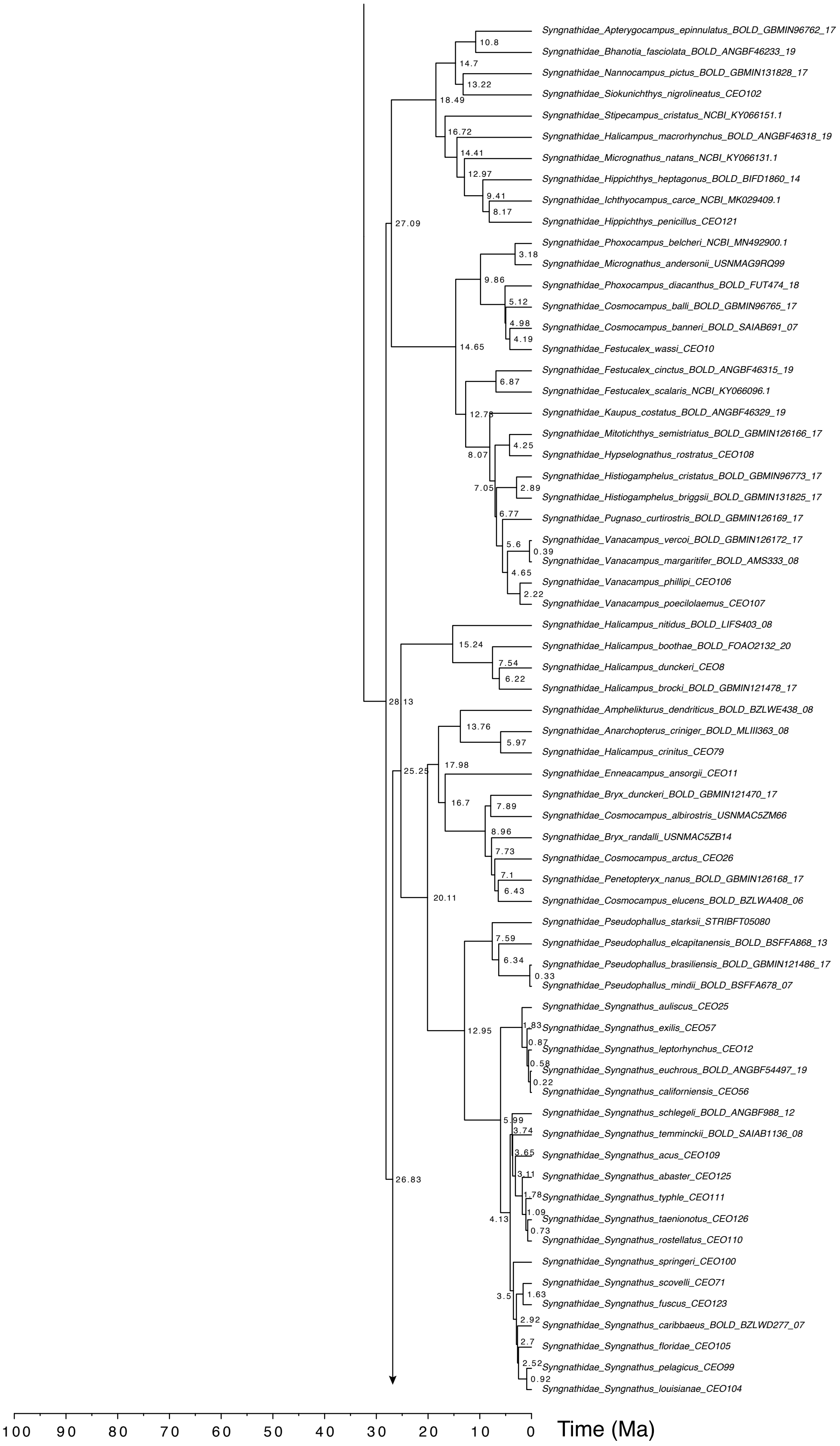
The origin of Syngnatharia was estimated to be 86.85 Ma in the “master tree”, with a range between 91.17–84.33 Ma across all 28 trees. Likewise, using the “alternative tree”, the estimated age was 86.01 Ma (88.47–84.44 Ma across all trees) (Figs. S3 and S4). For syngnatharian suborders, Dactylopteroidei originated at 81 Ma (84.82–74 Ma) using RelTime vs. 78.72 Ma (80.68–60.24 Ma) using MCMCTree; Syngnathoidei at 82.1 Ma (82.36–65.84 Ma) in RelTime vs. 82.34 Ma (83.9–77.47 Ma) in MCMCTree; Callionymoidei at 67.35 Ma (74.43–53.88 Ma) in RelTime vs. 68.47 Ma (76.11–52.08 Ma) in MCMCTree; and Mulloidei/Mullidae at 24.46 Ma (30–22 Ma) in RelTime vs. 20.85 Ma (27.72–18.96 Ma) in MCMCTree. At the family level, Centriscidae was dated at 59.58 Ma (67.52–54.8 Ma) in RelTime vs. 52.01 Ma (63.8–51.46 Ma) in MCMCTree; Fistulariidae at 32.25 Ma (35.96–27.34 Ma) in RelTime vs. 31.13 Ma (33.17–27.89 Ma) in MCMCTree; and Aulostomidae at 4.74 Ma (5.09–3.86 Ma) in RelTime vs. 4.32 Ma (4.48–3.91 Ma) in MCMCTree (Figs. S3 and S4). Other families have quite different ages depending on the method used. For example, due to the long stem of Dactylopteridae and Pegasidae, the ages are younger for Dactylopteridae in MCMCTree (15.49 Ma, 20.94–10.94 Ma vs. 26.83 Ma, 28.94–20.81 in RelTime) and for Pegasidae in RelTime (17.78 Ma, 20.31–15.19 Ma vs. 37.32 Ma, 41.04–31.95 Ma in MCMCTree). RelTime produced younger divergences for Syngnathidae 52.61 Ma (59.29–33.57 Ma) in RelTime vs. 63.27 (Ma; 66.02–55.2 Ma) in MCMCTree, for Solenostomidae 2.85 Ma (3.51–2.17 Ma) in RelTime vs. 9.84 Ma (11.34–6.88 Ma) in MCMCTree, and for Callionymidae 18.48 Ma (64.59–15.78 Ma) in RelTime vs. 46.21 Ma (71.35–36.45 Ma) in MCMCTree. Refer to Fig. S4 for a visual comparison.

We believe that for Syngnatharia, RelTime appears to outperform MCMCTree, as this clade is non-clock-like (Figs. S2 and S5). The branching patterns obtained with RelTime are closer to the inferred phylograms (see Fig. S5), showing short internodes near the base of the clade that highlight the rapid radiation nature among early syngnatharian lineages (Betancur-R. et al. 2017; Alfaro et al. 2018; Santaquiteria et al. 2021). Additionally, RelTime trees tend to capture more recent diversification bursts observed in the RAxML phylogram, whereas speciation events for those clades tend to be more spread in MCMCTree trees, presumably due to the implementation of the constant birth-death tree prior (see Fig. S5 for an example of one of the subsets). Therefore, we ran all our comparative analyses twice, using the trees inferred in MCMCTree and RelTime, but reported the RelTime results in the main text and the MCMCTree results here in Appendix B.


Bramidae_Taractichthys_longipinnis_CEO67
Pegasidae_Eurypegasus_draconis_CEO49
17.78
Pegasidae_Pegasus_laternarius_BOLD_ANGBF45878_19
8.99
Pegasidae_Pegasus_tetrabelos_CSIROGT7822
5.12
Pegasidae_Pegasus_volitans_CEO58
7
81
Dactylopteridae_Dactylopterus_volitans_CEO70
26.83
Dactylopteridae_Dactyloptena_gilberti_BOLD_GBMNB11840_20
2.66
Dactylopteridae_Dactyloptena_papilio_CSIROGT6501
9.25
Dactylopteridae_Dactyloptena_orientalis_CEO33
5.55
Dactylopteridae_Dactyloptena_macracantha_BOLD_FOAG327_08
4.6
Dactylopteridae_Dactyloptena_peterseni_CEO69
4.56
Dactylopteridae_Dactyloptena_tiltoni_BOLD_GBMTG315_16
Mullidae_Upeneus_nigromarginatus_USNMAG9RE99
9.92
Mullidae_Upeneus_doriae_NCBI_KU170640.1
5.6
Mullidae_Upeneus_sulphureus_USNMAG9RC75
3.51
Mullidae_Upeneus_vittatus_NCBI_JN312941.1
4.41
Mullidae_Upeneus_quadrilineatus_CSIROIN02060
1.24
Mullidae_Upeneus_suahelicus_NCBI_KP293705.1
0.3
Mullidae_Upeneus_supravittatus_NCBI_MG572223.1
19.63
Mullidae_Upeneus_moluccensis_KU6747
10.3
Mullidae_Upeneus_luzonius_NCBI_KY675471.1
0.39
Mullidae_Upeneus_subvittatus_NCBI_KY372338.1
16.84
Mullidae_Upeneus_parvus_NCBI_KR296730.1
Mullidae_Upeneus_sundaicus_CSIROGT4266
13.53
Mullidae_Upeneus_margarethae_BOLD_FOAO2196_20
2.52
Mullidae_Upeneus_filifer_CSIROCMR003702A
9.98
2.75
Mullidae_Upeneus_japonicus_KU10324
0.53
Mullidae_Upeneus_torres_CSIROGT1917
3.26
Mullidae_Upeneus_guttatus_CSIROGT4122
84.78
2.41
Mullidae_Upeneus_heemstra_BOLD_SBF197_11
0.44
Mullidae_Upeneus_tragula_CEO48
24.46
2.82
Mullidae_Upeneus_pori_NCBI_KM538624.1
1.41
Mullidae_Upeneus_lombok_BOLD_FOAJ836_09
2.19
Mullidae_Upeneus_australiae_CSIROGT4135
1.61
Mullidae_Upeneus_asymmetricus_CSIROIN01738
Mullidae_Mulloidichthys_flavolineatus_KU5644
3.8
Mullidae_Mulloidichthys_ayliffe_NCBI_KF489649.1
4.31
Mullidae_Mulloidichthys_dentatus_CEO43
13a
2.88
Mullidae_Mulloidichthys_martinicus_CEO27
13b
2.88
Mullidae_Mulloidichthys_vanicolensis_CEO44
Mullidae_Upeneichthys_stotti_CSIROGT5950
3.31
Mullidae_Upeneichthys_lineatus_CEO63
21.5
2.52
Mullidae_Upeneichthys_vlamingii_CSIROGT266
10
15.94
Mullidae_Mullus_argentinae_UNMDPT0134
2.06
Mullidae_Mullus_barbatus_NCBI_KM538418.1
0.8
Mullidae_Mullus_surmuletus_NCBI_KJ205294.1
0.47
Mullidae_Mullus_auratus_CEO95
17.61
Mullidae_Pseudupeneus_grandisquamis_CEO41
4.58
Mullidae_Pseudupeneus_prayensis_USNMAD9NF17
3.22
1
94.46
Mullidae_Pseudupeneus_maculatus_CEO55
Mullidae_Parupeneus_barberinoides_CEO30
12.26
Mullidae_Parupeneus_fraserorum_NCBI_DQ885034.1
Mullidae_Parupeneus_indicus_USNMAG9RD89
5.36
2.09
Mullidae_Parupeneus_barberinus_CEO47
3.72
Mullidae_Parupeneus_inayatae_CSIROIN00502
80.16
-1.27
4.83
Mullidae_Parupeneus_pleurostigma_KU743
2.52
Mullidae_Parupeneus_jansenii_CSIROGT4602
2.31
Mullidae_Parupeneus_heptacanthus_NCBI_KJ202184.1
0.56
Mullidae_Parupeneus_procerigena_KU6817
4.25
Mullidae_Parupeneus_chrysopleuron_NCBI_KY371910.1
3.12
Mullidae_Parupeneus_cyclostomus_CEO45
Mullidae_Parupeneus_macronemus_NCBI_MF123976.1
0.43
2
86.85
Mullidae_Parupeneus_multifasciatus_CEO46
3.63
1.95
Mullidae_Parupeneus_trifasciatus_NCBI_FJ459576.1
1.42
Mullidae_Parupeneus_insularis_USNMAG5NQ87
1.15
Mullidae_Parupeneus_crassilabris_USNMAG9RE44
2.83
Mullidae_Parupeneus_chrysonemus_NCBI_DQ521024.1
1.93
Mullidae_Parupeneus_margaritatus_BOLD_GBMNB11949_20
0.78
Mullidae_Parupeneus_forsskali_FMNHRS1304T54
2.28
Mullidae_Parupeneus_spilurus_BOLD_AMS282_08
0.67
Mullidae_Parupeneus_rubescens_CEO66
1.1
Mullidae_Parupeneus_ciliatus_NCBI_MK657905.1
0.74
Mullidae_Parupeneus_porphyreus_NCBI_MG816715.1
100
90
80
70
60
50
40
30
20
10
0
Time (Ma)


8
67.35
Draconettidae_Draconetta_xenica_CSIROGT11264
Callionymidae_Synchiropus_calauropomus_CSIROGT4799
5.77
Callionymidae_Synchiropus_papilio_NCBI_KF265064.1
5.37
Callionymidae_Synchiropus_phasis_BOLD_FOAO1789_19
1.07
Callionymidae_Synchiropus_apricus_CSIROGT882
7.54
Callionymidae_Synchiropus_kamoharai_NCBI_KF265026.1
0.52
Callionymidae_Synchiropus_monacanthus_NCBI_JF494647.1
1.54
Callionymidae_Synchiropus_altivelis_CEO148
0.64
Callionymidae_Synchiropus_masudai_CEO74
5.02
Callionymidae_Synchiropus_atrilabiatus_CEO85
2.17
Callionymidae_Synchiropus_phaeton_USNM405116
0.86
Callionymidae_Synchiropus_goodenbeani_CEO82
18.48
0.52
Callionymidae_Synchiropus_agassizii_CEO81
Callionymidae_Anaora_tentaculata_CEO61
Callionymidae_Callionymus_goodladi_BOLD_FOAN155_11
7.09
Callionymidae_Callionymus_variegatus_NCBI_KF265034.1
8.91
Callionymidae_Callionymus_marquesensis_USNMAG7PU28
8.81
Callionymidae_Callionymus_simplicicornis_NCBI_MK658663.1
17.49
8.74
Callionymidae_Callionymus_delicatulus_CEO94
7.79
Callionymidae_Callionymus_maculatus_NCBI_KJ204766.1
3.25
Callionymidae_Callionymus_lyra_NCBI_KJ204756.1
2.1
Callionymidae_Callionymus_reticulatus_NCBI_KJ204773.1
Callionymidae_Callionymus_beniteguri_NCBI_HM180832.1
15.46
Callionymidae_Callionymus_marleyi_BOLD_TZMSC053_05
9.23
Callionymidae_Diplogrammus_infulatus_NCBI_JF493381.1
6.52
Callionymidae_Diplogrammus_xenicus_NCBI_KF265030.1
4.86
Callionymidae_Diplogrammus_goramensis_CEO93
13.47
Callionymidae_Dactylopus_dactylopus_CSIROGT4079
Callionymidae_Synchiropus_splendidus_CEO135
12.64
1.77
Callionymidae_Synchiropus_picturatus_CEO23
14.03
9.72
Callionymidae_Synchiropus_lateralis_NCBI_KF265041.1
5.11
Callionymidae_Synchiropus_rameus_CSIROGT4623
6.2
Callionymidae_Synchiropus_marmoratus_BOLD_DSFSE734_08
4.99
Callionymidae_Synchiropus_ocellatus_CEO24
1.1
Callionymidae_Synchiropus_stellatus_CEO78
3.6
Callionymidae_Synchiropus_morrisoni_USNMAC1VE19
0.96
Callionymidae_Synchiropus_moyeri_CEO89
Callionymidae_Callionymus_lunatus_CEO4
13.66
0.78
Callionymidae_Callionymus_valenciennei_NCBI_JF952835.1
0.13
Callionymidae_Callionymus_ornatipinnis_BOLD_ANGBF52671_19
1.36
Callionymidae_Paracallionymus_costatus_NCBI_JF494077.1
Callionymidae_Callionymus_planus_NCBI_KF265063.1
1.29
Callionymidae_Callionymus_sagitta_NCBI_KT951740.1
1.26
1.17
Callionymidae_Callionymus_sokonumeri_NCBI_KY371215.1
0.28
Callionymidae_Callionymus_formosanus_NCBI_MG637186.1
0.64
Callionymidae_Callionymus_kaianus_NCBI_KP266849.1
0.52
Callionymidae_Callionymus_regani_NCBI_KF489517.1
0.49
Callionymidae_Callionymus_moretonensis_CSIROGT6463
Callionymidae_Eleutherochir_opercularis_NCBI_KF265061.1
3.75
12.95
Callionymidae_Callionymus_calcaratus_CEO76
3.94
Callionymidae_Callionymus_limiceps_CEO152
3.56
Callionymidae_Callionymus_meridionalis_NCBI_KF265054.1
2.71
Callionymidae_Callionymus_sublaevis_NCBI_HQ955968.1
7.22
Callionymidae_Callionymus_curvicornis_CEO154
5.78
Callionymidae_Draculo_celetus_NCBI_HQ945925.1
3.82
Callionymidae_Callionymus_filamentosus_NCBI_KM538274.1
6.77
Callionymidae_Callionymus_huguenini_NCBI_KF265051.1
4.45
Callionymidae_Draculo_mirabilis_CEO155
5.96
Callionymidae_Callionymus_russelli_CSIRODGQ0686b
12.45
3.94
Callionymidae_Callionymus_belcheri_CEO149
Callionymidae_Callionymus_virgis_NCBI_KF265056.1
5.51
Callionymidae_Callionymus_octostigmatus_NCBI_KF265055.1
8.65
Callionymidae_Tonlesapia_tsukawakii_NCBI_KF265050.1
7.42
Callionymidae_Callionymus_grossi_CEO150
10.62
Callionymidae_Callionymus_bairdi_USNMAC4YY15
Callionymidae_Synchiropus_kiyoae_NCBI_KF265045.1
9.94
6.19
Callionymidae_Synchiropus_postulus_NCBI_JF494652.1
Callionymidae_Callionymus_decoratus_NCBI_DQ521003.1
8.81
1.47
Callionymidae_Callionymus_persicus_BOLD_GBMNB11750_20
5.07
Callionymidae_Callionymus_sereti_CSIROGT4601
2.12
Callionymidae_Callionymus_australis_CSIROGT893
4.81
Callionymidae_Callionymus_japonicus_NCBI_KY371266.1
1.38
Callionymidae_Callionymus_afilum_NCBI_HQ955962.1
0.42
Callionymidae_Callionymus_ogilbyi_CSIROGT4778
0.41
Callionymidae_Callionymus_scaber_CEO64
100
90
80
70
60
50
40
30
20
10
0
Time (Ma)


Aulostomidae_Aulostomus_chinensis_BOLD_GBMTG1083_16
12
4.74
Aulostomidae_Aulostomus_maculatus_CEO156
6
65.49
Fistulariidae_Fistularia_tabacaria_BOLD_BZLWB466_06
32.25
Fistulariidae_Fistularia_petimba_CEO88
4.48
Fistulariidae_Fistularia_commersonii_CEO40B
1.33
Fistulariidae_Fistularia_corneta_CEO62
5
78.36
Centriscidae_Aeoliscus_punctulatus_BOLD_TZSAL132_04
5.8
Centriscidae_Aeoliscus_strigatus_CEO32
14.47
Centriscidae_Centriscus_cristatus_CSIROGT4449
2.05
Centriscidae_Centriscus_scutatus_CEO59
9
59.58
Centriscidae_Centriscops_humerosus_CSIROGT8057
10.69
Centriscidae_Notopogon_lilliei_CSIROGT7994
Centriscidae_Notopogon_xenosoma_BOLD_FMVIC157_08
8.6
0.16
Centriscidae_Notopogon_fernandezianus_CSIROGT902
7.91
Centriscidae_Macroramphosus_gracilis_CEO68
1.59
Centriscidae_Macroramphosus_scolopax_CEO28
Solenostomidae_Solenostomus_cyanopterus_CEO37
2.85
Solenostomidae_Solenostomus_paradoxus_CEO42B
Syngnathidae_Nerophis_ophidion_CEO112
11
16.55
Syngnathidae_Nerophis_lumbriciformis_BOLD_ANGBF51386_19
82.1
7.63
Syngnathidae_Entelurus_aequoreus_CEO113
Syngnathidae_Doryrhamphus_janssi_CEO2
15.5
Syngnathidae_Hippichthys_cyanospilos_BOLD_SBF396_11
9.49
Syngnathidae_Doryrhamphus_japonicus_BOLD_ANGBF46309_19
5.52
Syngnathidae_Doryrhamphus_melanopleura_CSIROGT3241
2.89
35.05
Syngnathidae_Doryrhamphus_excisus_CEO137
20.44
Syngnathidae_Leptoichthys_fistularius_BOLD_FMVIC693_08
10.57
Syngnathidae_Maroubra_perserrata_BOLD_GBMIN126164_17
5.77
Syngnathidae_Heraldia_nocturna_BOLD_GBMIN121479_17
14.77
Syngnathidae_Dunckerocampus_baldwini_BOLD_GBMIN131820_17
5.45
Syngnathidae_Dunckerocampus_boylei_CEO51
7.48
Syngnathidae_Dunckerocampus_chapmani_BOLD_RESIC376_11
23.49
6.78
Syngnathidae_Dunckerocampus_dactyliophorus_CEO35
4.32
Syngnathidae_Dunckerocampus_pessuliferus_CEO140
3
78.29
Syngnathidae_Choeroichthys_brachysoma_BOLD_MBFA326_07
6.66
Syngnathidae_Choeroichthys_sculptus_CEO3
Syngnathidae_Coelonotus_leiaspis_BOLD_BIFD1966_14
10.03
Syngnathidae_Microphis_cuncalus_CEO17
21.77
9.73
Syngnathidae_Microphis_manadensis_BOLD_ANGBF46347_19
0.31
Syngnathidae_Microphis_deocata_CEO18
11.79
Syngnathidae_Microphis_jagorii_BOLD_ANGBF46344_19
9.05
Syngnathidae_Doryichthys_martensii_CEO19
6.84
Syngnathidae_Doryichthys_boaja_CEO141
10.29
Syngnathidae_Microphis_retzii_BOLD_BIFD2378_14
9.08
Syngnathidae_Microphis_aculeatus_NCBI_KY066127.1
2.66
Syngnathidae_Microphis_lineatus_STRIBFT18819
0.46
Syngnathidae_Microphis_brachyurus_CEO116
Syngnathidae_Syngnathoides_biaculeatus_CEO101
Syngnathidae_Solegnathus_hardwickii_CSIRODGQ0572a
16.56
3.66
Syngnathidae_Solegnathus_lettiensis_BOLD_ANGBF46382_19
1.56
52.61
Syngnathidae_Solegnathus_guentheri_CSIROGT192
9.37
Syngnathidae_Solegnathus_spinosissimus_CSIROGT7988
8.94
Syngnathidae_Phyllopteryx_taeniolatus_CEO83
3.93
31.74
Syngnathidae_Phycodurus_eques_CEO50
Syngnathidae_Stigmatopora_narinosa_BOLD_GBMIN131830_17
5.83
Syngnathidae_Stigmatopora_argus_BOLD_GBMIN121489_17
4.15
Syngnathidae_Stigmatopora_nigra_CEO153
14.67
Syngnathidae_Corythoichthys_nigripectus_BOLD_GOAIL343_17
3.3
Syngnathidae_Corythoichthys_schultzi_CEO72
5.15
Syngnathidae_Corythoichthys_benedetto_BOLD_RESIC058_11
0.97
Syngnathidae_Corythoichthys_amplexus_CEO80
32.42
3.42
Syngnathidae_Corythoichthys_haematopterus_CEO34
1.52
Syngnathidae_Corythoichthys_intestinalis_CEO86
0.59
Syngnathidae_Corythoichthys_conspicillatus_BOLD_MBFA990_07
0
Syngnathidae_Corythoichthys_flavofasciatus_CEO92
100
90
80
70
60
50
40
30
20
10
0
Time (Ma)


Syngnathidae_Apterygocampus_epinnulatus_BOLD_GBMIN96762_17
10.8
Syngnathidae_Bhanotia_fasciolata_BOLD_ANGBF46233_19
14.7
Syngnathidae_Nannocampus_pictus_BOLD_GBMIN131828_17
13.22
Syngnathidae_Siokunichthys_nigrolineatus_CEO102
18.49
Syngnathidae_Stipecampus_cristatus_NCBI_KY066151.1
16.72
Syngnathidae_Halicampus_macrorhynchus_BOLD_ANGBF46318_19
14.41
Syngnathidae_Micrognathus_natans_NCBI_KY066131.1
12.97
Syngnathidae_Hippichthys_heptagonus_BOLD_BIFD1860_14
9.41
Syngnathidae_Ichthyocampus_carce_NCBI_MK029409.1
8.17
27.09
Syngnathidae_Hippichthys_penicillus_CEO121
Syngnathidae_Phoxocampus_belcheri_NCBI_MN492900.1
3.18
Syngnathidae_Micrognathus_andersonii_USNMAG9RQ99
9.86
Syngnathidae_Phoxocampus_diacanthus_BOLD_FUT474_18
5.12
Syngnathidae_Cosmocampus_balli_BOLD_GBMIN96765_17
4.98
Syngnathidae_Cosmocampus_banneri_BOLD_SAIAB691_07
4.19
14.65
Syngnathidae_Festucalex_wassi_CEO10
Syngnathidae_Festucalex_cinctus_BOLD_ANGBF46315_19
6.87
Syngnathidae_Festucalex_scalaris_NCBI_KY066096.1
12.78
Syngnathidae_Kaupus_costatus_BOLD_ANGBF46329_19
Syngnathidae_Mitotichthys_semistriatus_BOLD_GBMIN126166_17
4.25
8.07
Syngnathidae_Hypselognathus_rostratus_CEO108
Syngnathidae_Histiogamphelus_cristatus_BOLD_GBMIN96773_17
7.05
2.89
Syngnathidae_Histiogamphelus_briggsii_BOLD_GBMIN131825_17
6.77
Syngnathidae_Pugnaso_curtirostris_BOLD_GBMIN126169_17
5.6
Syngnathidae_Vanacampus_vercoi_BOLD_GBMIN126172_17
0.39
Syngnathidae_Vanacampus_margaritifer_BOLD_AMS333_08
4.65
Syngnathidae_Vanacampus_phillipi_CEO106
2.22
Syngnathidae_Vanacampus_poecilolaemus_CEO107
Syngnathidae_Halicampus_nitidus_BOLD_LIFS403_08
15.24
Syngnathidae_Halicampus_boothae_BOLD_FOAO2132_20
7.54
Syngnathidae_Halicampus_dunckeri_CEO8
6.22
Syngnathidae_Halicampus_brocki_BOLD_GBMIN121478_17
28.13
Syngnathidae_Amphelikturus_dendriticus_BOLD_BZLWE438_08
13.76
Syngnathidae_Anarchopterus_criniger_BOLD_MLIII363_08
5.97
Syngnathidae_Halicampus_crinitus_CEO79
17.98
25.25
Syngnathidae_Enneacampus_ansorgii_CEO11
Syngnathidae_Bryx_dunckeri_BOLD_GBMIN121470_17
16.7
7.89
Syngnathidae_Cosmocampus_albirostris_USNMAC5ZM66
8.96
Syngnathidae_Bryx_randalli_USNMAC5ZB14
7.73
Syngnathidae_Cosmocampus_arctus_CEO26
7.1
Syngnathidae_Penetopteryx_nanus_BOLD_GBMIN126168_17
6.43
20.11
Syngnathidae_Cosmocampus_elucens_BOLD_BZLWA408_06
Syngnathidae_Pseudophallus_starksii_STRIBFT05080
7.59
Syngnathidae_Pseudophallus_elcapitanensis_BOLD_BSFFA868_13
6.34
Syngnathidae_Pseudophallus_brasiliensis_BOLD_GBMIN121486_17
0.33
Syngnathidae_Pseudophallus_mindii_BOLD_BSFFA678_07
Syngnathidae_Syngnathus_auliscus_CEO25
1.83
Syngnathidae_Syngnathus_exilis_CEO57
12.95
0.87
Syngnathidae_Syngnathus_leptorhynchus_CEO12
0.58
Syngnathidae_Syngnathus_euchrous_BOLD_ANGBF54497_19
0.22
Syngnathidae_Syngnathus_californiensis_CEO56
Syngnathidae_Syngnathus_schlegeli_BOLD_ANGBF988_12
5.99
3.74
Syngnathidae_Syngnathus_temminckii_BOLD_SAIAB1136_08
3.65
Syngnathidae_Syngnathus_acus_CEO109
26.83
3.11
Syngnathidae_Syngnathus_abaster_CEO125
1.78
Syngnathidae_Syngnathus_typhle_CEO111
1.09
4.13
Syngnathidae_Syngnathus_taenionotus_CEO126
0.73
Syngnathidae_Syngnathus_rostellatus_CEO110
Syngnathidae_Syngnathus_springeri_CEO100
Syngnathidae_Syngnathus_scovelli_CEO71
1.63
3.5
Syngnathidae_Syngnathus_fuscus_CEO123
2.92
Syngnathidae_Syngnathus_caribbaeus_BOLD_BZLWD277_07
2.7
Syngnathidae_Syngnathus_floridae_CEO105
2.52
Syngnathidae_Syngnathus_pelagicus_CEO99
0.92
Syngnathidae_Syngnathus_louisianae_CEO104
100
90
80
70
60
50
40
30
20
10
0
Time (Ma)

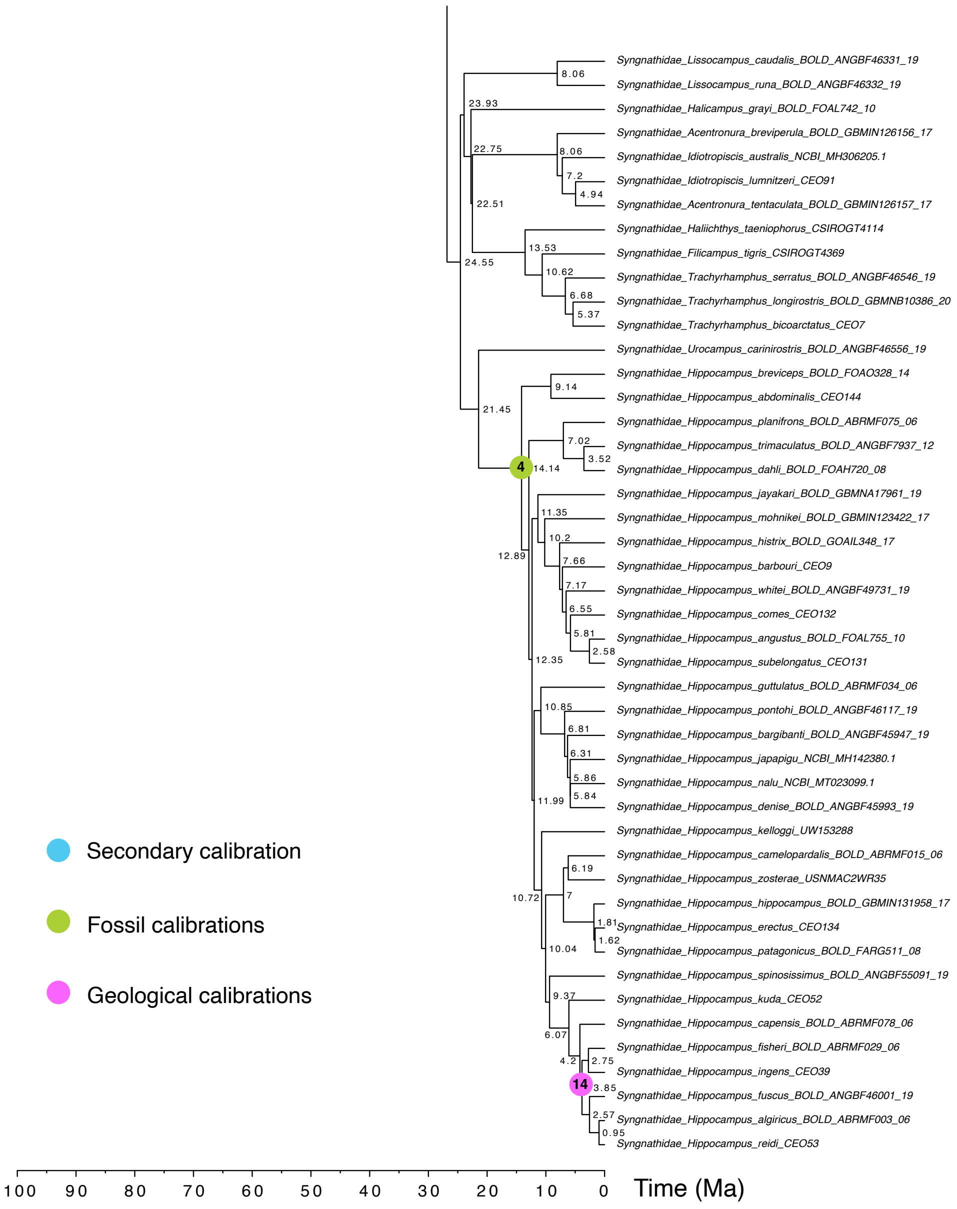


Figure S2. Time-calibrated phylogeny for 323 species of Syngnatharia and one outgroup (*Taractichthys longipinnis*). The topology reflects the "master tree", a maximum-likelihood RAxML tree topology based on 932 UCEs (75% completeness matrix), inferred in RelTime and used as a backbone constraint to place species with only COI sequences, later calibrated via *congruification* analysis in TreePL. Circles at nodes indicate the 14 calibration points used for RelTime and MCMCTree analyses, which are colored according to the type of calibration used (see Extended Material and Methods, Table S2).

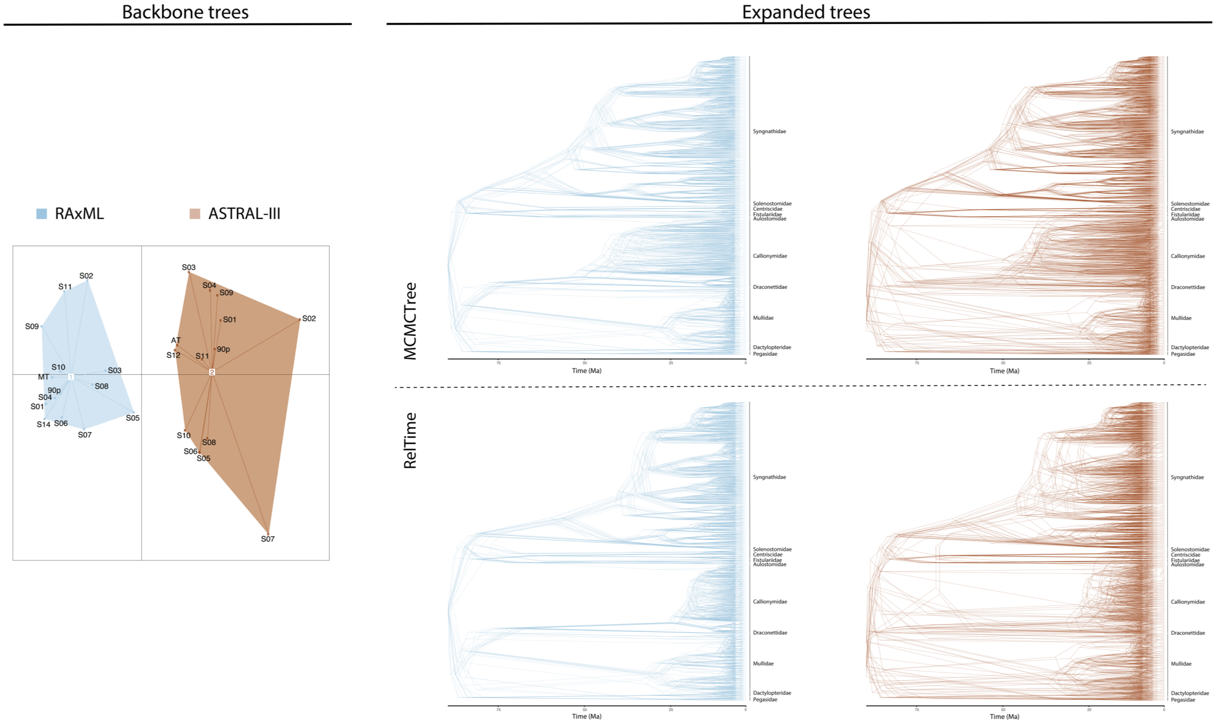


Figure S3. Uncertainty in phylogenetic relationships and diverge times. **Tree space plot (left)** for the 28 backbone trees (163 species) estimated with RAxML and ASTRAL-III. The trees are primarily grouped based on phylogenetic method used (RAxML vs. ASTRAL-III) followed by genomic subset. The average (centroid) tree in tree space is represented as "1" for RAxML and "2" for ASTRAL-III. MT: master tree; AT: ASTRAL-III tree based on the 75% completeness matrix; 90p: trees based on 90% completeness matrix; S: trees generated with gene subsets. **Density trees (right)** for the expanded trees (323 species) inferred with RAxML and ASTRAL-III, and time-calibrated with RelTime and MCMCTree.

Figure S4. Divergence times of syngnatharian suborders and families obtained from the 28 trees calibrated with RelTime and MCMCTree and compared to two previous studies: Santaquiteria *et al.* (2021) which also used MCMCTree, and Stiller *et al.* (2022) which used BEAST. The inclusion of misidentified *F. tabacaria* (actual identity *F. petimba*) in Stiller *et al.* (2022), and the lack of *F. tabacaria* in Santaquiteria *et al.* (2021), resulted in significant differences in the estimated ages of the family Fistulariidae compared to those inferred here (indicated with a red arrow). Additionally, due to heterotachy within the suborder Dactylopteroidei, we also observe differences between RelTime and Bayesian (MCMCTree and BEAST) methods (indicated with black arrows). It should be noted that the family Draconettidae is not illustrated here, as only one species was included in this study as well as in Santaquiteria *et al.* (2021).

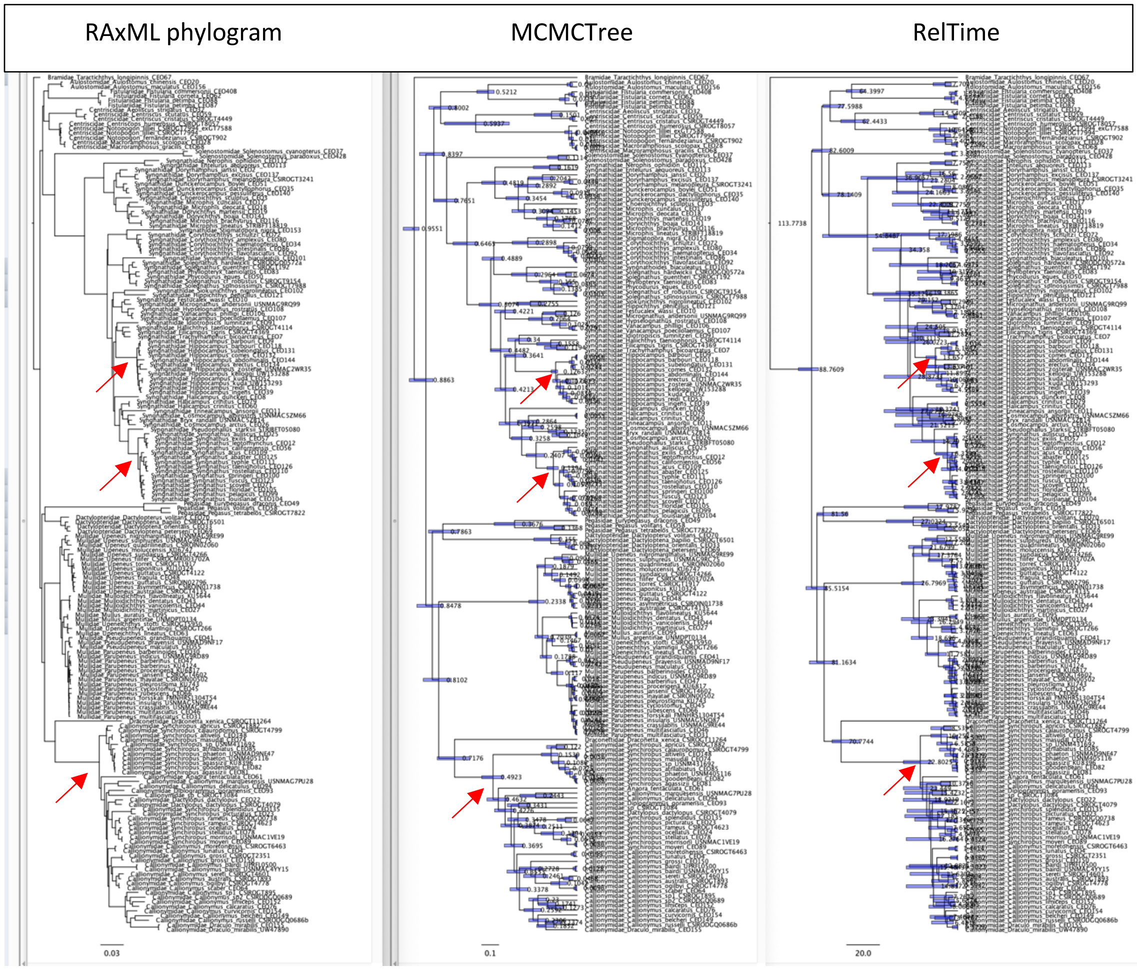


Figure S5. Phylogenetic trees obtained based on subset 1 illustrating the branching patterns obtained with the RAxML phylogram (left) and the chronograms for that topology as estimated with MCMCTree (middle) and RelTime (right). The phylogram shows that Syngnatharia poses a challenge for time-calibration analyses due to significant departures from a strict clock model, (e.g., compare Pegasidae vs. Dactylopteridae), and also because it features many rapid evolutionary radiations (e.g., Callionymidae, *Hippocampus*, *Syngnathus*; indicated with red arrows). The RelTime tree tends to capture the diversification bursts seen in the RAxML phylogram, whereas speciation events tend to be more spread in the MCMCTree tree, presumably due to implementation of the constant birth-death tree prior. However, MCMCTree does outperform RelTime in error bar estimation (errors bars are much longer bar in the RelTime tree).

Table S5. Summary statistics of the 12 biogeographic models implemented in BioGeoBEARS based on the “master tree” (RelTime) and “alternative tree” (MCMCTree). The best-fitting model for each scheme is indicated in bold.

Models	**LnL**	**Number of parameters**	**d**	**e**	**j**	**w**	**AICc**	**AICc weight**
"master tree"								
DEC	-1056	2	0.061	0.03	0	1	2115	6.90E-17
DEC+j	-1048	3	0.054	0.024	0.031	1	2103	3.80E-14
DEC+w	-1039	3	0.041	0.022	0	0.32	2084	5.60E-10
DEC+j+w	-1190	4	0.018	0.038	0.0018	0.028	2388	4.90E-76
DIVALIKE	-1080	2	0.069	0.031	0	1	2165	1.20E-27
DIVALIKE+j	-1073	3	0.063	0.026	0.027	1	2152	8.80E-25
DIVALIKE+w	-1101	3	0.037	0.019	0	0.21	2208	6.80E-37
DIVALIKE+j+w	-1227	4	0.021	0.04	0.0082	0.035	2461	5.30E-92
BAYAREALIKE	-1094	2	0.056	0.095	0	1	2191	2.10E-33
BAYAREALIKE+j	**-1017**	**3**	**0.034**	**0.038**	**0.11**	**1**	**2041**	**1**
BAYAREALIKE+w	-1106	3	0.035	0.077	0	0.27	2218	3.10E-39
BAYAREALIKE+j+w	-1115	4	0.0054	0.045	0.02	0.0006	2238	1.70E-43
"alternative tree"								
DEC	-1127	2	0.038	0.02	0	1	2257	2.10E-28
DEC+j	-1114	3	0.034	0.016	0.037	1	2234	3.20E-23
DEC+w	-1099	3	0.026	0.013	0	0.33	2204	9.20E-17
DEC+j+w	-1249	4	0.0099	0.025	0.0009	0.014	2505	3.20E-82
DIVALIKE	-1161	2	0.047	0.024	0	1	2325	4.10E-43
DIVALIKE+j	-1146	3	0.041	0.017	0.038	1	2298	3.90E-37
DIVALIKE+w	-1165	3	0.031	0.02	0	0.25	2335	2.70E-45
DIVALIKE+j+w	-1305	4	0.012	0.031	0.0068	0.018	2618	1.20E-106
BAYAREALIKE	-1147	2	0.051	0.08	0	1	2299	2.10E-37
BAYAREALIKE+j	**-1062**	**3**	**0.022**	**0.029**	**0.11**	**1**	**2130**	**1**
BAYAREALIKE+w	-1172	3	0.021	0.056	0	0.23	2351	1.10E-48
BAYAREALIKE+j+w	-1199	4	0.01	0.056	0.012	0.011	2407	7.10E-61

LnL: LogLikelihood; d: dispersal; e: extinction; j: founder-speciation; w: dispersal matrix power exponential; AICc: Corrected Akaike Information Criterion.

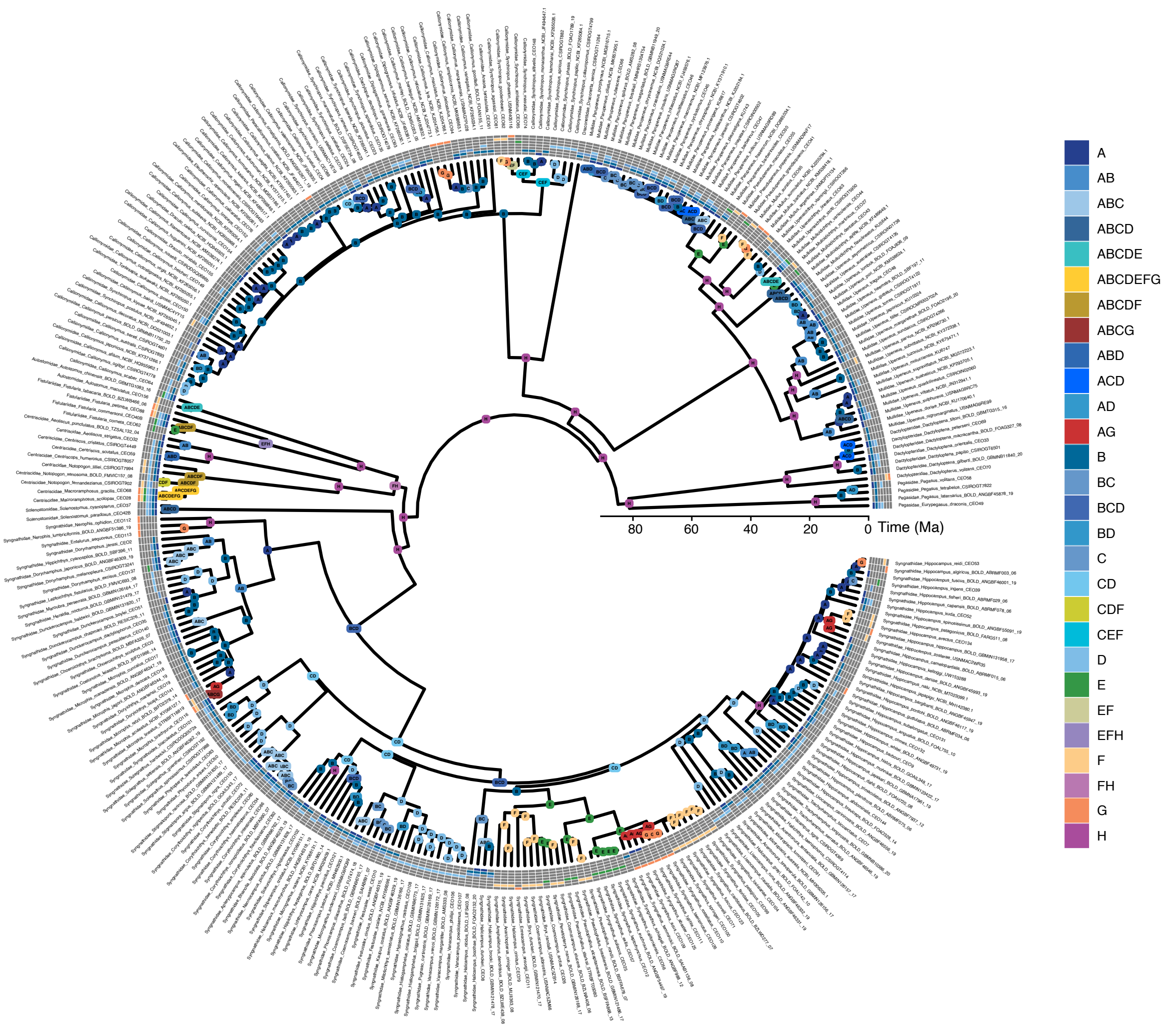


Figure S6. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model (BAYAREA+*j*) applied to the 28 RelTime trees using the "master tree" as a fixed topology in BioGeoBEARS. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; G: Eastern Atlantic; and H: Tethys sea. Ma: millions of years.

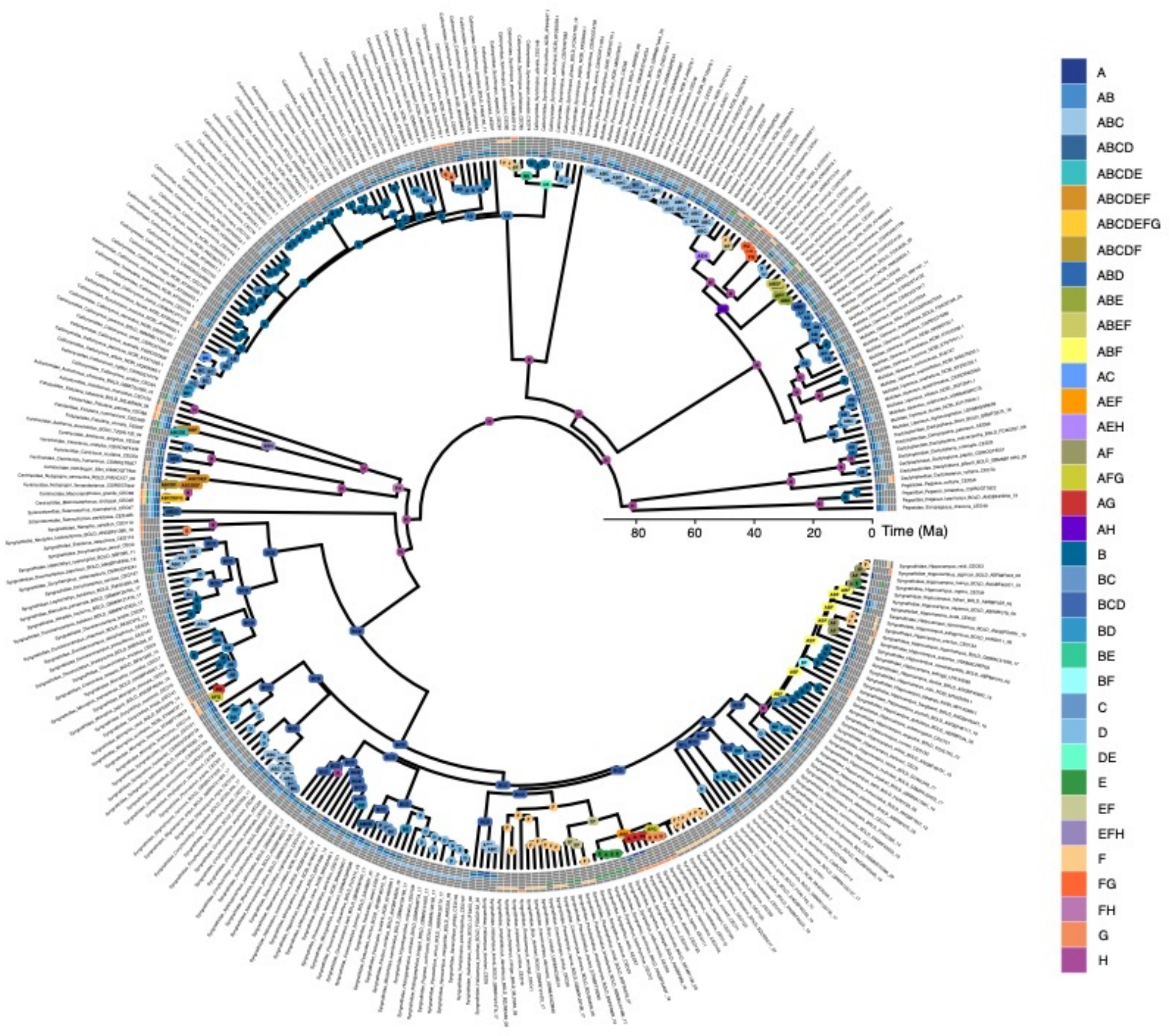


Figure S7. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model (BAYAREA-*j*) applied to the 28 RelTime trees using the "master tree" as a fixed topology in BioGeoBEARS. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; G: Eastern Atlantic; and H: Tethys sea. Ma: millions of years.

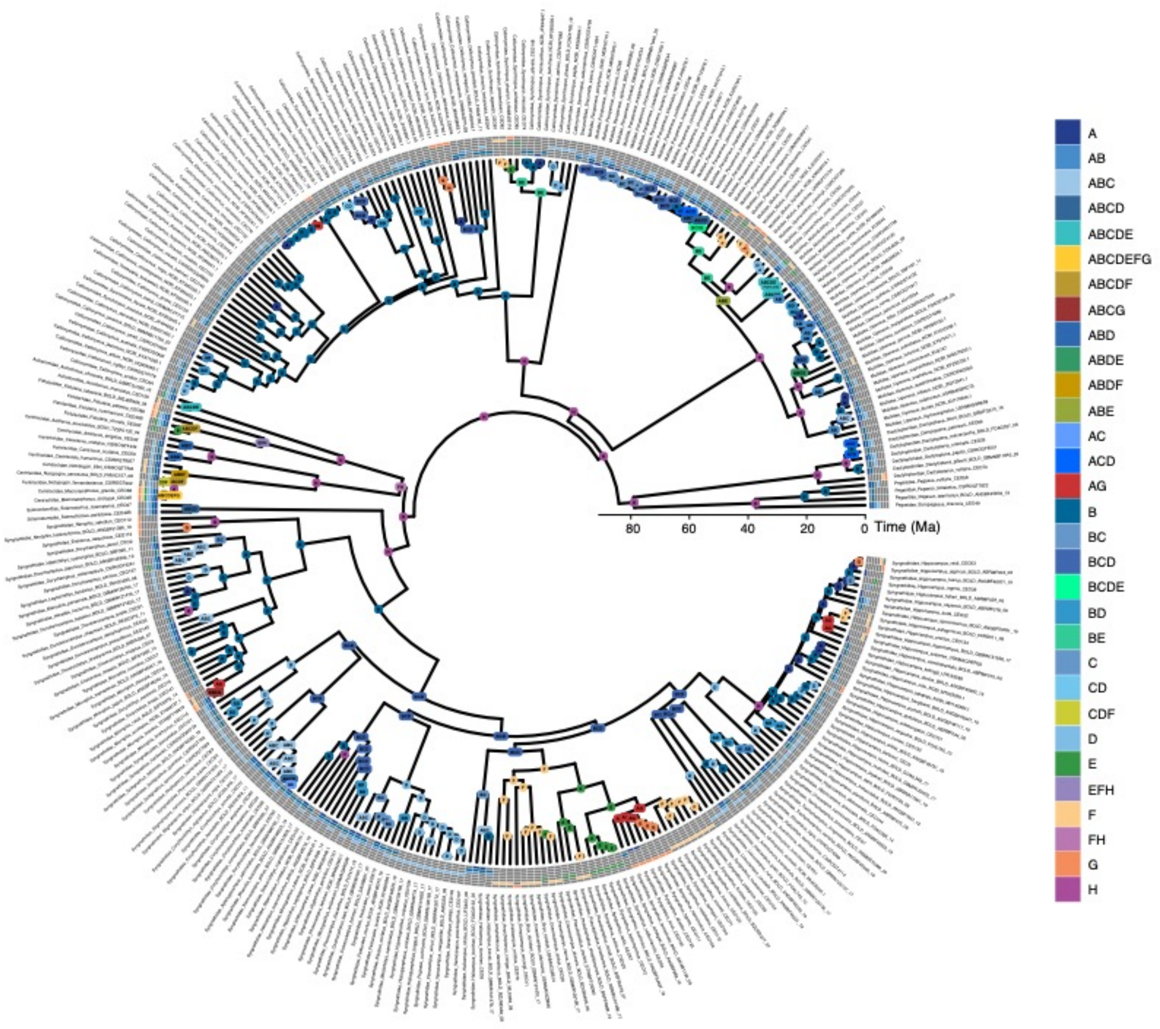


Figure S8. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model (BAYAREA+*j*) applied to the 28 MCMCTree trees using the "alternative tree" as a fixed topology in BioGeoBEARS. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; G: Eastern Atlantic; and H: Tethys sea. Ma: millions of years.

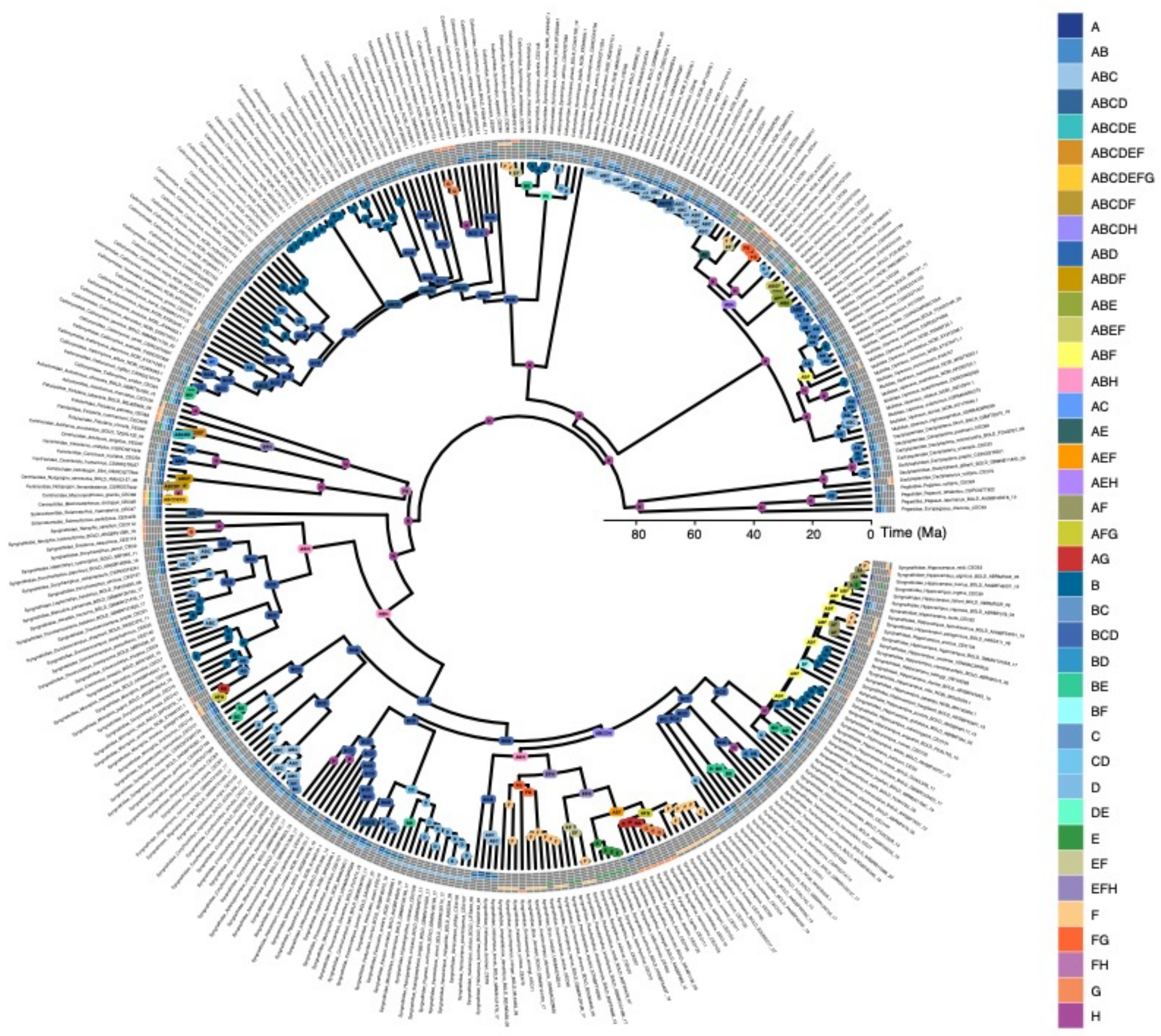


Figure S9. Ancestral range estimations for Syngnatharia using the best-supported biogeographic model (BAYAREA-*j*) applied to the 28 MCMCTree trees using the "alternative tree" as a fixed topology in BioGeoBEARS. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; G: Eastern Atlantic; and H: Tethys sea. Ma: millions of years.

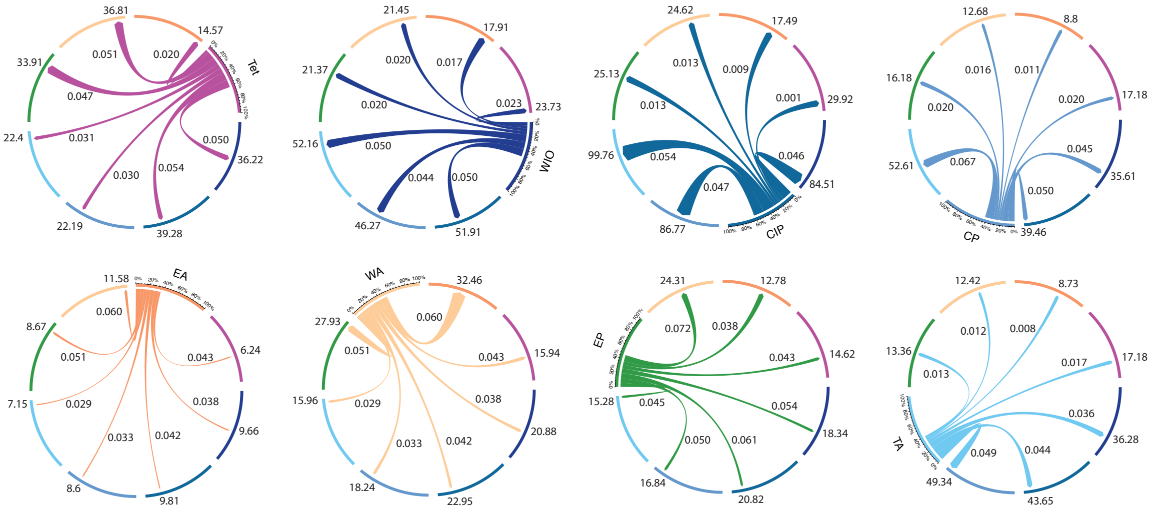


Figure S10. Number of dispersal events and dispersal rates estimated for different biogeographic regions. Chord diagrams denote dispersal events outside each region; line width represents the percentage of lineages dispersing from a focal area to the rest of the areas. Number of lineages dispersed and dispersal rates are depicted outside and inside the diagrams, respectively. These numbers are averaged between 100 biogeographic stochastic histories estimated with the "master tree" and the best-fit biogeographic model, BAYAREA-*j*. Tet: Tethys Sea; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; WA: Western Atlantic; and EA: Eastern Atlantic.

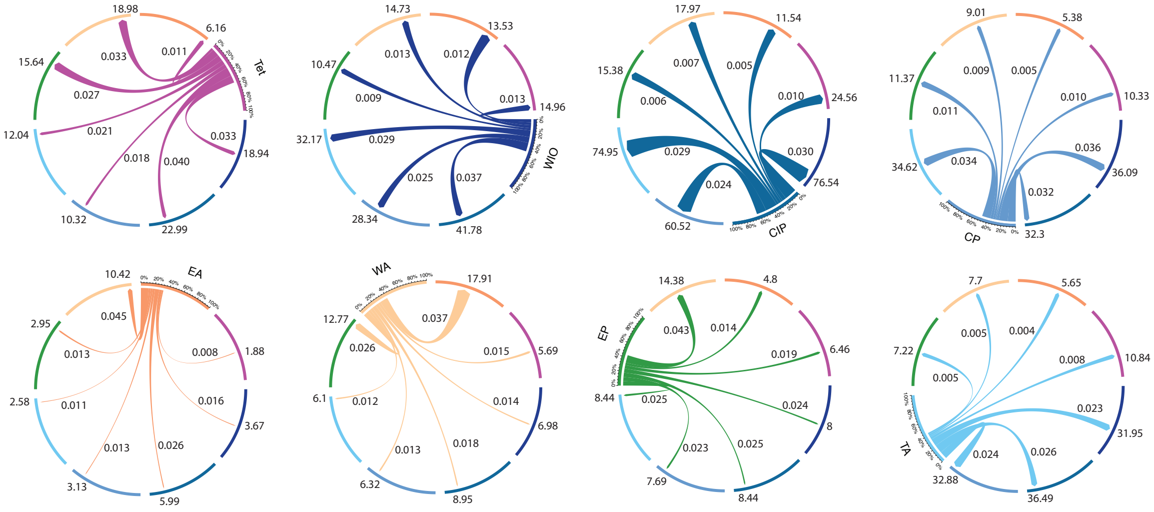


Figure S11. Number of dispersal events and dispersal rates estimated for different biogeographic regions. Chord diagrams denote dispersal events outside each region; line width represents the percentage of lineages dispersing from a focal area to the rest of the areas. Number of lineages dispersed and dispersal rates are depicted outside and inside the diagrams, respectively. These numbers are averaged between 100 biogeographic stochastic histories estimated with the "alternative tree" and the best-fit biogeographic model, BAYAREA+*j*. Tet: Tethys Sea; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; WA: Western Atlantic; and EA: Eastern Atlantic.

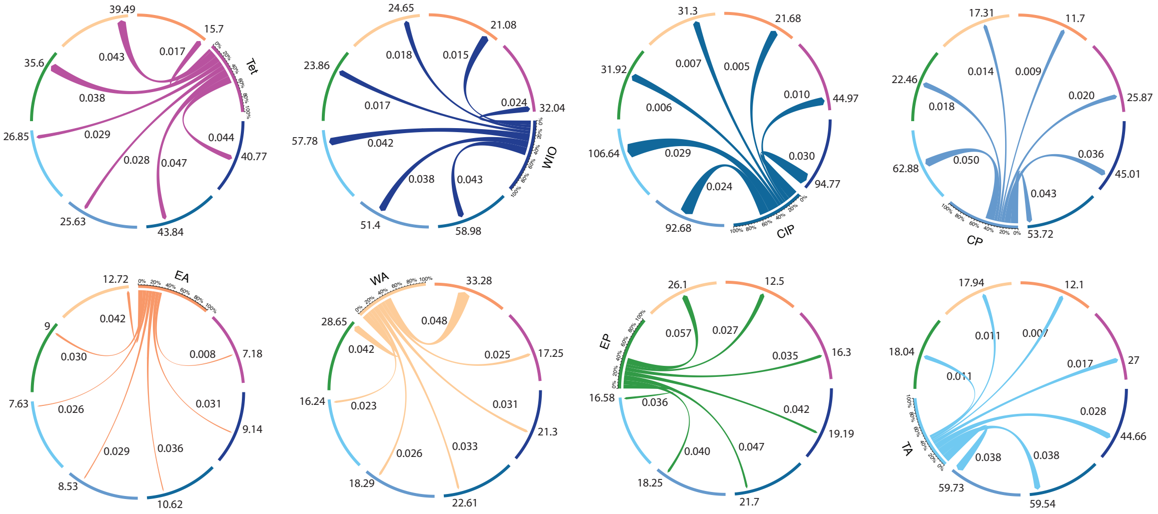


Figure S12. Number of dispersal events and dispersal rates estimated for different biogeographic regions. Chord diagrams denote dispersal events outside each region; line width represents the percentage of lineages dispersing from a focal area to the rest of the areas. Number of lineages dispersed and dispersal rates are depicted outside and inside the diagrams, respectively. These numbers are averaged between 100 biogeographic stochastic histories estimated with the "alternative tree" and the best-fit biogeographic model, BAYAREA-*j*. Tet: Tethys Sea; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; TA: Temperate Australasia; EP: Eastern Pacific; WA: Western Atlantic; and EA: Eastern Atlantic.

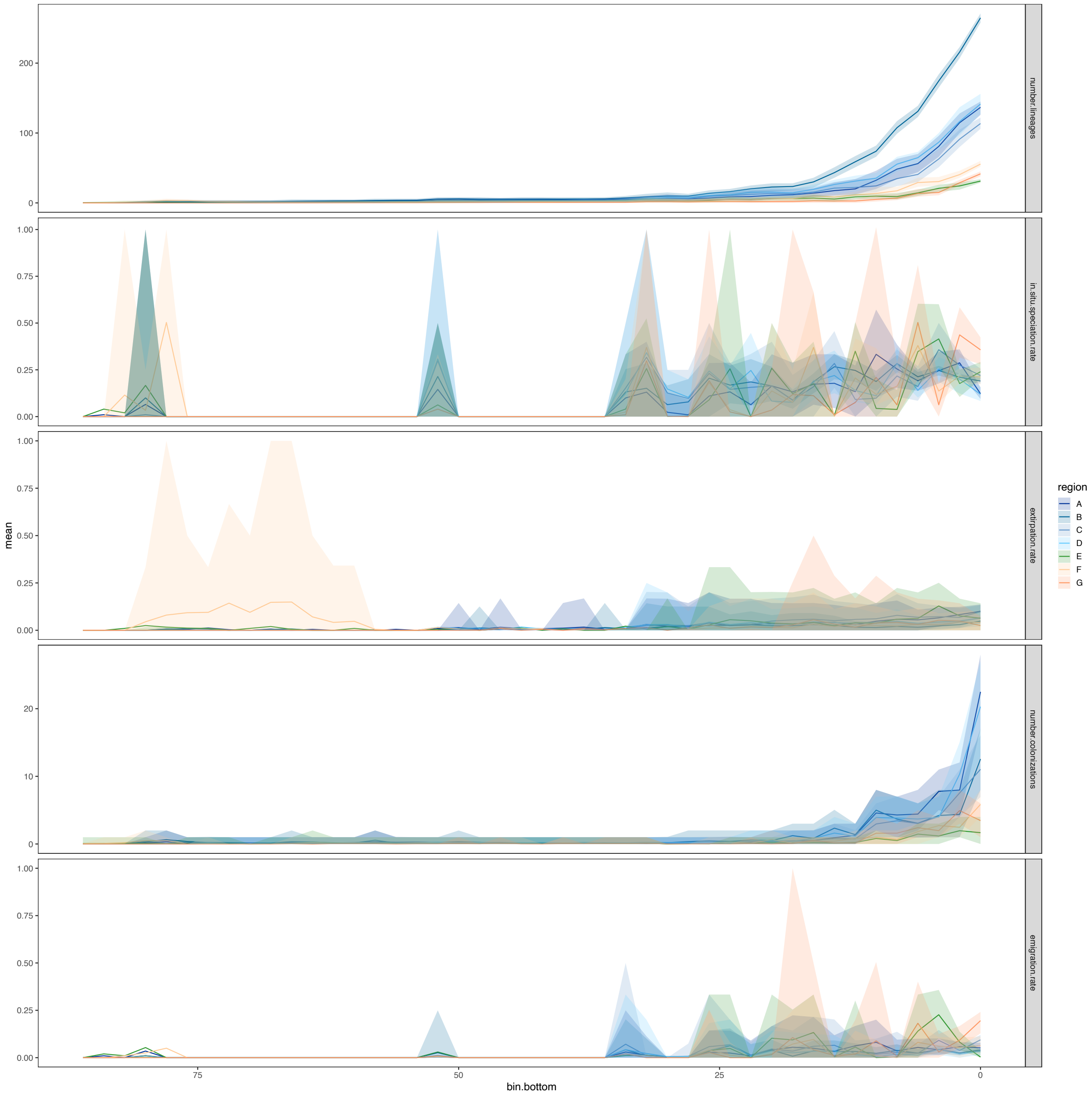


Figure S13. Frequency and timing of colonization events between the seven biogeographic regions based on 100 stochastic histories, as estimated using the "master tree" and the best-fit biogeographic model, BAYAREA+*j*. Lines represent mean values and the shaded area denotes the 95% confidence interval (CI). Note that "emigration" refers to jump dispersal. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; and G: Eastern Atlantic.

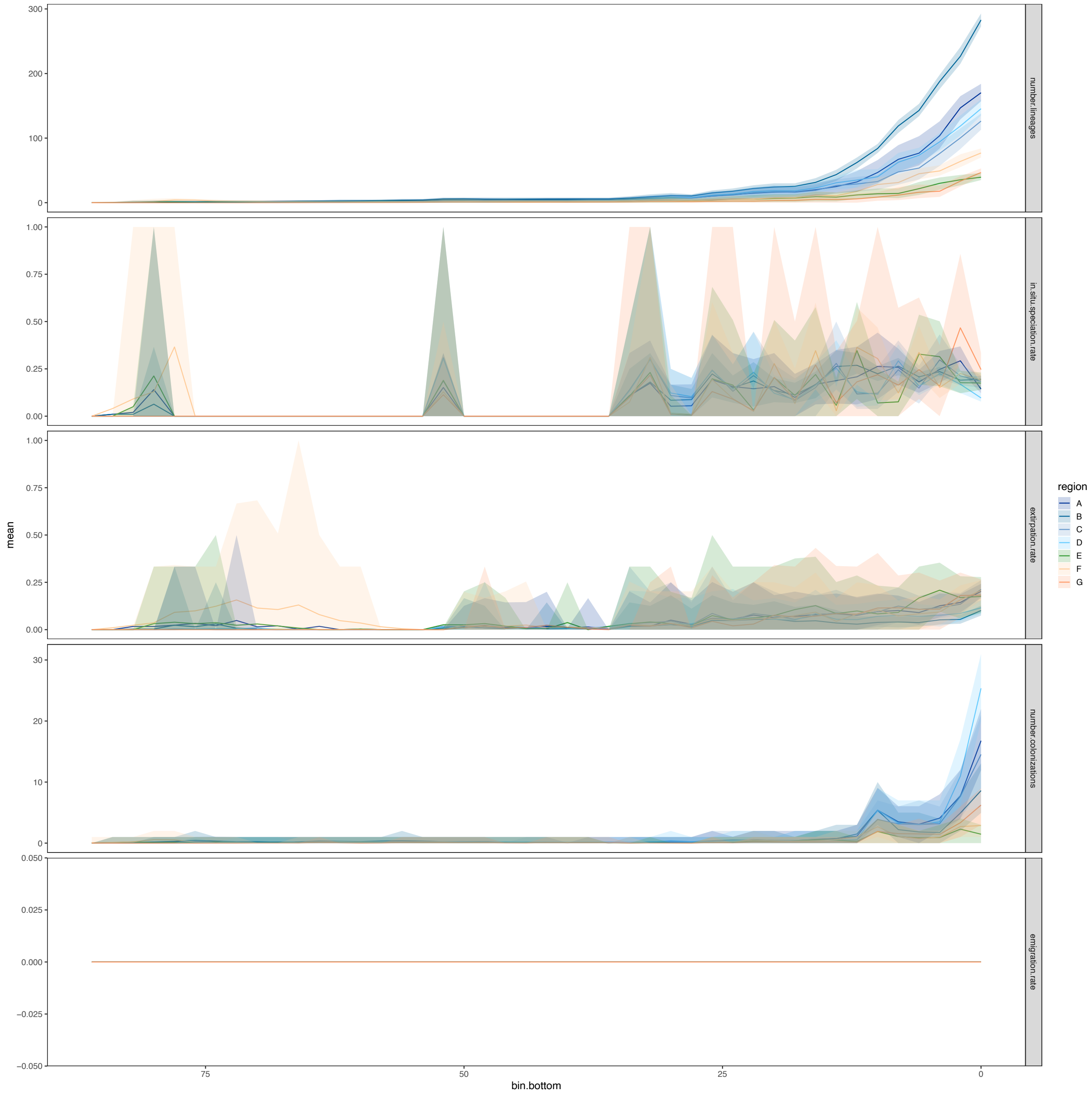


Figure S14. Frequency and timing of colonization events between the seven biogeographic regions based on 100 stochastic histories, as estimated using the "master tree" and the best-fit biogeographic model, BAYAREA-*j*. Lines represent mean values and the shaded area denotes the 95% confidence interval (CI). Note that "emigration" refers to jump dispersal. In biogeographic models without jump dispersal, lineages only leave a region via extirpation. Therefore, emigration rates in this analysis are zero, as the *j* parameter was not used. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; and G: Eastern Atlantic.

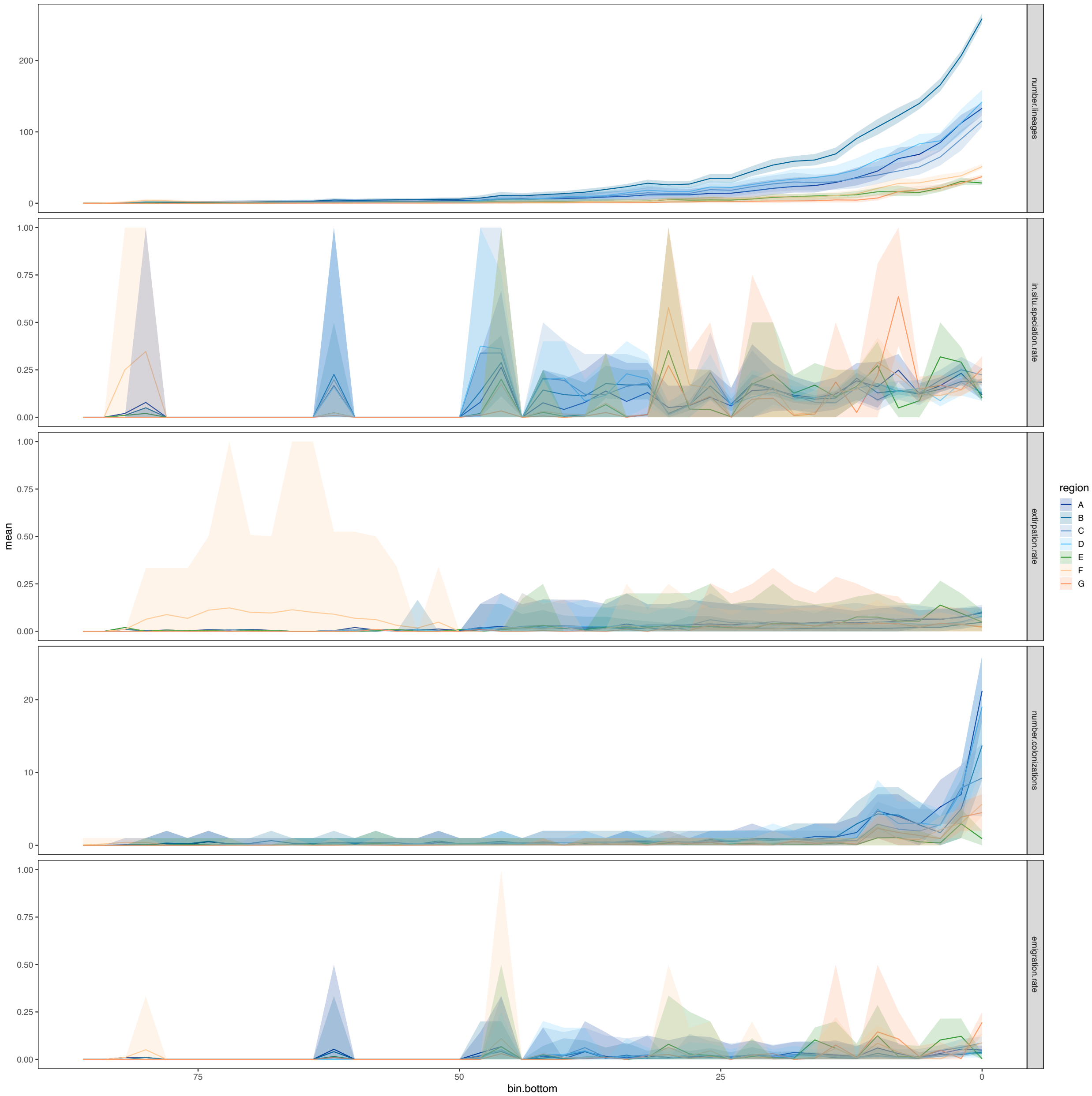


Figure S15. Frequency and timing of colonization events between the seven biogeographic regions based on 100 stochastic histories, as estimated using the "alternative tree" and the best-fit biogeographic model, BAYAREA+*j*. Lines represent mean values and the shaded area denotes the 95% confidence interval (CI). Note that "emigration" refers to jump dispersal. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; and G: Eastern Atlantic.

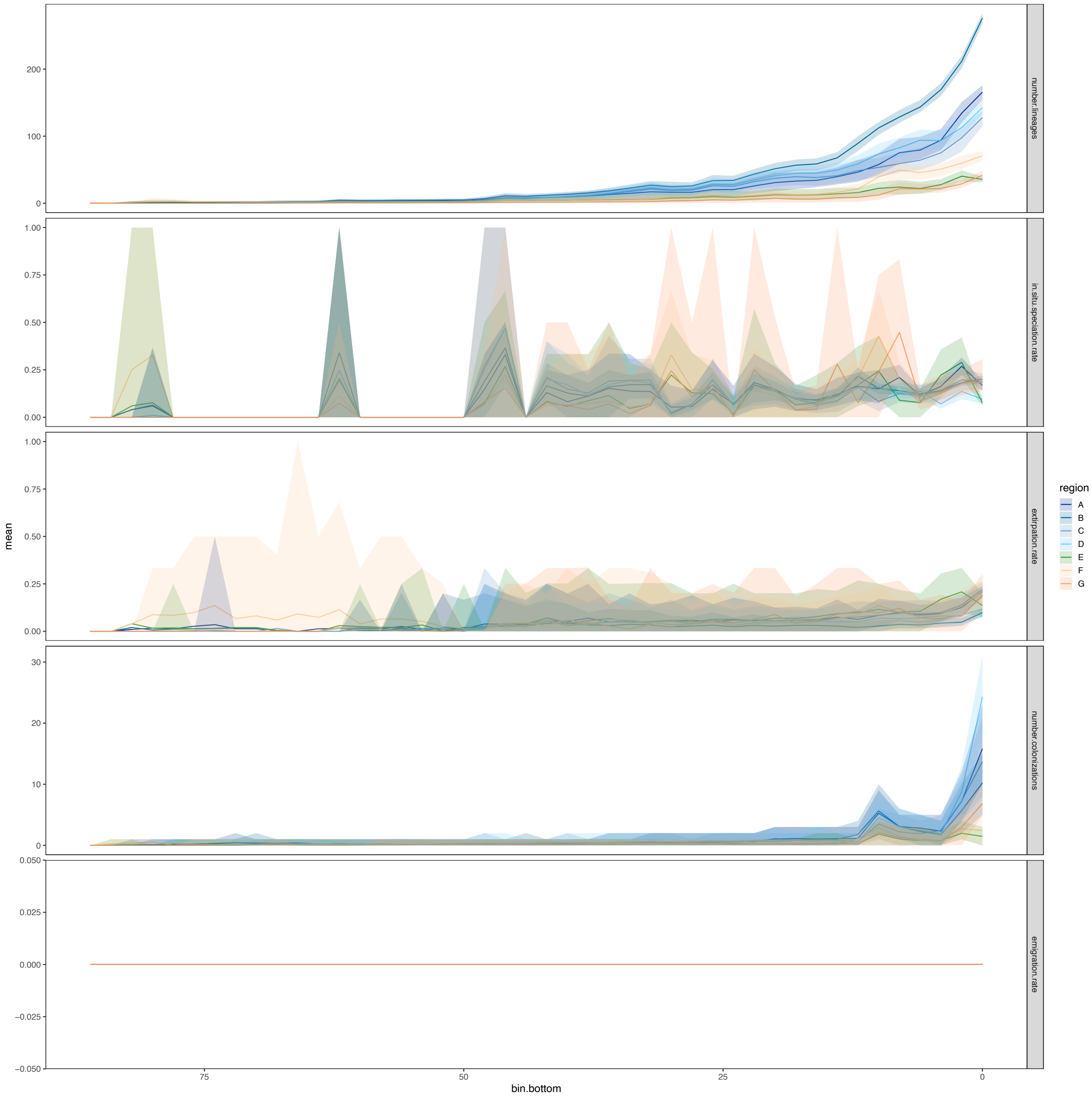


Figure S16. Frequency and timing of colonization events between the seven biogeographic regions based on 100 stochastic histories, as estimated using the "alternative tree" and the best-fit biogeographic model, BAYAREA-*j*. Lines represent mean values and the shaded area denotes the 95% confidence interval (CI). Note that "emigration" refers to jump dispersal. In biogeographic models without jump dispersal, lineages only leave a region via extirpation. Therefore, emigration rates in this analysis are zero, as the *j* parameter was not used. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; and G: Eastern Atlantic.

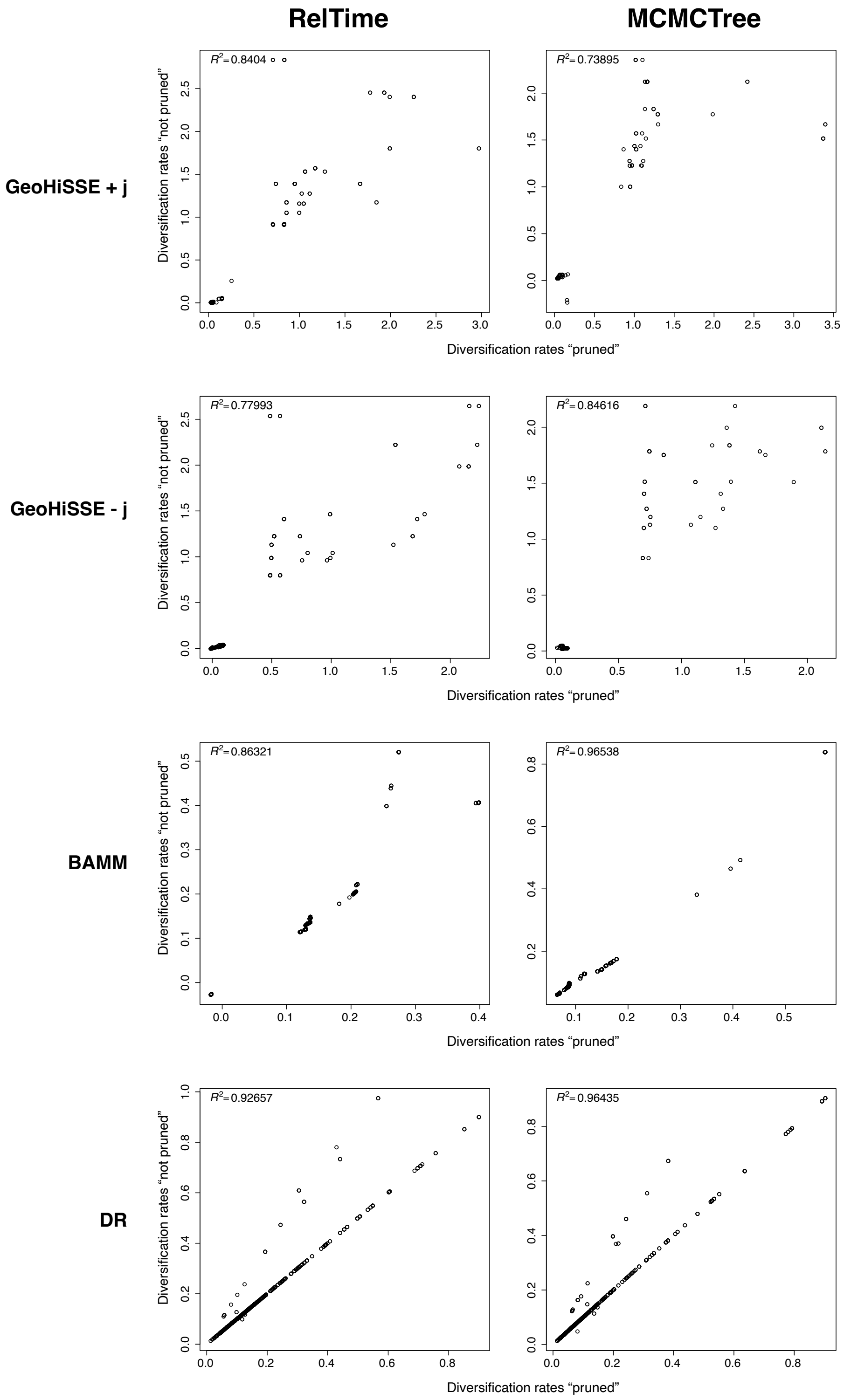


Figure S17. Correlation of lineage diversification rates between phylogenies using all 323 taxa (denoted as “not pruned”) and after excluding potential instances of taxonomic over-splitting (denoted as “pruned”; 311 taxa) and dated with two calibration methods, RelTime and MCMCTree. Diversification rates were analyzed using three different methods (GeoHiSSE with and without the *j* parameter, BAMM and DR).

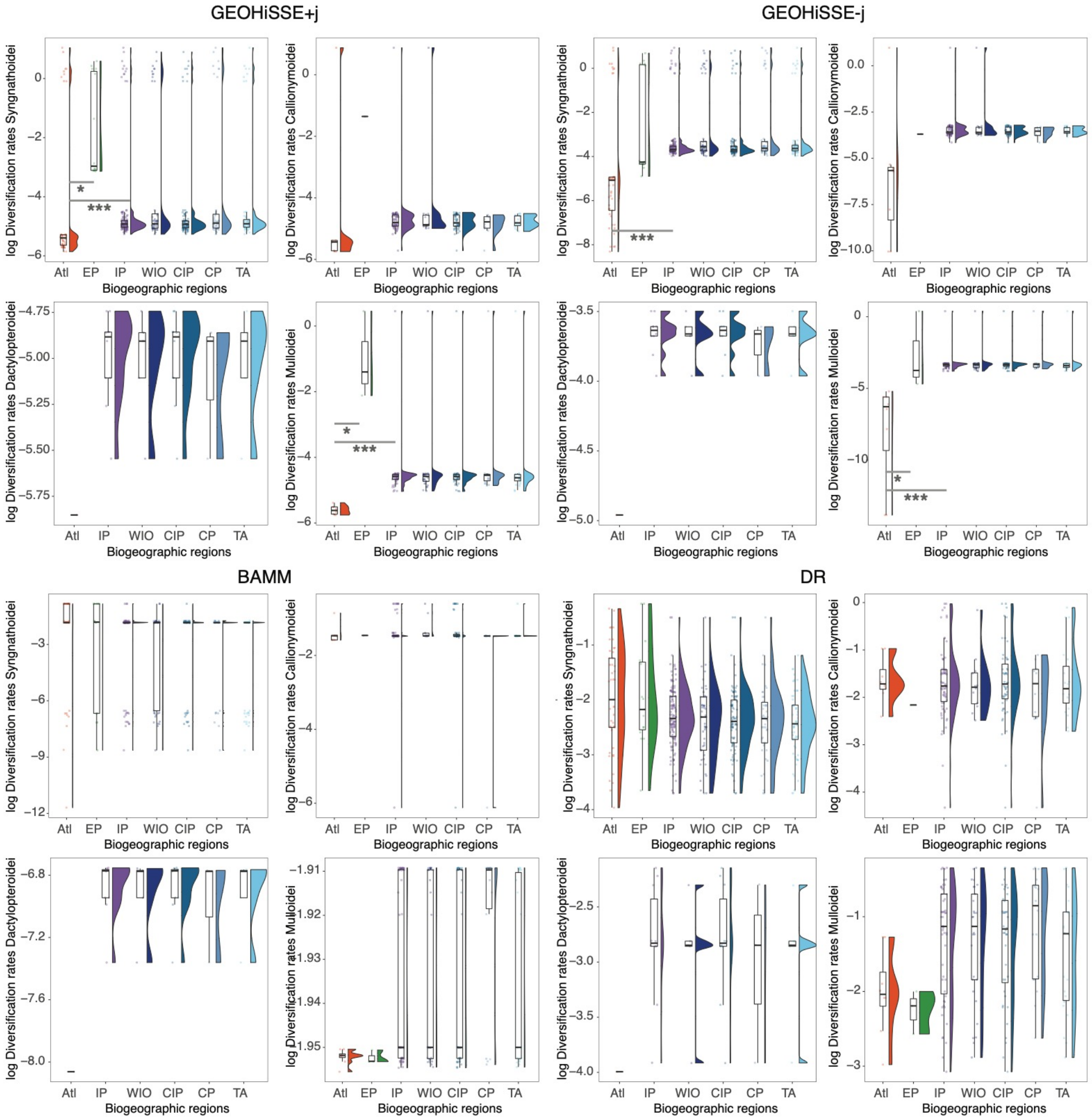


Figure S18. Average log-transformed rates of lineage diversification for syngnatharian suborders, as estimated using three different approaches (GeoHiSSE±*j*, BAMM, and DR) and depicted using raincloud plots (half-violin plots and boxplots). Average rates were estimated based on values obtained with the 28 RelTime trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Note the several outliers with high diversification rates (1.05–2.83; log-transformed 0.05–1.04) on results obtained using GeoHiSSE±*j*. These outliers correspond to species with widespread distributions. Widespread species within the order Syngnathoidei: *Aulostomus chinensis, Centriscops humerosus, Macroramphosus gracilis, Macroramphosus scolopax, Notopogon fernandezianus, Notopogon lilliei, Fistularia commersonii, Fistularia petimba, Doryrhamphus excisus, Microphis brachyurus, Syngnathus acus, Syngnathus temminckii*; within Callionymoidei: *Paracallionymus costatus*; and within Mulloidei: *Mulloidichthys vanicolensis*. Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; see Table S6 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

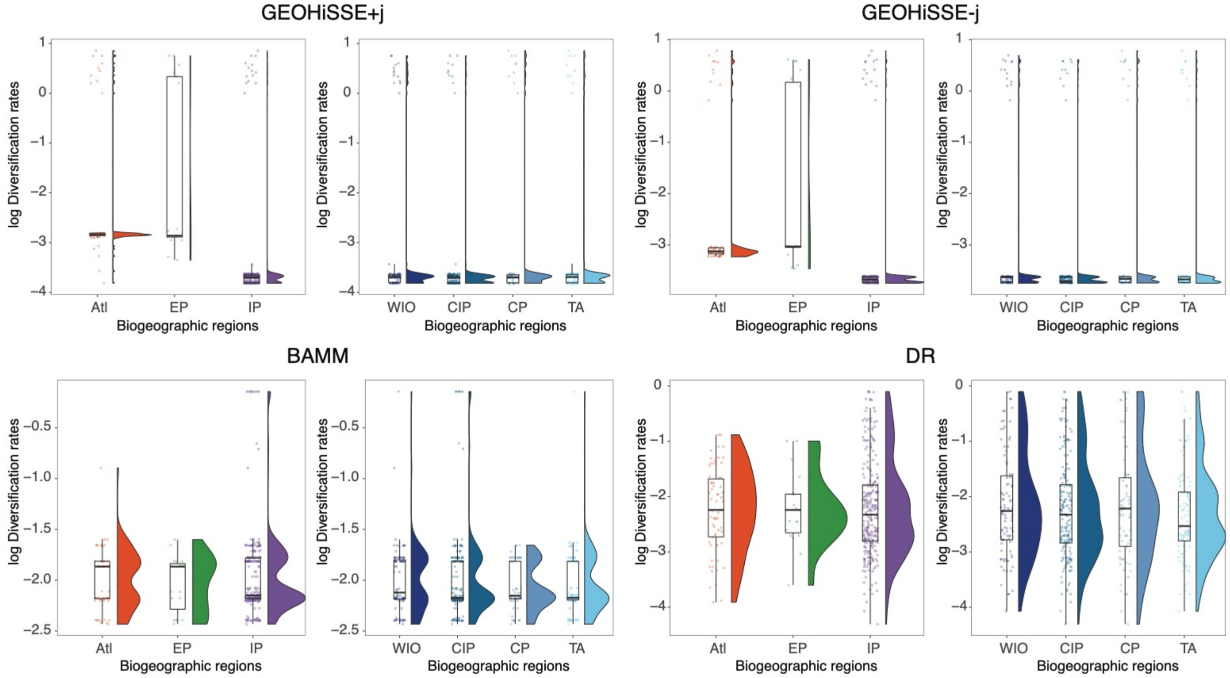


Figure S19. Average log-transformed rates of lineage diversification for all syngnatharians, as estimated using three different approaches (GeoHiSSE±*j*, BAMM, and DR) and depicted using raincloud plots (half-violin plots and boxplots). Average rates were estimated based on values obtained with the 28 MCMCTree trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Note the several outliers with high diversification rates (1.05–2.83; log-transformed 0.05–1.04) on results obtained using GeoHiSSE±*j*. These outliers correspond to species with widespread distributions. Widespread species within the order Syngnathoidei: *Aulostomus chinensis, Centriscops humerosus, Macroramphosus gracilis, Macroramphosus scolopax, Notopogon fernandezianus, Notopogon lilliei, Fistularia commersonii, Fistularia petimba, Doryrhamphus excisus, Microphis brachyurus, Syngnathus acus, Syngnathus temminckii*; within Callionymoidei: *Paracallionymus costatus*; and within Mulloidei: *Mulloidichthys vanicolensis*. Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; see Table S6 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

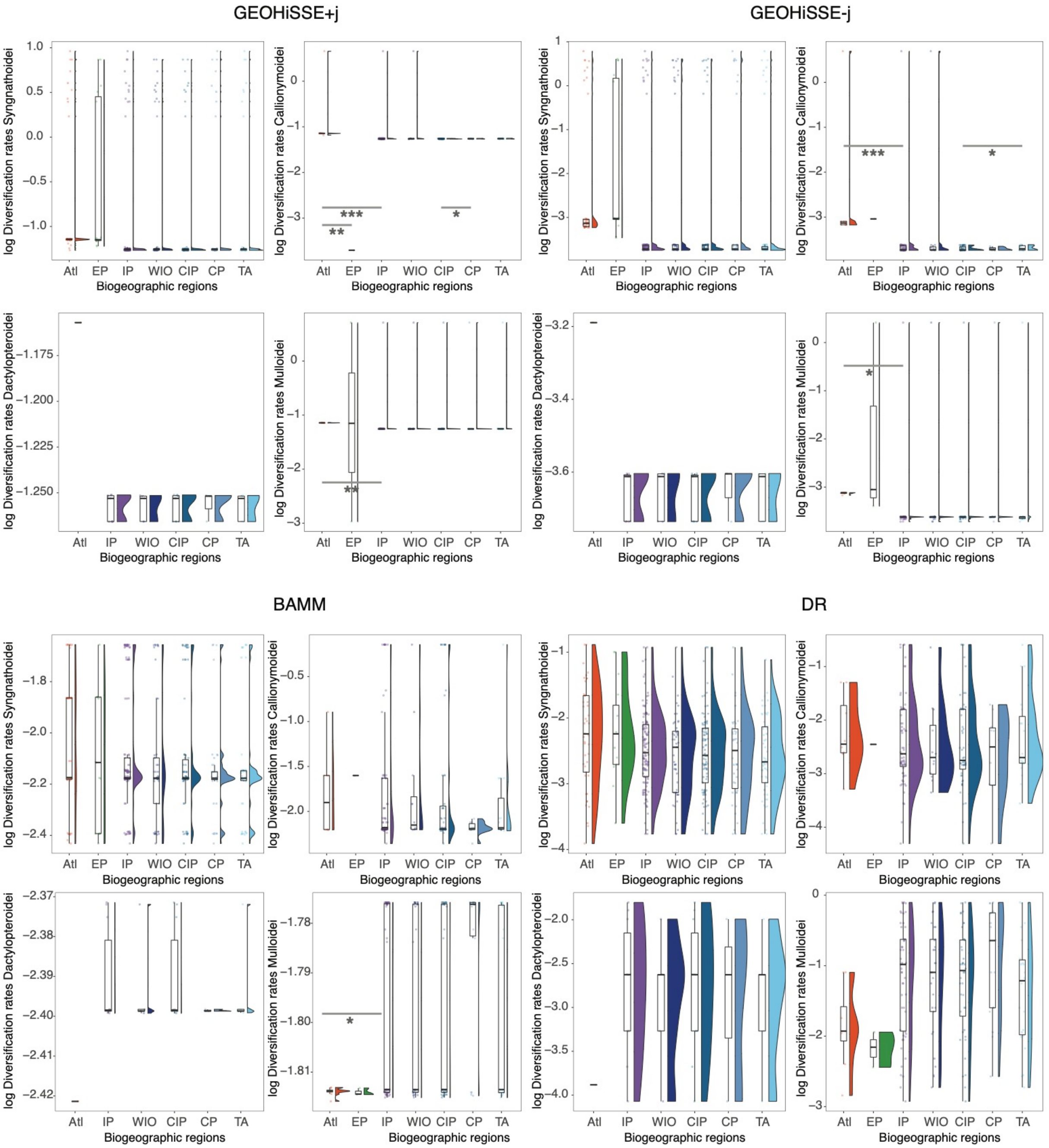


Figure S20. Average log-transformed rates of lineage diversification for syngnatharian suborders, as estimated using three different approaches (GeoHiSSE±*j*, BAMM, and DR) and depicted using raincloud plots (half-violin plots and boxplots). Average rates were estimated based on values obtained with the 28 MCMCTree trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Note the several outliers with high diversification rates (1.05–2.83; log-transformed 0.05–1.04) on results obtained using GeoHiSSE±*j*. These outliers correspond to species with widespread distributions. Widespread species within the order Syngnathoidei: *Aulostomus chinensis, Centriscops humerosus, Macroramphosus gracilis, Macroramphosus scolopax, Notopogon fernandezianus, Notopogon* *lilliei, Fistularia commersonii, Fistularia petimba, Doryrhamphus excisus, Microphis brachyurus, Syngnathus acus, Syngnathus temminckii*; within Callionymoidei: *Paracallionymus costatus*; and within Mulloidei: *Mulloidichthys vanicolensis*. Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$; see Table S6 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

Table S6. Statistical analysis of syngnatharian diversification rates across biogeographic regions using BAMM, DR, and GeoHiSSE methods. Red indicates statistical significance ($p \leq 0.05$).

RelTime	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**		
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	
BAMM	0.8771	0.1631	0.9618	0.1496	0.1783	0.8733	0.1138	0.6891	0.02749	0.1343	
DR	0.6883	0.8509	0.4299	0.5057	0.3742	0.6279	0.2059	0.8529	0.6989	0.8741	
GeoHiSSE *-j*	4.01E-07	0.1357	6.06E-05	0.2927	0.05351	0.1497	0.113	0.423	0.0002174	0.6914	
GeoHiSSE *+j*	9.36E-06	0.155	0.0002378	0.2445	0.2691	0.06992	0.1138	0.4288	0.003298	0.361	
Areas:	**GeoHiSSE *-j***	**GeoHiSSE *+j***	**GeoHiSSE *-j***	**GeoHiSSE *+j***					**BAMM**	**GeoHiSSE *-j***	**GeoHiSSE *+j***
At.:EP	0.047	0.0033	0.33	0.04565					1	0.013	0.0341
At.:IP	9.50E-08	7.60E-06	2.20E-05	0.00013					0.051	5.10E-05	0.0023
EP:IP	1	1	0.74	1					0.369	1	1

MCMCTree	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**		
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	
BAMM	0.01728	0.153	0.07078	0.3081	0.3985	0.8781	0.7518	0.5913	0.01126	0.1183	
DR	0.7575	0.8982	0.2331	0.7194	0.3671	0.6366	0.2049	0.8312	0.7688	0.87	
GeoHiSSE *-j*	0.9382	0.5932	0.3272	0.2458	2.62E-05	0.02042	0.1138	0.7217	0.02951	0.4119	
GeoHiSSE *+j*	0.2936	0.5665	0.7062	0.1279	6.28E-05	0.01169	0.113	0.6912	0.0068	0.5506	
Areas:		**BAMM**	**GeoHiSSE *-j***	**GeoHiSSE *+j***	**GeoHiSSE *-j***	**GeoHiSSE *+j***			**BAMM**	**GeoHiSSE *-j***	**GeoHiSSE *+j***
At.:EP		1			1	0.0081			1	1	0.0621
At.:IP		0.116			1.10E-05	3.00E-05			0.024	0.035	0.0053
EP:IP		0.082			0.12	0.4079			0.205	0.705	1
CP:CIP					0.129	0.04					
TA:CIP					0.049	0.06					
WIO:CIP					1	1					
TA:CP					1	1					
WIO:TA					1	1					
WIO:CP					1	0.74					

Morphological disparity and rates by region

Contemporary trait disparity analyses across regions (using sum of variances obtained from pPC scores based on the "master tree") reveal that syngnatharian head morphology is more disparate in the IP (0.055) and EP (0.057) compared to the Atl. (0.048), while body morphology varies more in the Atl. (0.039) and the EP (0.045) than in the IP (0.030) (Fig. 2.4). Disparity among subareas within the IP is rather similar although head shape has lower disparity than body morphology [head: WIO (0.058); CP (0.058); CIP (0.056); TA (0.056), body: WIO (0.034); TA (0.032); CIP (0.028); CP (0.028)] (Fig. 2.4). Suborder-level analyses do not show congruent patterns, with disparity across biogeographic regions varying depending on the clade (Fig. S21). We find similar results when using pPC scores based on the RelTime tree (Fig. 2.4 and Fig. S21), pPC scores based on the MCMCTree tree (Fig. S22), and uncorrected PC scores (Fig. S23). Morphospace occupation analyses show a substantial overlap between the three major realms and within IP subareas for both head-only and full-body shape datasets (Fig. 2.4, Fig. S24 for details on the remaining pPC axes). For head morphology (Fig. 2.4A, see Appendix 2), the greatest morphospace overlap is between IP and Atl. (Jaccard= 0.236, Sørenson=0.382), followed by EP and Atl. (Jaccard= 0.217, Sørenson=0.358), and IP and EP (Jaccard= 0.094, Sørenson=0.172). Additionally, there is substantially higher head shape similarity within IP subareas (Jaccard= 0.453–0.634, Sørenson=0.624–0.776), specially between CIP and TA (Jaccard= 0.634, Sørenson= 0.776). All suborders also show a higher degree of overlap between these two IP subareas (Jaccard= 0.222–0.648, Sørenson=0.364–0.786), although syngnathoid species also exhibit similarly high overlap between WIO and CP (Jaccard= 0.624, Sørenson=0.769) and between WIO and CIP (Jaccard= 0.504, Sørenson=0.670). Regardless of the clade (Syngnatharia vs. suborders), morphospace overlap based on body morphology is lower than that of head shape (Fig. 2.4). The greatest overlap in body shape (Fig. 2.4B, see Appendix 2) is found between EP and Atl. (Jaccard= 0.135, Sørenson=0.238) followed by IP and Atl. (Jaccard= 0.094, Sørenson=0.173), and IP and EP (Jaccard= 0.024, Sørenson=0.047). Within the IP, the highest overlap exists between WIO and CIP (Jaccard= 0.371, Sørenson=0.541) and CIP and TA (Jaccard= 0.271, Sørenson=0.427). Morphospace overlap calculated for each suborder's body shape shows similar patterns as the head, where the overlap between CIP and TA is the highest (Jaccard= 0.115–0.650, Sørenson=0.207–0.787), with the exception of Syngnathoidei whose overlap is higher between WIO and CIP (Jaccard= 0.506, Sørenson=0.672) and between WIO and CP (Jaccard= 0.324, Sørenson=0.490) (Figs. S25 and S26). Analyses conducted using pPC scores on the "alternative tree" show almost identical results (Figs. S24, S27, and S28). See Appendix 2 for hypervolume overlap statistic values across analyses based on either all syngnatharians or by suborder.

Multivariate disparity-through-time analyses indicate that a significant proportion of syngnatharian morphological variation originated early in the history of Syngnatharia (ca. 87–78 Ma), followed by a steady reduction in disparity within subclades (Fig. 2.4 and S29). The best supported evolutionary model for the head-only morphology was BMOUi or EBOUi, where an initial BM or EB morphological evolution is followed by OU later in the history of the clade (Akaike information criterion weight or AICw values 0.61 and 0.36, respectively; Fig. S30). For the full body morphology, we obtained similar results when using pPC scores based on RelTime trees (AICw 0.62 and 0.24, respectively; Fig. S30). However, when we used PC scores based on RelTime trees or a combination of PC and pPC scores with MCMCTree trees, BM emerged as the most supported morphological model (AICw 0.82, 0.96 and 0.96, respectively; Fig. S31). Morphospace clustering tends to show adaptive peaks that differentiate suborder- or family-level lineages for both head-only and full-body datasets (Fig. S32). In the head-only dataset, PC1 of the morphospace accounts for 59% of the variation, primarily associated with snout elongation and head depth (Fig. S32). PC2 explains an additional 30%, potentially attributed to the distance between the rostral tip of the premaxilla and the caudal end of the maxilla. In contrast, for the full-body dataset, PC1, contributing to 62%, is linked to the position of the dorsal fin and body depth depression (Fig. S32). However, the pattern for PC2 is less clear, as it explains only 15% of the variation. Surprisingly, body elongation does not appear to be the main factor driving variation along PC1 as observed in other fishes (Claverie and Wainwright 2014).

Morphological evolutionary rates estimated based on pPC scores show rate constancy between the major realms (head: Atl., EP and IP median of ~0.0005, body: Atl. 0.000034; EP and IP 0.000024) and within subareas of the IP (head: WIO, CP, CIP, and TA ~0.0003, body: WIO and CP 0.000019; CIP 0.000024 and TA 0.000022; Fig. 3B). When analyzing all species together, no significant differences are found between biogeographic regions (Table S7). However, within the suborder Syngnathoidei, significant differences exist between the Atl. and the IP (p=0.011) for head-only, a result mostly driven by the high morphological rates in the Atl. for the genus *Syngnathus* (Fig. S33, Table S7). Using PC scores, we observe similar results, except in the suborder Syngnathoidei, in which there is a significant difference, although not high, for the head-only dataset between WIO and TA subareas (p=0.035) and between the Atl. and IP for the full-body dataset (p=0.011); Fig. S34, Table S7). Finally, analyses performed on MCMCTree trees yield similar results when using pPC scores (Fig. S35), while when using PC scores (Fig. S36), Syngnathoidei show significance at the body morphology between the Atl. and EP (p=0.041) and the Atl. and IP (p=0.0002; Table S7).

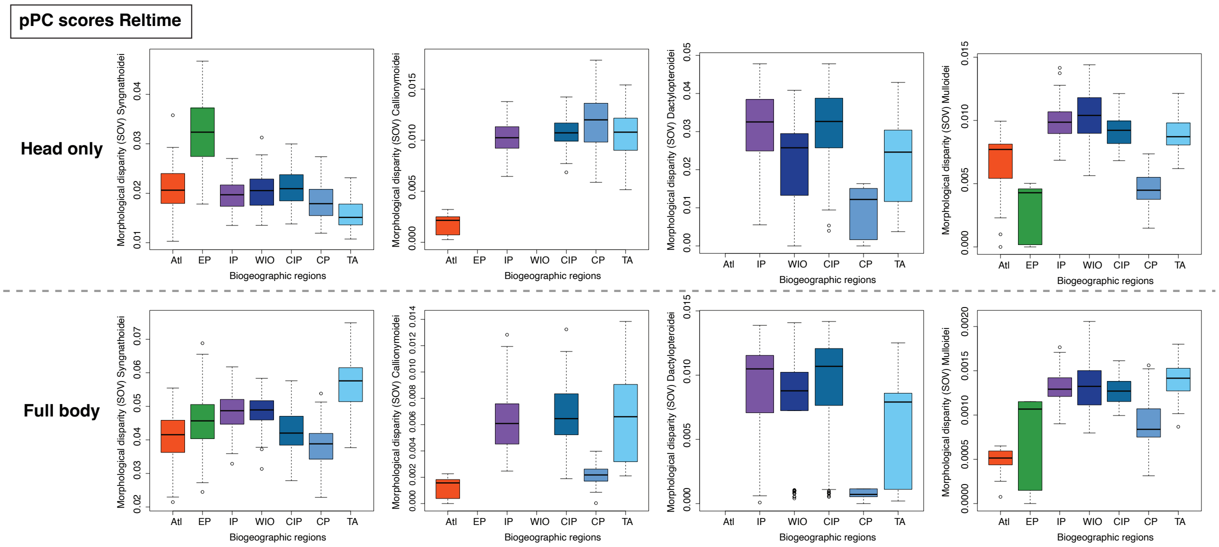


Figure S21. Contemporary morphological disparity across regions for syngnatharian suborders as estimated using sum of variances (SOV). Estimates are based on pPC scores and the “master tree” for head-only and full-body datasets. Note that disparity cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

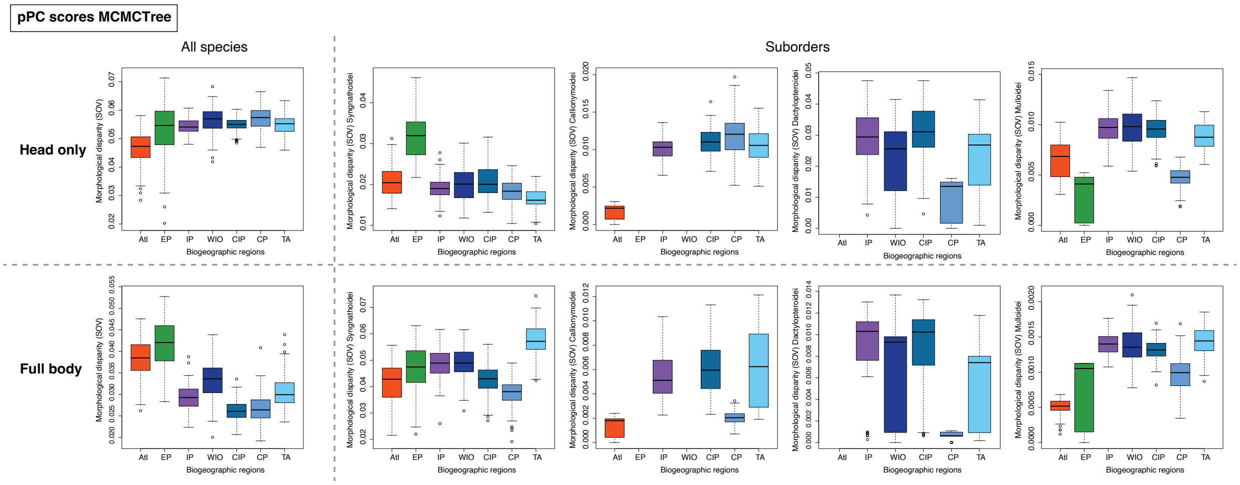


Figure S22. Contemporary morphological disparity across regions for syngnatharian (all species or by suborder) as estimated using sum of variances (SOV). Estimates are based on pPC scores and the "alternative tree" for head-only and full-body datasets. Note that disparity cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

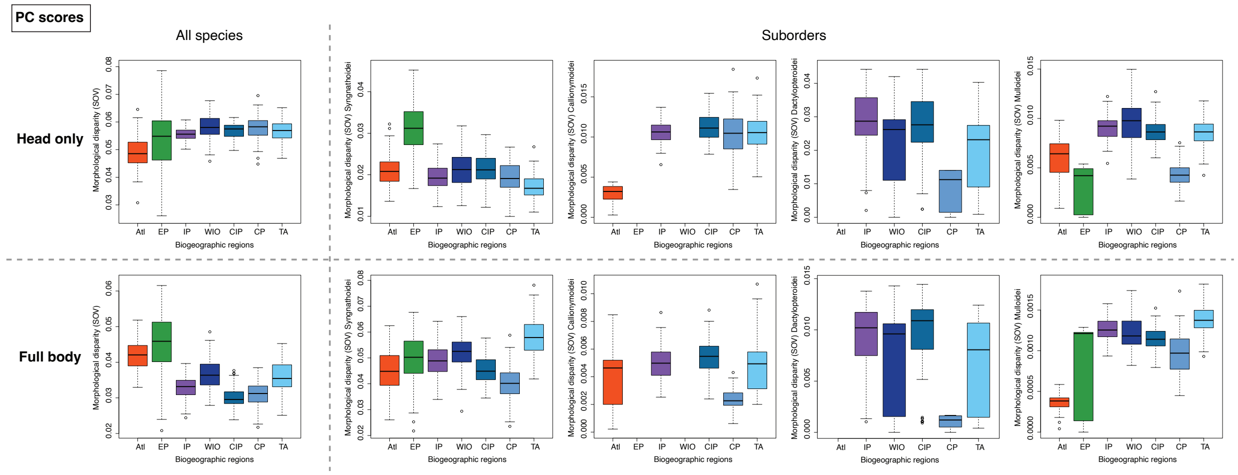
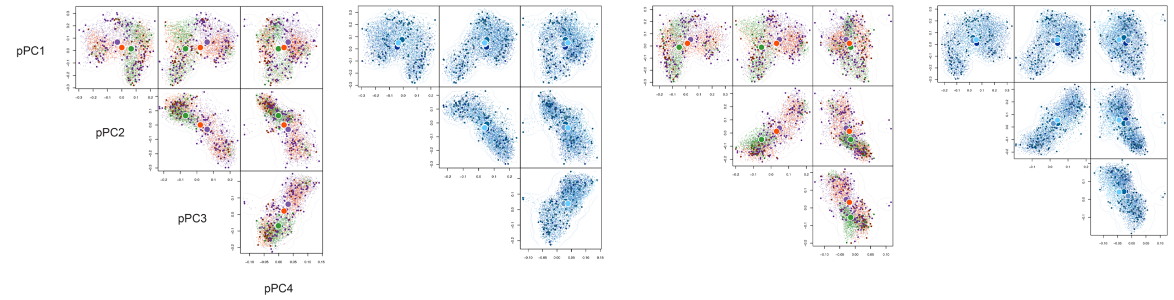
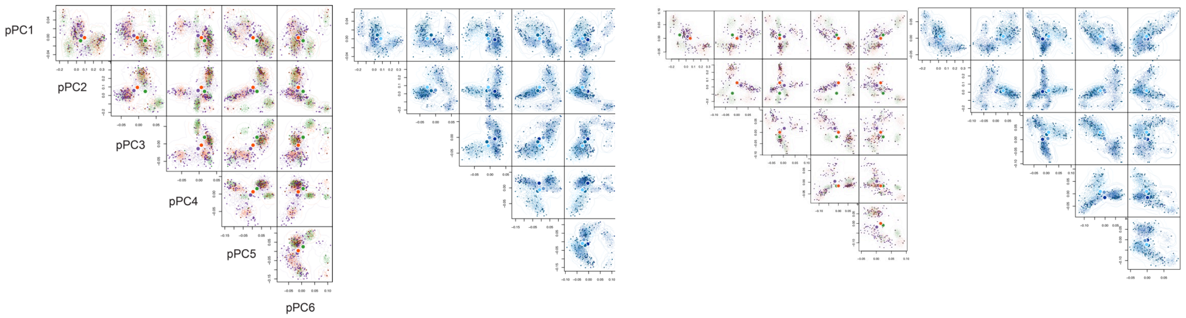


Figure S23. Contemporary morphological disparity across regions for syngnatharian (all species or by suborder) as estimated using sum of variances (SOV). Estimates are based on PC scores for head-only and full-body datasets. Note that disparity cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.


Head only (pPC scores Reltime); Atl EP IP WIO CIP CP TA

Head only (pPC scores MCMCTree); Atl EP IP WIO CIP CP TA


Full body (pPC scores Reltime); Atl EP IP WIO CIP CP TA

Full body (pPC scores MCMCTree); Atl EP IP WIO CIP CP TA

Figure S24. Hypervolumes based on pPC scores obtained with the "master tree" and "alternative tree" for syngnatharians, showing the overlap for head-only (four-dimensional: pPCs 1 to 4) and full-body (six-dimensional: pPCs 1 to 6) datasets across the three major oceanic realms (Atl: Atlantic, EP: Eastern Pacific, IP: Indo-Pacific) and within the Indo-Pacific (WIO: Western Indian Ocean, CIP: Central Indo-Pacific, CP: Central Pacific, TA: Temperate Australasia). Detailed hypervolume overlap statistics (Jaccard, Sørenson, Fraction unique 1, Fraction unique 2) can be found in Appendix 2.

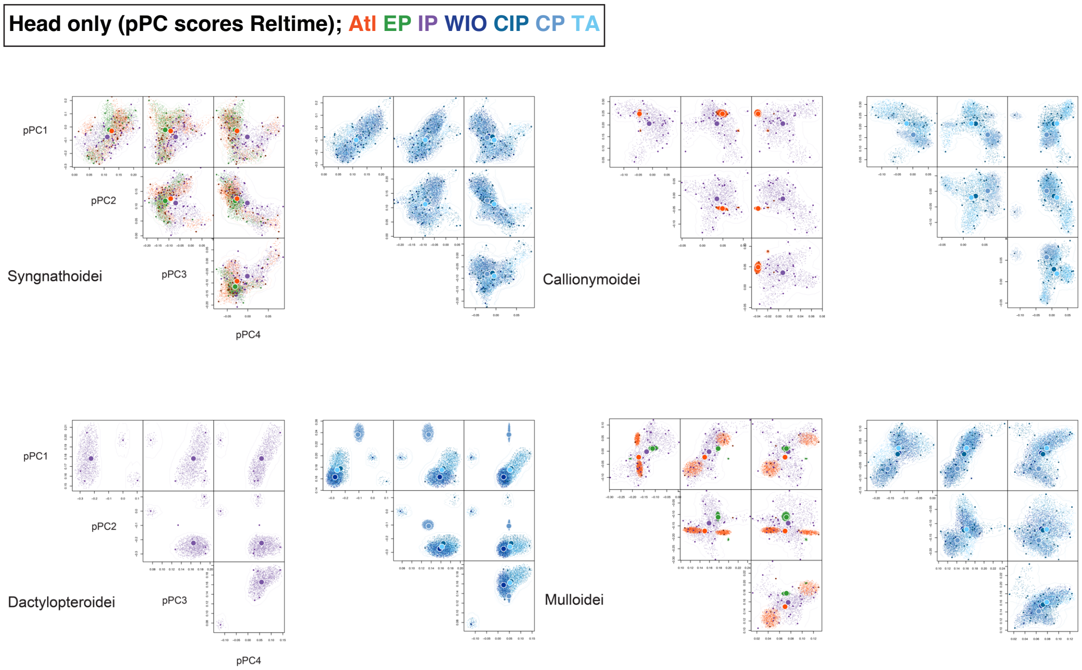


Figure S25. Hypervolumes based on pPC scores obtained with the "master tree" for syngnatharian suborders showing the overlap for head-only (four-dimensional: pPCs 1 to 4) dataset across the three major oceanic realms (Atl: Atlantic, EP: Eastern Pacific, IP: Indo-Pacific) and within the Indo-Pacific (WIO: Western Indian Ocean, CIP: Central Indo-Pacific, CP: Central Pacific, TA: Temperate Australasia). Hypervolumes cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Detailed hypervolume overlap statistics (Jaccard, Sørenson, Fraction unique 1, Fraction unique 2) can be found in Appendix 2.

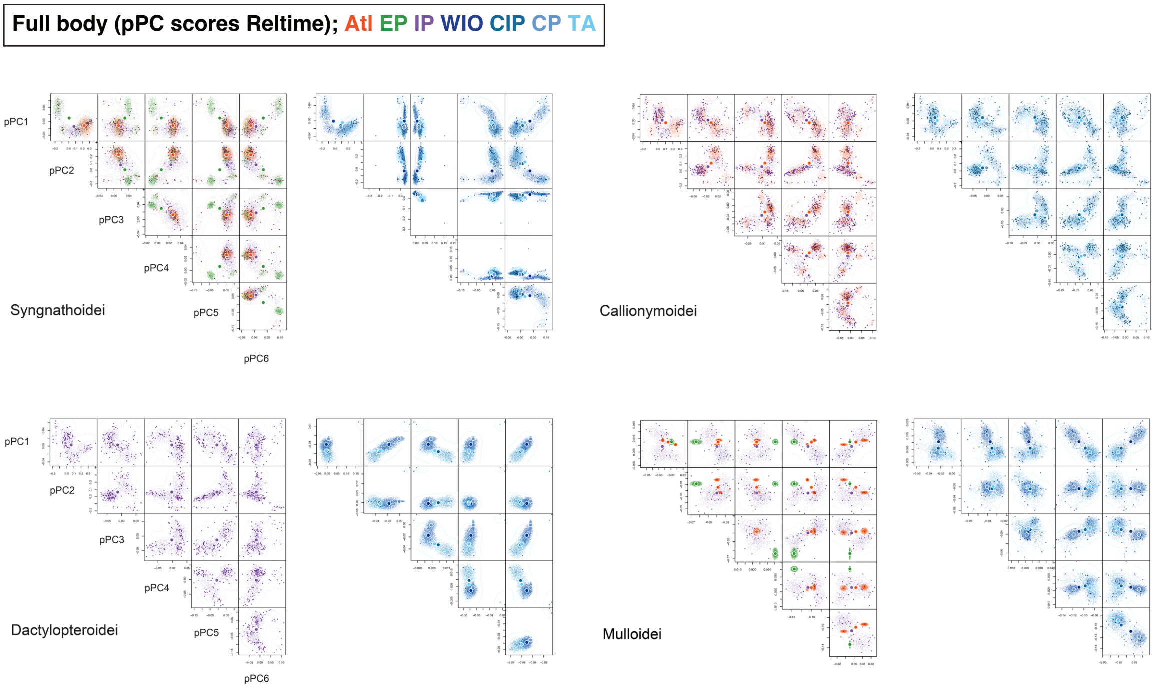


Figure S26. Hypervolumes based on pPC scores obtained with the "master tree" for syngnatharian suborders showing the overlap for full-body (six-dimensional: pPCs 1 to 6) dataset across the three major oceanic realms (Atl: Atlantic, EP: Eastern Pacific, IP: Indo-Pacific) and within the Indo-Pacific (WIO: Western Indian Ocean, CIP: Central Indo-Pacific, CP: Central Pacific, TA: Temperate Australasia). Hypervolumes cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Detailed hypervolume overlap statistics (Jaccard, Sørenson, Fraction unique 1, Fraction unique 2) can be found in Appendix 2.

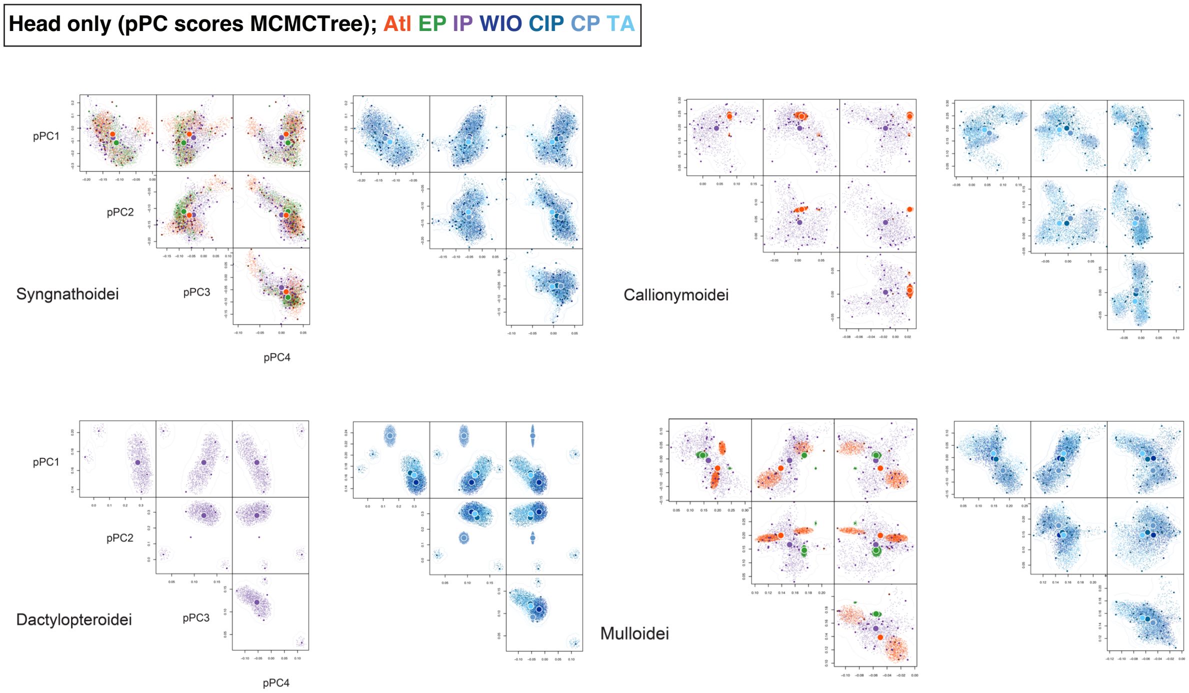


Figure S27. Hypervolumes based on pPC scores obtained with the "alternative tree" for syngnatharian suborders showing the overlap for head-only (four-dimensional: pPCs 1 to 4) dataset across the three major oceanic realms (Atl: Atlantic, EP: Eastern Pacific, IP: Indo-Pacific) and within the Indo-Pacific (WIO: Western Indian Ocean, CIP: Central Indo-Pacific, CP: Central Pacific, TA: Temperate Australasia). Hypervolumes cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Detailed hypervolume overlap statistics (Jaccard, Sørenson, Fraction unique 1, Fraction unique 2) can be found in Appendix 2.

Full body (pPC scores MCMCTree); Atl EP IP WIO CIP CP TA

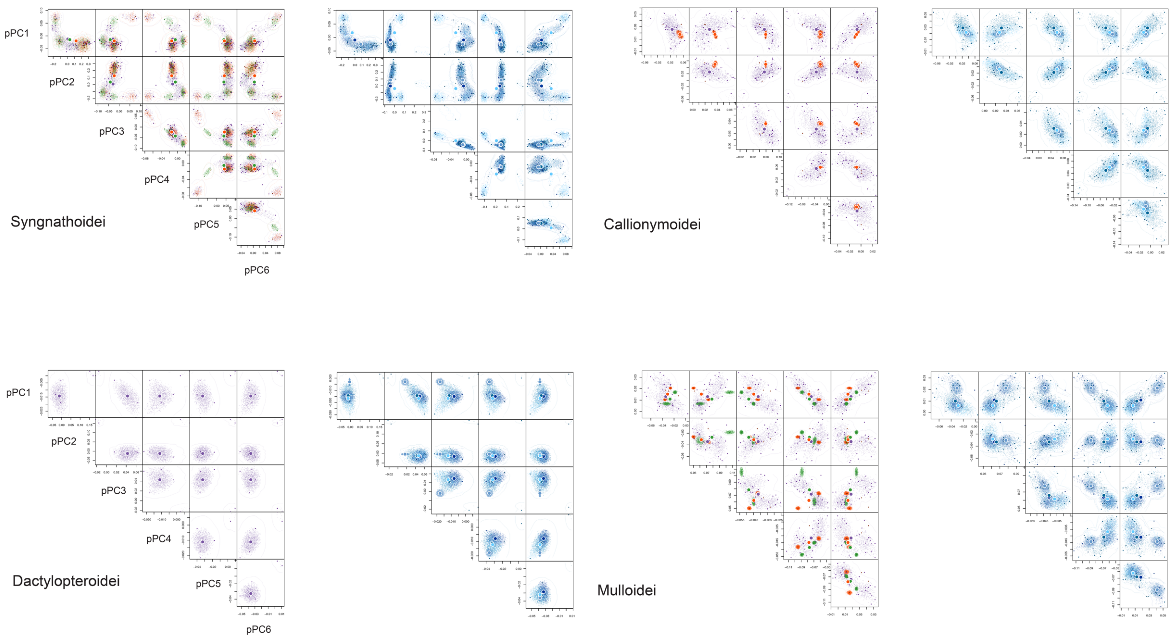


Figure S28. Hypervolumes based on pPC scores obtained with the "alternative tree" for syngnatharian suborders showing the overlap for full-body (six-dimensional: pPCs 1 to 6) dataset across the three major oceanic realms (Atl: Atlantic, EP: Eastern Pacific, IP: Indo-Pacific) and within the Indo-Pacific (WIO: Western Indian Ocean, CIP: Central Indo-Pacific, CP: Central Pacific, TA: Temperate Australasia). Hypervolumes cannot be calculated for biogeographic regions with fewer than three species (Callionymoidei: 1 species in EP and 2 species in WIO; Dactylopteroidei: 1 species in Atlantic). Detailed hypervolume overlap statistics (Jaccard, Sørenson, Fraction unique 1, Fraction unique 2) can be found in Appendix 2.

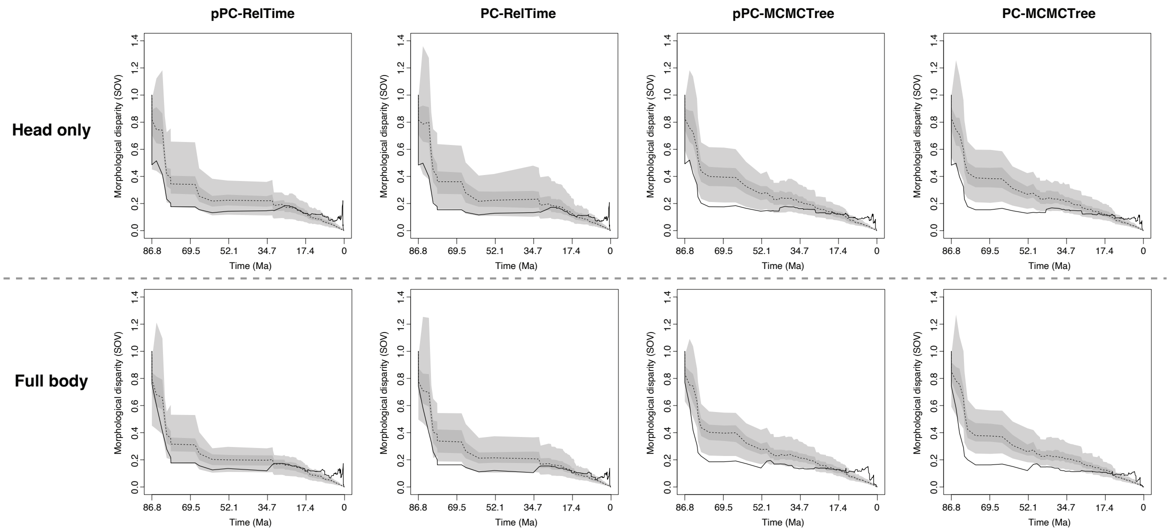


Figure S29. Disparity through time plots for syngnatharians as estimated using sum of variances (SOV) obtained from PC and pPC scores based on the "master tree" and "alternative tree" for head-only and full-body datasets.

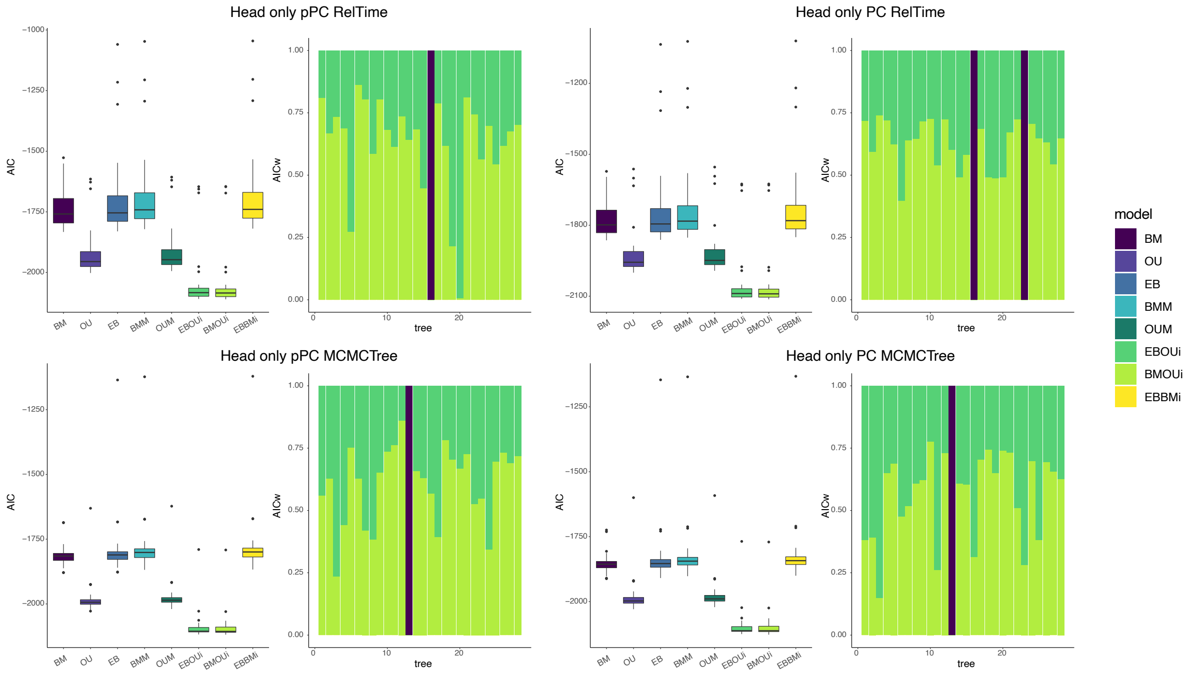


Figure S30. Model-fitting of eight morphological evolution models. The plots show the distribution of the Akaike information criterion (AIC) values and AIC weights for each tree, utilizing PC and pPC scores based on the 28 RelTime and 28 MCMCTree trees for the head-only dataset (first four axes). The models assessed are BM (single-rate Brownian motion), OU (single-regime Ornstein-Uhlenbeck), EB (early burst), BMM (multi-regime BM), OUM (multi-regime OU), EBOUi (EB followed by independent rates OU), BMOUi (BM followed by independent rates OU), and EBBMi (EB followed by independent rates BM). The best supported evolutionary model for the head-only morphology is EB or BM, followed by OU.

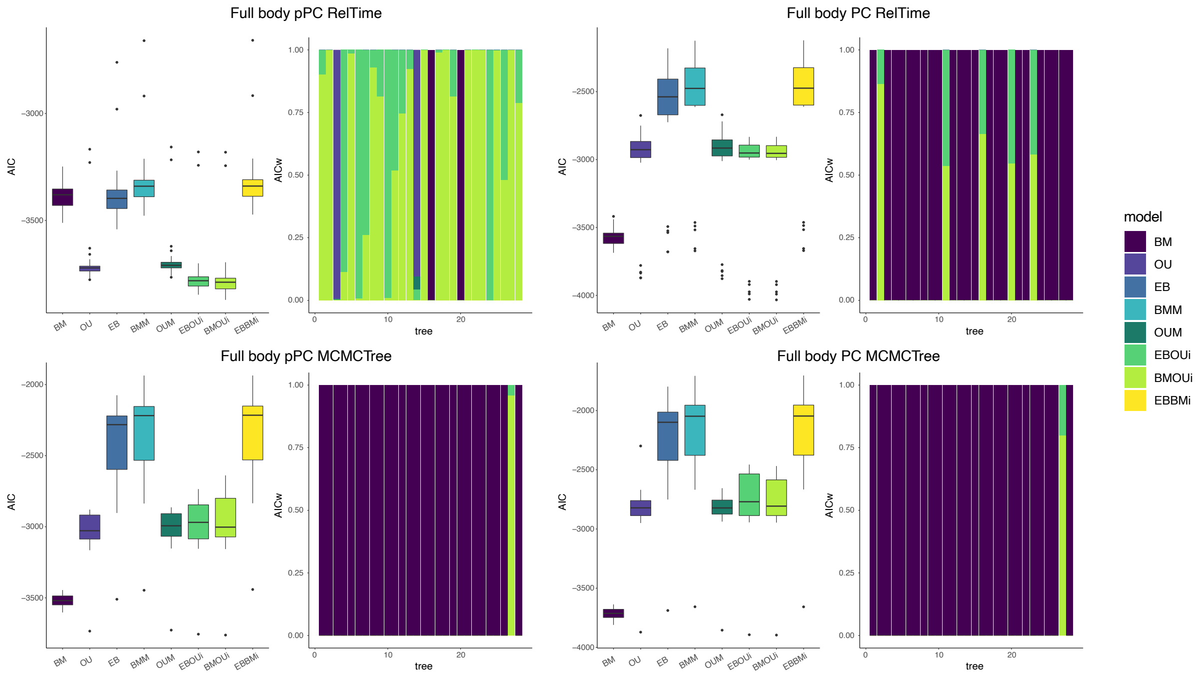


Figure S31. Model-fitting of eight morphological evolution models. The plots show the distribution of the Akaike information criterion (AIC) values and AIC weights for each tree, utilizing PC and pPC scores based on the 28 RelTime and 28 MCMCTree trees for the head-only dataset (first six axes). The models assessed are BM (single-rate Brownian motion), OU (single-regime Ornstein-Uhlenbeck), EB (early burst), BMM (multi-regime BM), OUM (multi-regime OU), EBOUi (EB followed by independent rates OU), BMOUi (BM followed by independent rates OU), and EBBMi (EB followed by independent rates BM). The best supported evolutionary model for the full body morphology is EB or BM, followed by OU when using pPC scores based on RelTime trees. However, when we used PC scores or a combination of PC and pPC scores with MCMCTree trees, BM emerged as the most supported morphological model.

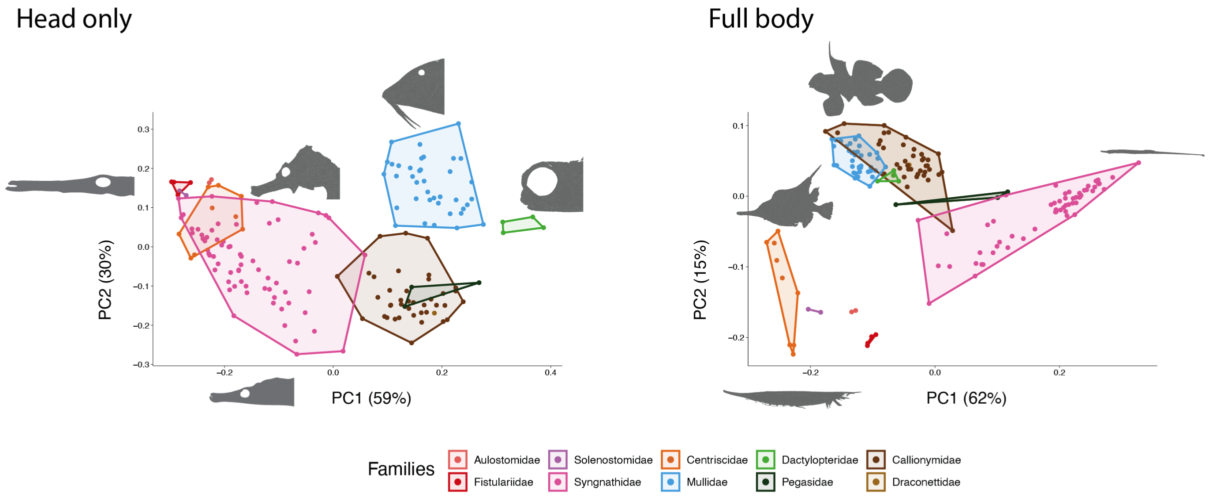


Figure S32. Morphospace of principal components 1 and 2 for head-only and full-body datasets. Colored polygons represent each syngnatharian family. The percentage of variance explained for each PC are indicated in the axes.

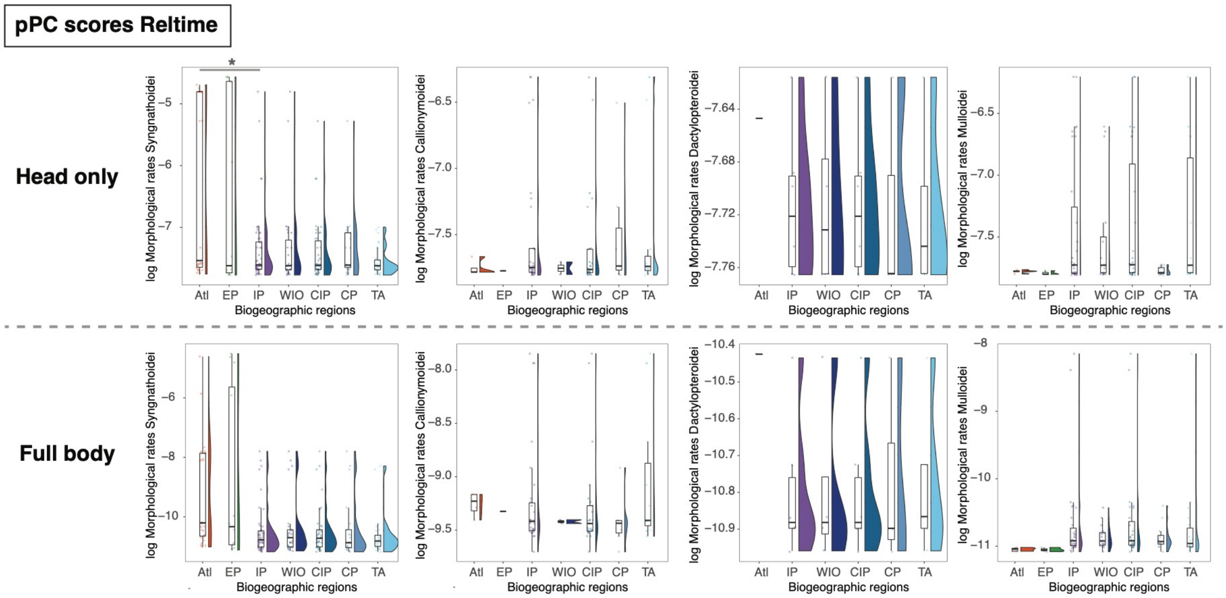


Figure S33. Average log-transformed rates of morphological evolution for syngnatharian suborders, as estimated in BAMM and depicted using raincloud plots (half-violin plots and boxplots) for head-only and full-body datasets. Average rates were estimated based on values obtained with pPC scores from 28 RelTime trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$; see Table S7 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

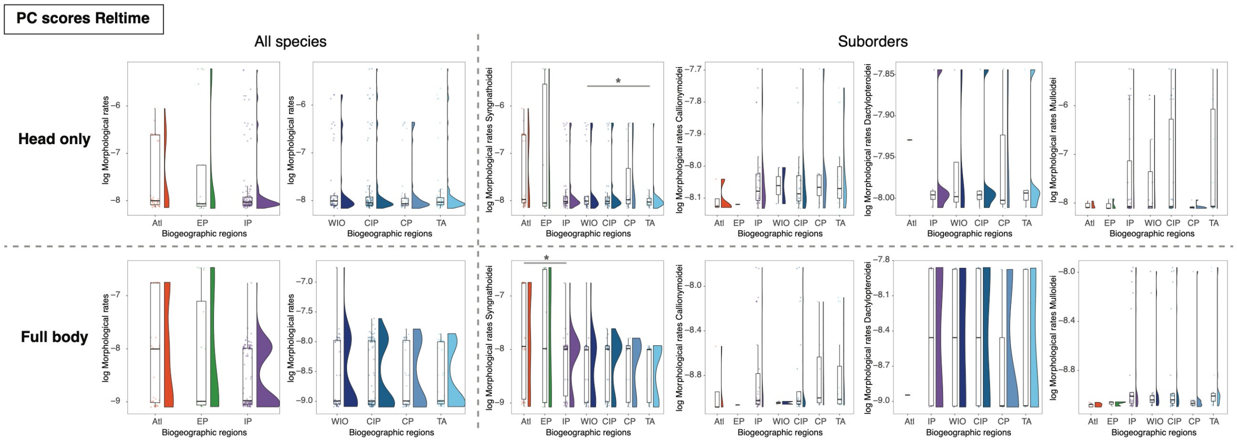


Figure S34. Average log-transformed rates of morphological evolution for syngnatharian (all species or by suborder) as estimated in BAMM and depicted using raincloud plots (half-violin plots and boxplots) for head-only and full-body datasets. Average rates were estimated based on values obtained with PC scores from 28 RelTime trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$; see Table S7 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

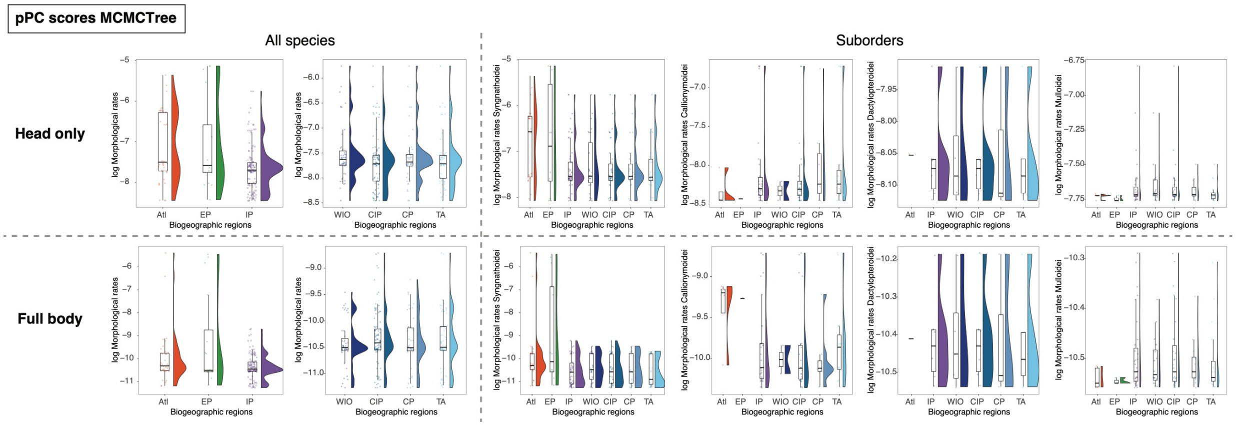


Figure S35. Average log-transformed rates of morphological evolution for syngnatharian (all species or by suborder) as estimated in BAMM and depicted using raincloud plots (half-violin plots and boxplots) for head-only and full-body datasets. Average rates were estimated based on values obtained with pPC scores from 28 MCMCTree trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Dots represent mean tip rates for each species. See Table S7 for statistics results. Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

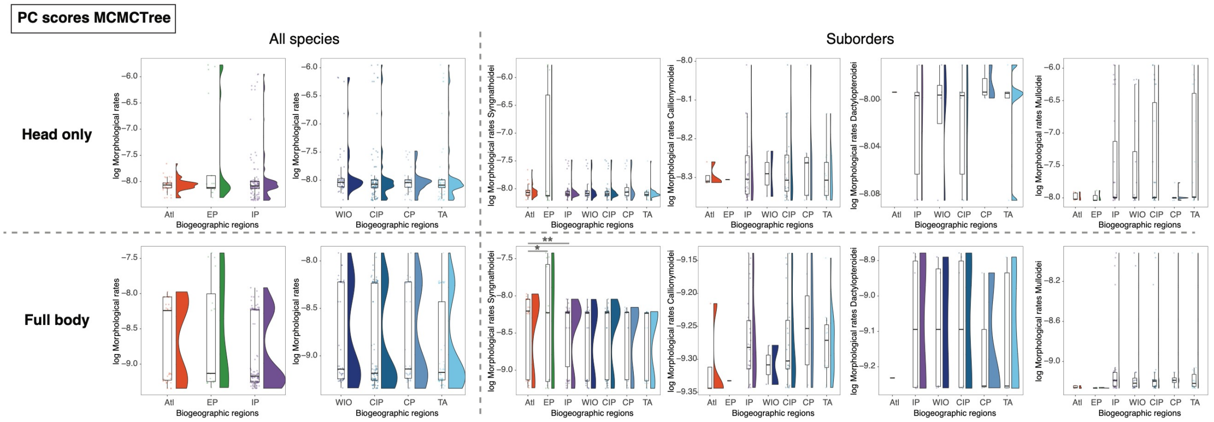


Figure S36. Average log-transformed rates of morphological evolution for syngnatharian (all species or by suborder) as estimated in BAMM and depicted using raincloud plots (half-violin plots and boxplots) for head-only and full-body datasets. Average rates were estimated based on values obtained with PC scores from 28 MCMCTree trees and are shown across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots). Dots represent mean tip rates for each species. Asterisks indicate statistical significance between regions (* $p \leq 0.05$, ** $p \leq 0.01$; see Table S7 for more details). Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

Table S7. Statistical analysis of syngnatharian rates of morphological evolution estimated in BAMM across biogeographic regions for head-only and full-body datasets. Red indicates statistical significance ($p \leq 0.05$).

RelTime pPC	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**	
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**
Head only	0.9465	0.9554	0.007648	0.09124	0.09256	0.9295	0.3173	0.8865	0.2798	0.2091
Full body	0.4503	0.9679	0.09186	0.3825	0.9294	0.5311	0.3173	0.9082	0.5103	0.7012
Areas:			**Head only**							
A:EP			1							
A:IP			0.011							
EP:IP			0.118							

RelTime PC	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**	
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**
Head only	0.9786	0.9798	0.05164	0.008639	0.6896	0.8739	0.3173	0.9453	0.1053	0.6384
Full body	0.2922	0.839	0.008494	0.07343	0.5228	0.904	1	0.8178	0.9161	0.8142
Areas:			**Head only**	**Full body**						
A:EP				1						
A:IP				0.011						
EP:IP				0.144						
CP:CIP			0.194							
TA:CIP			1							
WIO:CIP			0.113							
TA:CP			0.064							
WIO:TA			0.035							
WIO:CP			1							

MCMCTree pPC	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**	
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**
Head only	0.1892	0.8428	0.04879	0.6976	0.6584	0.6905	0.3173	0.96	0.5514	0.2022
Full body	0.3343	0.9241	0.1109	0.7176	0.4628	0.4914	0.3173	0.9775	0.5912	0.6733
Areas:			**Head only**							
A:EP			0.148							
A:IP			0.082							
EP:IP			1							

MCMCTree PC	**All species**		**Syngnathoidei**		**Callionymoidei**		**Dactylopteroidei**		**Mulloidei**	
	p-value between realms	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**	**p-value between realms**	**p-value within IP**
Head only	0.8267	0.9948	0.02673	0.08525	0.09582	0.8524	0.3173	0.8976	0.2615	0.367
Full body	0.1656	0.9718	0.0005421	0.6477	0.4004	0.9016	1	0.8151	0.9294	0.3803
Areas:			**Head only**	**Full body**						
A:EP			0.079	0.04166						
A:IP			0.052	0.00024						
EP:IP			1	1						

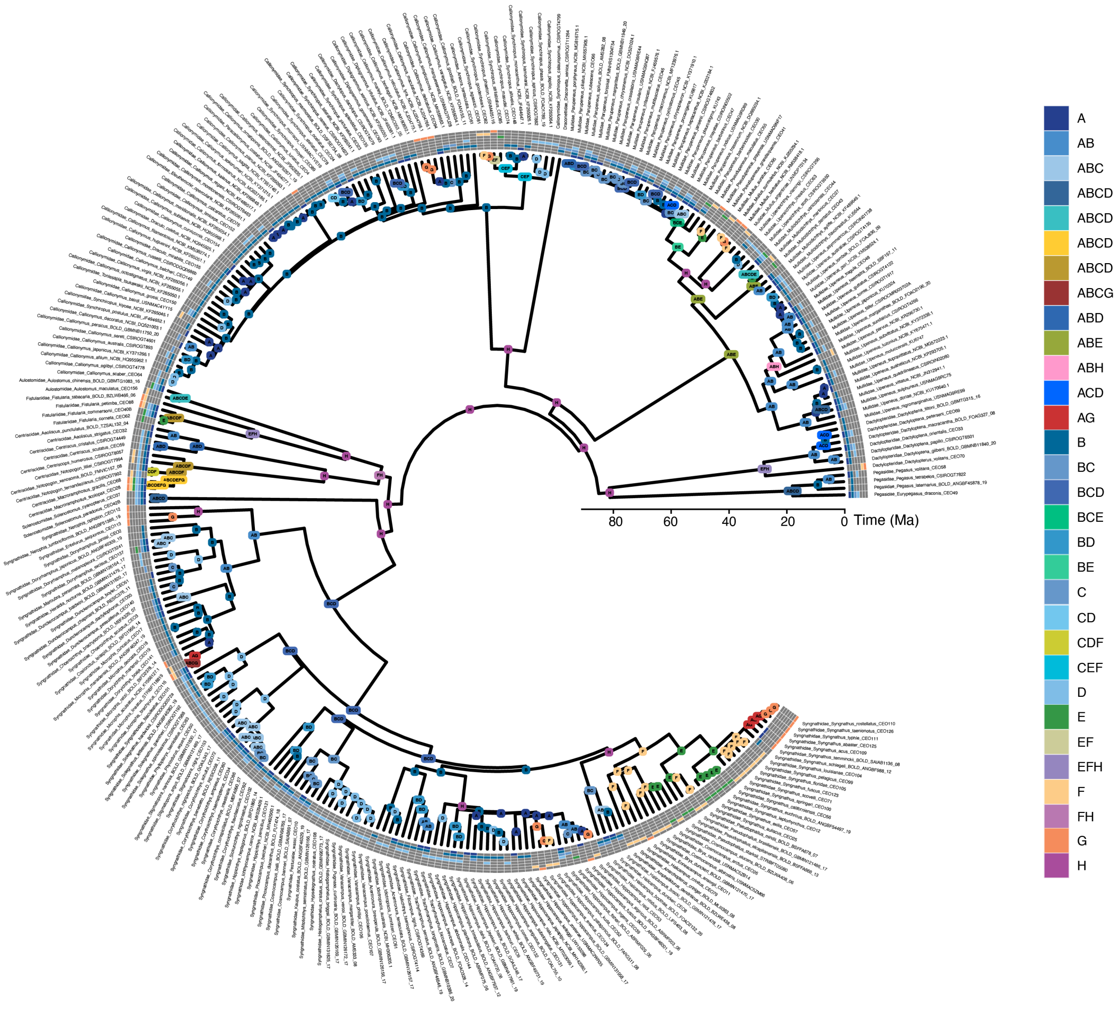


Figure S37. Ancestral range estimations for 297 syngnatharian species after removing the 26 syngnathid species with incongruent placement between this study and Stiller *et al.* (2022). This new phylogeny was ran using the best-supported biogeographic model (BAYAREA+*j*) in BioGeoBEARS. Given that removed species have mostly tipward placements, this analysis produced similar results compared to those using the complete tree with 313 species (see Fig. S3 for comparison). Boxes at each node and tip are color-coded by area, or areas with the highest ML probability. A: Western Indian Ocean; B: Central Indo-Pacific; C: Central Pacific; D: Temperate Australasia; E: Eastern Pacific; F: Western Atlantic; G: Eastern Atlantic; and H: Tethys sea. Ma: millions of years.

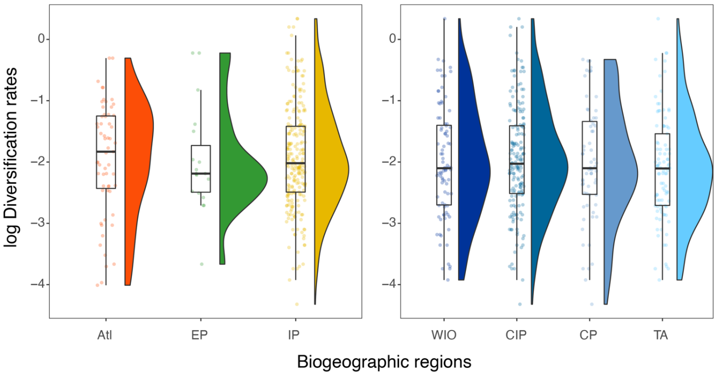


Figure S38. Average log-transformed rates of lineage diversification for all syngnatharians across the three major oceanic realms (left plots) and within Indo-Pacific subareas (right plots), as depicted using raincloud plots (half-violin plots and boxplots). The rates, which are constant across all biogeographic regions, were estimated in DR based on the 297-taxon phylogeny after removing the 26 syngnathid species with incongruent placement between this study and Stiller *et al.* (2022). Dots represent mean tip rates for each species. Atl: Atlantic; EP: Eastern Pacific; IP: Indo-Pacific; WIO: Western Indian Ocean; CIP: Central Indo-Pacific; CP: Central Pacific; and TA: Temperate Australasia.

Supplementary References

Adams C.G., Gentry A.W., Whybrow P.J. 1983. Dating the terminal Tethyan event. Utr. Micropaleontol. Bull. 30:273–298.

Adams D.C., Collyer M.L. 2018. Phylogenetic ANOVA: Group-clade aggregation, biological challenges, and a refined permutation procedure. Evolution (N. Y). 72:1204–1215.

Adams D.C., Collyer M.L., Kaliontzopoulou A., Baken E. 2021. Geomorph: Software for geometric morphometric analyses. R package version 3.3.2. See https//cran.r-project.org/package=geomorph.

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Bachmayer F. 1980. Eine fossile Schlangennadel (Syngnathidae) aus dem Leithakalk (Badenien) von St. Margarethen, Burgenland (Österreich). Ann. des Naturhistorischen Museums Wien. 83:29–33.

Bannikov A.F., Carnevale G. 2012. A long-bodied centriscoid fish from the basal Eocene of Kabardino-balkaria, northern Caucasus, Russia. Naturwissenschaften.

Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Beluzhenko E. V. 2002. Stratigrafiya sredne-verkhnemiotsenovykh i pliotsenovykh otlozheniy mezhdurech'ya Psekups–Belaya (Severo-Za- padnyj Kavkaz). Statya 1. Sredniy miotsen. Bull. Moscow So- ciety Nat. Geol. Ser. 77:47-59 (In Russian).

Betancur-R. R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17.

Blonder B., Morrow C.B., Maitner B., Harris D.J., Lamanna C., Violle C., Enquist B.J., Kerkhoff A.J. 2018. New approaches for delineating n-dimensional hypervolumes. Methods Ecol. Evol. 9:305–319.

Blot J. 1980. The ichthyofauna of the deposits of Monte Bolca (Province de Vérone, Italie). Cat. systématique présentat l'état actuel des Rech. Concern. cette faune. Bull Mus nation d'Hist nat, Paris, 4e Ser. sec. C. 2:339–396.

Bouckaert R.R. 2010. DensiTree: Making sense of sets of phylogenetic trees. Bioinformatics. 26:1372–1373.

Bray D.J., Gomon M.F. 2021. Fishes of Australia. Museums Victoria and OzFishNet. Available from http://fishesofaustralia.net.au/.

Caetano D.S., O'Meara B.C., Beaulieu J.M. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution (N. Y). 72:2308–2324.

Caldwell M.K. 1962. Development and distribution of larval and ju- venile fishes of the family Mullidae of the western North Atlantic. Fish. Bull. 62:403–457.

Calzoni P., Amalfitano J., Giusberti L., Marramà G., Carnevale G. 2023. Eocene Rhamphosidae (Teleostei: Syngnathiformes) from the Bolca Lagerstätte, Italy. Riv. It. Paleont. Strat. 129:573–607.

Carnevale G., Bannikov A.F. 2019. A dragonet (Teleostei, callionymoidei) from the eocene of monte bolca, italy. Boll. della Soc. Paleontol. Ital. 58:295–307.

Carnevale G., Bannikov A.F., Landini W., Sorbini C. 2006. Volhynian (early Sarmatian sensu lato) fishes from Tsurevsky, North Caucasus, Russia. 80:684–699.

Clavel J., Escarguel G., Merceron G. 2015. mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. Methods Ecol. Evol.

Claverie T., Wainwright P.C. 2014. A morphospace for reef fishes: Elongation is the dominant axis of body shape evolution. PLoS One. 9.

Coates A.G., Obando J.A. 1996. The geologic evolution of the Central American Isthmus. Evolution and Environment in Tropical America. .

Ding W.N., Ree R.H., Spicer R.A., Xing Y.W. 2020. Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora. Science (80-.). 369:578–581.

Dupin J., Matzke N.J., Särkinen T., Knapp S., Olmstead R.G., Bohs L., Smith S.D. 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. J. Biogeogr. 44:887–899.

Eastman J.M., Harmon L.J., Tank D.C. 2013. Congruification: Support for time scaling large phylogenetic trees. Methods Ecol. Evol. 4:688–691.

Felsenstein J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Friedman M., Carnevale G. 2018. The Bolca Lagerstätten: shallow marine life in the Eocene. J. Geol. Soc. London.

Friedman M., Feilich K.L., Beckett H.T., Alfaro M.E., Faircloth B.C., Černý D., Miya M., Near T.J., Harrington R.C. 2019. A phylogenomic framework for pelagiarian fishes (Acanthomorpha: Percomorpha) highlights mosaic radiation in the open ocean. Proc. R. Soc. B Biol. Sci.

Froese R., Pauly D. 2021. FishBase. Available from www.fishbase.org.

Garland T., Ives A.R. 2000. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

Gavrilov Y., Shcherbinina E., Oberhänsli H. 2003. Paleocene- Eocene boundary events in the northeastern Peri-Tethys. Geol. Soc. Am. Spec. Pap. 369:147–168.

GBIF. 2022. Global Biodiversity Information Facility. Available from https://www.gbif.org/.

Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Syst. Biol. 60:451–465.

Guillerme T. 2018. dispRity: A modular R package for measuring disparity. Methods Ecol. Evol. 9:1755–1763.

Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2008. GEIGER: Investigating evolutionary radiations. Bioinformatics. 24:129–131.

Harrington R.C., Faircloth B.C., Eytan R.I., Smith W.L., Near T.J., Alfaro M.E., Friedman M. 2016. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. BMC Evol. Biol. 16:1–14.

Hedman M.M. 2010. Constraints on clade ages from fossil outgroups. Paleobiology. 36:16–31.

IUCN. 2021. The IUCN Red List of Threatened Species. Available from https://www.iucnredlist.org.

Jetz W., Thomas G.H., Joy J.B., Hartmann K., Mooers A.O. 2012. The global diversity of birds in space and time. Nature. 491:444–448.

Jombart T., Kendall M., Almagro-Garcia J., Colijn C. 2017. treespace: Statistical exploration of landscapes of phylogenetic trees. Mol. Ecol. Resour. 17:1385–1392.

Jordan D.S., Evermann B.W. 1898. Fishes NORTH AND MIDDLE AMERICA. Bull. United States Natl. Museum.:1241–2183.

Klaus K. V., Matzke N.J. 2020. Statistical comparison of trait-dependent biogeographical models indicates that Podocarpaceae dispersal is influenced by both seed cone traits and geographical distance. Syst. Biol. 69:61–75.

Kulbicki M., Parravicini V., Bellwood D.R., Arias-Gonzàlez E., Chabanet P., Floeter S.R., Friedlander A., McPherson J., Myers R.E., Vigliola L., Mouillot D. 2013. Global biogeography of reef fishes: A hierarchical quantitative delineation of regions. PLoS One. 8.

Landis M.J., Matzke N.J., Moore B.R., Huelsenbeck J.P. 2013. Bayesian analysis of biogeography when the number of areas is large. Syst. Biol. 62:789–804.

Lloyd G.T., Bapst D.W., Friedman M., Davis K.E. 2016. Probabilistic divergence time estimation without branch lengths: Dating the origins of dinosaurs, avian flight and crown birds. Biol. Lett.

Longo S.J., Faircloth B.C., Meyer A., Westneat M.W., Alfaro M.E., Wainwright P.C. 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Mol. Phylogenet. Evol. 113:33–48.

Matzke N.J. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. R Packag. version 0.2.

Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

Matzke N.J. 2019. BioGeoBEARS - Run BioGeoBEARS on multiple trees. Available from https://github.com/nmatzke/BioGeoBEARS/blob/master/R/BioGeoBEARS_on_multiple_trees_v1.R.

Matzke N.J. 2022. Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. J. Biogeogr. 49:1805–1824.

Mello B., Tao Q., Barba-Montoya J., Kumar S. 2021. Molecular dating for phylogenies containing a mix of populations and species by using Bayesian and RelTime approaches. Mol. Ecol. Resour. 21:122–136.

Miller E.C., Martinez C.M., Friedmand S.T., Wainwright P.C., Price S.A., Luke Tornabene. 2022. Alternating regimes of shallow and deep-sea diversification explain a species-richness paradox in marine fishes. Proc. Natl. Acad. Sci.:1–12.

Mirarab S., Warnow T. 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics.

Montes C., Cardona A., Jaramillo C., Pardo A., Silva C., Valencia V., Ayala C., Pérez-Angel L., Rodriguez-Parra L., Ramirez V., Niño H. 2015. Middle Miocene closure of the Central American Seaway. Science (80-.). 348:226–229.

Nielsen E. 1960. A new Eocene teleost from Denmark. Meddelelser fra Dansk Geol. Foren. 14:247–252.

O'Dea, Aaron, Lessios H.A., Coates A.G., Eytan R.I., Restrepo-Moreno S.A., Cione A.L., Collins L.S., de Queiroz A., Farris D.W., Norris R.D., Stallard R.F., Woodburne M.O., Aguilera O., Aubry M.-P., Berggren W.A., Budd A.F., Cozzuol M.A., Coppard S.E., Duque-Caro H., Finnegan S., Gasparini G.M., Grossman E.L., Johnson K.G., Keigwin L.D., Knowlton N., Leigh E.G., Leonard-Pingel J.S., Marko P.B., Pyenson N.D., Rachello-Dolmen P.G., Soibelzon E., Soibelzon L., Todd J.A., Vermeij G.J., Jackson J.B.C. 2016. Formation of the Isthmus of Panama. Sci. Adv. 2:1–12.

OBIS. 2021. Data from the Ocean Biogeographic Information System. Intergovernmental Oceanographic Commission of UNESCO. .

Papazzoni C.A., Carnevale G., Fornaciari E., Giusberti L., Trevisani E. 2014. The Pesciara-Monte Postale Fossil- Lagerstätte: 1. Biostratigraphy, sedimentology and depositional model. In Papazzoni C.A., Giusberti L., Carnevale G., Roghi G., Bassi D. & Zorzin R. (eds), The Bolca Fossil-Lagerstätten: A window into the Eocene World. Rend. della Soc. Paleontol. Ital. 4:29–36.

Parham J.F., Donoghue P.C.J., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver J.E., Van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L., Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J. 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61:346–359.

Parin N., Micklich N. 1996. Fossil gasterosteiformes from the lower oligocene of Frauenweiler (Baden-Württemberg, Germany) I. New information on the morphology and systematics of the genus Aeoliscus Jordan &Starks 1902. Paläontologische Zeitschrift. 70:521–545.

Peterson R.D., Sullivan J.P., Hopkins C.D., Santaquiteria A., Dillman C.B., Pirro S., Betancur-R R., Arcila D., Hughes L.C., Ortí G. 2022. Phylogenomics of Bony-Tongue Fishes (Osteoglossomorpha) Shed Light on the Craniofacial Evolution and Biogeography of the Weakly Electric Clade (Mormyridae). Syst. Biol. 71:1032–1044.

Pietsch T. 1978. Evolutionary Relationships of the Sea Moths (Teleostei : Pegasidae) with a Classification of Gasterosteiform Families Author (s): Theodore W . Pietsch Published by : American Society of Ichthyologists and Herpetologists (ASIH) Stable URL : https://w. Copeia.:517–529.

Plummer M., Best N., Cowles K., Vines K. 2006. {CODA}: Convergence Diagnosis and Output Analysis for {MCMC}. R News. 6:7–11.

Pohlert T. 2021. Package ‘PMCMRplus’. R Package Version 4.1. Available online: https://cran.r-project.org/web/packages/PMCMRplus/index.html (accessed on 16 October 2022). .

Popov Y.A. 2017. First record of the pipefish Nerophis zapfei Bachmayer (Syngnathidae, Gasterosteiformes) from the Middle Miocene of Northern Moldova. Paleontol. J. 51:533–541.

Price M.N., Dehal P.S., Arkin A.P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One.

QGIS Development Team. 2009. QGIS Geographic Information System. .

Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One. 9.

Rabosky D.L. 2016. Reproductive isolation and the causes of speciation rate variation in nature. Biol. J. Linn. Soc. 118:13–25.

Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.

Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5:701–707.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol.

Ranwez V., Douzery E.J.P., Cambon C., Chantret N., Delsuc F. 2018. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol.

Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45:741–749.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

dos Reis M., Yang Z. 2019. Bayesian molecular clock dating using genome-scale datasets. Methods in Molecular Biology. .

Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol.

Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero A., Walker Jr. H.J., Carpenter K.E., Sheaves M., Ortí G., Arcila D., Betancur-R. R. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. Proc. Natl. Acad. Sci. U. S. A.

Rögl F. 1999. Palaeogeographic considerations for Mediterranean and Paratethys seaways. Ann. Naturhist. Mus. Wien. 99A:279–310.

Ronquist F. 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. Syst. Biol. 46:195–203.

Santaquiteria A., Siqueira A.C., Duarte-Ribeiro E., Carnevale G., White W.T., Pogonoski J.J., Baldwin C.C., Ortí G., Arcila D., Ricardo B.R. 2021. Phylogenomics and Historical Biogeography of Seahorses, Dragonets, Goatfishes, and Allies (Teleostei: Syngnatharia): Assessing Factors Driving Uncertainty in Biogeographic Inferences. Syst. Biol. 70:1145–1162.

Schmid H.P., Harzhauser M., Kroh A. 2001. Hypoxic Events on a Middle Miocene Carbonate Platform of the Central Paratethys (Austria, Badenian, 14 Ma), with contribution by Coric S., Rögl F. & Schultz O. Ann. des Naturhistorisches Museum Wien. 95A:127–177.

Schmitz B., Peucker-Ehrenbrink B., Heilmann-Clausen C., Åberg G., Asaro F., Lee C.T.A. 2004. Basaltic explosive volcanism, but no comet impact, at the Paleocene-Eocene boundary: High-resolution chemical and isotopic records from Egypt, Spain and Denmark. Earth Planet. Sci. Lett. 225:1–17.

Smith S.A., O'Meara B.C. 2012. TreePL: Divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 28:2689–2690.

Spalding M.D., Fox H.E., Allen G.R., Davidson N., Ferdaña Z.A., Finlayson M., Halpern B.S., Jorge M.A., Lombana A., Lourie S.A., Martin K.D., McManus E., Molnar J., Recchia C.A., Robertson J. 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. Bioscience. 57:573–583.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.

Stecher G., Tamura K., Kumar S. 2020. Molecular evolutionary genetics analysis (MEGA) for macOS. Mol. Biol. Evol. 37:1237–1239.

Steininger F., Rögl F. 1979. The paratethys history. A contribution towards the Neogene geodynamics of the alpine orogene. Ann. Geol. des Pays Hell. 3:1153–1165.

Stiller J., Short G., Hamilton H., Saarman N., Longo S., Wainwright P., Rouse G.W., Simison W.B. 2022. Phylogenomic analysis of Syngnathidae reveals novel relationships, origins of endemic diversity and variable diversification rates. BMC Biol. 20:1–21.

Storey M., Duncan R.A., Swisher C.C. 2007. Paleocene-Eocene thermal maximum and the opening of the northeast Atlantic. Science (80-.). 316:587–589.

Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kumar S. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. U. S. A. 109:19333–19338.

Tamura K., Tao Q., Kumar S. 2018. Theoretical foundation of the reltime method for estimating divergence times from variable evolutionary rates. Mol. Biol. Evol. 35:1770–1782.

Uyeda J.C., Caetano D.S., Pennell M.W. 2015. Comparative Analysis of Principal Components Can be Misleading. Syst. Biol. 64:677–689.

Volta G. 1796. Ittiolit. Veronese del Mus. Bozziano. 2.

Wickham H. 2008. Elegant Graphics for Data Analysis: ggplot2. .

Xing Y., Ree R.H. 2017. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. Proc. Natl. Acad. Sci. U. S. A. 114:E3444–E3451.

Yang C., Zheng Y., Tan S., Meng G., Rao W., Yang C., Bourne D.G., O'Brien P.A., Xu J., Liao S., Chen A., Chen X., Jia X., Zhang A. bing, Liu S. 2020. Efficient COI barcoding using high throughput single-end 400 bp sequencing. BMC Genomics. 21:1–10.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics.

Appendix C

Supplementary Material for Evolution of planktivory in extant and fossil acanthuriforms and the genomic basis of this trophic transition

Supplementary Materials and Methods

DNA extractions, exon capture, sequencing, and assembly

We generated new exon capture data from tissue samples extracted from museum voucher specimens for a total of 57 acanthuriforms (*Acanthurus tractus* duplicated) and 9 outgroups from closely related families: Chaetodontidae, 3; Pomacanthidae, 2; Caproidae, 1; Ephippidae, 1; and Scatophagidae, 2. DNA from tissue samples was extracted in a 96-well plate format on a GenePrep, following manufacturer's instructions at the Laboratory of Analytical Biology at the Smithsonian National Museum of Natural History. High quality DNA extractions were sent to Arbor Biosciences for library preparation and exon capture to target the 1,105 single-copy exons developed for the FishLife project (Hughes et al. 2018) using the Eupercaria-specific probe set (Hughes et al. 2020), that also includes PCR-based 29 legacy markers (mtDNA and nuclear genes) commonly used for fish phylogenetics (e.g., Li et al. 2007; Betancur-R et al. 2013; Broughton et al. 2013). Enriched libraries were sequenced using one lane of the Illumina HiSeq 4000 platform with paired-end 100bp at the University of Chicago Genomics facility. The raw sequence data was assembled and aligned using the bioinformatic pipeline developed by Hughes et al. (2020; available at https://github.com/lilychughes/FishLifeExonCapture/). The final step of the pipeline generate alignments in their correct reading frames for each exon using MACSE v. 2.03 (Ranwez et al. 2018).

Quality control

For each sample we selected the COI and CYTB genes and blasted them against the Barcode of Life Database (BOLD) and National Center for Biotechnology Information (NCBI) repositories for verification of species identifications using "bold_identification" python script (Yang et al. 2020). We also visually inspected all the alignments to adjust the reading frames, remove poor-quality reads and correct misaligned sections in Geneious Prime v. 2021.2 (Kearse et al. 2012). 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* (Borowiec 2016). After these quality control steps, we removed a total of 9 markers with low quality, as well as 98 markers with more than 50% missing data (present in fewer than 32 species). This resulted in a reduced molecular matrix consisting of 998 genes for 56 ingroup species (including a duplicated *A. tractus*) out of the 86 extant acanthuriforms (~65.1%), representing all genera and families, along with 9 outgroup species as outlined above.

Taxonomic sampling augmentation

To increase the number of species, we generated an expanded matrix combining the 91 newly sequenced species (reduced matrix) with sequences for up to 29 markers obtained from GenBank. We first individually aligned each legacy marker from GenBank using MACSE. We then aligned them with their corresponding legacy marker sequenced on our reduced matrix. We manually checked each gene to ensure it was in the correct reading frame and retained 22 out of the 29 markers, with 7 being common between both datasets (Table S1). We concatenated the legacy markers with the reduced matrix, placing the mitochondrial (mtDNA) markers at the end, resulting in a concatenated matrix of 1002 markers for 148 individuals. We conducted an initial assessments of phylogenetic relationships in FastTree-2 (Price et al. 2010) to identify possible cases of contamination and misidentification. We removed a total of 58 duplicates. After all phylogenomic inference analyses were conducted, we renamed *Prionurus punctatus* to *Prionurus laticlavius* as it is recognized as a junior subjective synonym of the latter (Ludt et al. 2019). Therefore, we removed the duplicated *P. laticlavius* specimen with only legacy markers. Furthermore, the placement of *Acanthurus tristis* in the phylogenetic trees could not be determined due to incomplete data, including only COI marker with a sequence length of 202 bp. Consequently, we excluded *A. tristis* from all trees for downstream comparative analyses. The final expanded molecular matrix comprises of 1002 genes for 80 ingroup species (~93%) and 9 outgroup species (65 spp. with FishLife exons and 25 spp. with legacy markers).

Table S1. Legacy markers and their corresponding exon IDs that are present in the FishLife dataset and/or in the sequences downloaded from GenBank.

Legacy marker	Exon ID	FishLife	GenBank
TBR1	E1541	YES	NO
RAG1	E1684	YES	YES
KIAA1239	E1728	YES	NO
MYH6	E1730	YES	YES
ENC1	E1732	YES	NO
PLAGL2	E1735	YES	YES
RIPK4	E1737	YES	NO
SH3PX3	E1738	YES	NO
SIDKEY	E1739	YES	NO
SREB2	E1740	YES	NO
ZIC1	E1741	YES	YES
SVEP1	E1746	YES	NO
GPR61	E1747	YES	NO
IRBP	E1748	YES	NO
RNF213	E1749	YES	NO
RHOD	E1750	YES	YES
UBEA3	E1752	YES	NO
UBEA3like (modified)	E1753	YES	NO
COI	Mitochondrial DNA	YES	YES
CYTB	Mitochondrial DNA	YES	YES
12S	Mitochondrial DNA	NO	YES
16S	Mitochondrial DNA	NO	YES

Phylogenomic inference

For each assembled molecular matrix, reduced and expanded, we inferred maximum likelihood (ML) trees and multispecies coalescent species trees. First, we determined the best-fitting partition scheme for each matrix using PartitionFinder2 (Lanfear et al. 2017) based on a priori by-codon partitions for each protein-coding marker, and two partitions for each of the ribosomal markers (12S and 16S). We estimated concatenation-based ML trees in RAxML v. 8.2.11 (Stamatakis 2014) using the best-fit partitioning schemes selected via the Bayesian Information Criterion (BIC) and the GTRGAMMA model. Using the *raxml-ng* (extension of RAxML for supercomputers; Kozlov et al. 2019) we ran 30 independent ML searches and used 100 nonparametric bootstrapping to assess edge support. To infer species trees while accounting for incomplete lineage sorting (ILS), we initially estimated individual gene trees in RAxML using by-codon partitions. All mtDNA markers were grouped into a single locus alignment, with by-codon partitions applied specifically for protein-coding genes and two partitions for 12S and 16S, as explained above. After inferring best trees from multiple runs and bootstrap support (BS) values, we collapsed gene tree branches with low BS (<33%). We then conducted multispecies coalescent species-tree analyses with multi-locus bootstrapping in ASTRAL-III (Zhang et al. 2017) using collapsed gene trees as input to generate a species tree for each matrix. We also assessed gene concordance factors (Minh et al. 2020) by calculating the percentage of gene trees in the data matrix that support a specific branch in the concatenation-based (RAxML) and multispecies coalescent-based (ASTRAL-III) species trees inferred for both datasets (Minh et al. 2020).

Integration of fossils and extant species

We newly coded a morphological matrix consisting of 107 characters for 32 fossil and 19 extant acanthuriform species plus 5 extant outgroups (see below for the list of osteological characters and character states). To assess the phylogenetic placement of each species based on morphology, we inferred trees based on parsimony and ML approaches. We estimated the parsimony tree in TNT v. 1.5 (Goloboff and Catalano 2016) using a driven-search strategy (sectorial ratchet, tree-fusing methodologies) with default parameters. The ML morphological tree was estimated using the MULTIGAMMA and Mk models with 30 iterations in RAxML. We combined the morphological and the expanded molecular matrices for a total of 112 fossil and extant ingroup species and nine outgroups. We estimated the combined matrix in RAxML using the MULTIGAMMA and Mk models, 100 bootstraps and six partitions: five for the molecular sequences (one for each codon position of all nuclear and mtDNA protein-coding markers, plus two for 12S and 16S) and one for the morphological dataset. Taxa with polymorphic character states were coded as missing ("?") for RAxML, which cannot handle polymorphic characters.

Morphological characters

List of juvenile and adult (but not larval) 107 osteological characters (synapomorphies and autapomorphies) and character states used for distinguishing generic taxa in the following literature on acanthuroids and immediate outgroups: Randall (1955); Smith (1966); Tyler (1970), (1997); Mok (1977); Tyler et al. (1989); Guiasu and Winterbottom (1993); Winterbottom (1993); Winterbottom and McLennan (1993); Bannikov and Tyler (1995); Tyler and Sorbini (1998); Tyler and Bannikov (2000); Tyler and Micklich (2011), and personal observations by J.C.Tyler.

Skull

1) Skull bone surfaces. Relatively smooth to variously striated or ridged but not cancellous=0; cancellous, especially the frontal and supraoccipital, with pitted indentations and/or spongy-like deep sculpturing, connective tissue or fatty filled=1.

2) Ascending process of premaxilla, length. Moderate, about one-third to two-thirds length of alveolar process, whether to front edge of block-like ethmoid or to level of ca. one-third to one-half back on more elongate ethmoid=0; short, less than one-third length of alveolar process, not reaching behind anterior end of ethmoid=1; long, usually reaching more than one-half back on ethmoid=2.

3) Maxilla-premaxilla articulation. Moveable=0; slightly moveable=1; relatively immoveable=2.

4) Jaw protrusibility. Protrusibility well developed=0; only slightly protrusible=1; essentially no protrusibility, rotating around ethmoid (presumed for fossils based on premaxilla and maxilla conditions)=2.

5) Articular, size relative to dentary. Articular distinctly longer and larger than dentary=0; articular about equal in length and size to dentary=1; articular distinctly smaller than dentary=2.

6) Articular/dentary articulation. Broad anterior surface of articular attached firmly to dentary=0; area of contact between articular and dentary much reduced and joint flexible=1.

7) Palatine, position, and articulation. At least posterior portion of palatine alongside or immediately in front of lateral ethmoid and articulated to it, usually by a condylar process=0; palatine lies well forward of lateral ethmoid and no articulation with it=1.

8) Mesethmoid, position. Mostly situated anterior to lateral ethmoids, often with a convex and usually wedge-shaped surface anteriorly along which rostral cartilage slides during protrusion of upper jaw=0; not extended forward of lateral ethmoids, and inverted posteriorly with lateral walls that extend back into orbit to accommodate ascending processes of premaxillae=1.

9) Frontal, spiny process. Absent=0; a conical spine-like protuberance or prominent bump present on each frontal near anterodorsal margin of orbit=1.

10) Ectopterygoid shape. Elongate and extending ventrally at least along about one-half of anterior edge of quadrate=0; greatly reduced in size, its ventral region extending only a short distance along anterior edge of quadrate=1.

11) Symplectic length. About 25% or more of length of ventrolateral margin of quadrate=0; reduced in length, about 8% to 15% of quadrate length=1; symplectic absent=2.

12) Symplectic shape. A simple straight cyclindrical bone=0; with a gentle sigmoid flexure=1; not applicable= –.

13) Parietal. Present=0; absent=1.

14) Parasphenoid apophysis. Absent=0; present as a bony process or spur=1; present as a cartilage without bony process=2.

15) Preopercular margin. Serrate=0; smooth=1.

16) Infraorbitals, number. Six=0; seven=1; five=2; four=3; three=4; two=5.

17) Second infraorbital articulation with lachrymal. A close association between lachrymal and second infraorbital, and main body of lachrymal lies below the path of infraorbital ring=0; second infraorbital articulates loosely and less closely with lachrymal, often out of contact with one another, and point of connection is at posteroventral corner of lachrymal, and main body of lachrymal lies above the projected path of infraorbital ring=1; not applicable when second infraorbital absent.

18) Infraorbital series orientation and position, and lachrymal articulation. Lachrymal with a condylar articulation with lateral ethmoid, and its posterodorsal margin conforms to anteroventral border of orbit=0; infraorbital series turns anteriorly below lateral ethmoid and extends forward alongside of snout, with lachrymal displaced anteriorly and removed from anterior border of orbit, and not articulating with lateral ethmoid=1.

19) Suborbital shelf. Present=0; absent=1.

20) Supraoccipital crest, height when present. High, often forming triangular peak with thickened anterior edge=0; moderate height=1; absent or very low, often as short ridge=2.

21) Supraoccipital association with exoccipitals ventrally. Ventral extension of supraoccipital (spina occipitalis) embraced laterally by dorsal extensions of exoccipitals that are tightly attached to supraoccipital on each side of the extension=0; supraoccipital does not contact exoccipitals ventrally=1.

22) Supraoccipital, spina occipitalis. Well developed, extending ventrally between epiotics to dorsal margin of foramen magnum and embraced laterally by dorsal processes of exoccipitals=0; no spina occipitalis and epiotics meet synchondrally, broadly separating supraoccipital from exoccipitals and foramen magnum=1.

23) Hyomandibula, anterodorsal surface. Relatively flat and smooth=0; a distinct transverse ridge present=1.

24) Hyomandibula, dorsomedial flange, if present. Without a foramen=0; with a large foramen=1; not applicable= –.

25) Opercle, presence of dilator process at dorsal end. Poorly developed, and posterior edge of upper region of opercle slightly to moderately concave to moderately or distinctly convex=0; well developed as a tapering dorsal projection, and posterior edge of upper region of opercle only slightly convex to relatively straight=1.

26) Interopercle, shape. Approximately ovoid=0; broad posteriorly with dorsal portion continuing anteriorly as a narrow extension=1, or as a narrow ligamentous band =2.

Teeth

27) Teeth, articulation. Fixed, relatively inflexible=0; slightly bendable, flexible, often setiform, brush-like in multiple rows=1; especially bendable, flexible=2; teeth absent=3.

28) Teeth, shape. Relatively conical and of moderate length, with smooth edges at least in adults, and without prominent lobes, denticulate edges, or notches=0; compressed and spatulate through

most of their length, with well-developed denticulations/lobations=1; mostly conical to somewhat compressed at least distally, with denticulations variously small, minute or absent=2; stout, compressed, with one or two prominent notches=3; slender, elongate, sometimes laterally compressed, tapering to the end and without prominent lobes, denticulations, and notches=4; slender, elongate, with small lobes on one side only of distal region=5; slender, elongate, with trident notching at distal end=6; relatively stoutly conical, with shallow to moderately deep notches or widely spaced indentations, sometimes only on posterodorsal edge=7; simple, greatly reduced in size or absent, at least in adults=8.

Hyoid arch and branchiostegals

29) Branchiostegals, number. Seven (3+4)=0; six (2+4)=1; five (1+4)=2; four (0+4)=3.

30) First branchiostegal ray, shape. Slender, almost straight to curved scimitar-shaped=0; somewhat broadened but scimitar-shaped=1; broadly flattened, thin, sometimes irregular shaped=2.

31) Urohyal, shape. Broad triangular outline, with a moderately concave and sometimes irregular posterior edge, and the greatest depth about equal to or greater than the length, with the somewhat thickened anteroventral edge not laterally expanded as a flange=0; narrow triangular outline, only moderately expanding in depth posteriorly from the articular head, much longer than deep, and ventral edge moderately to prominently laterally expanded as a flange of increasing width posteriorly=1; relatively square plate, somewhat concave along posterior edge, and moderately expanded laterally as a flange along anteroventral edge=2; deeply and broadly concave along posterior edge, often sickle-shaped, much higher than long, with distance from anterior articular to deepest edge of concavity one-half or less than height=3; elongate ovoid, gradually increasing moderately in depth from articular head to gently rounded posterior end=4; enormously expanded from articular region of lower jaw to ventral end of cleithrum, with a thickened area vertically in middle of broad lateral surface, and a relatively straight posterior edge=5.

Vertebral column

32) Vertebrae, total number. Twenty four=0; twenty three=1; twenty two=2; twenty five=3; twenty six=4; twenty seven=5; twenty eight=6; twenty nine to thirty=7.

33) Vertebrae, abdominal number. Ten=0; nine=1.

34) Vertebrae, caudal number. Fourteen=0; thirteen=1; nineteen to twenty=2.

35) Vertebral formula. 10+14=0; 10+13=1; 9+13=2; 10+19-20=3; 9+15=4; 10+12=5.

36) Neural spine of first vertebra. Autogenous=0; fixed to its centrum=1.

37) Vacant interneural spaces, number (in most specimens, if about equally frequent in several spaces recorded as polymorphic). None=0; one=1; two=2; three=3.

38) Vacant interneural space, when only one space vacant, location of that space. First space=0; second space=1; third space=2; fourth space=3; fifth space=4; sixth space=5; seventh space=6; eighth space=7; not applicable= –.

39) Vacant interneural spaces, when two or more spaces vacant, number of spaces. Two=0; three=1; not applicable= –.

40) Vacant interneural spaces, when two or more spaces vacant, number of groups of spaces. One=0; two=1; three=2; not applicable= –.

41) Supraneurals, number. Three or more=0; two=1; one=2; none=3.

42) Pleural ribs, presence on anterior abdominal vertebrae. Present on first and immediately following centra=0; absent on first centrum but present on second and immediately following centra=1; absent on first two centra but present on third and immediately following centra=2; absent on first three centra but present on fourth and immediately following centra=3.

43) Pleural ribs, most posterior rib, centrum to which articulated. Last abdominal centrum=0; penultimate abdominal centrum=1; antepenultimate=2.

44) Pleural ribs, most posterior rib, length and shape. Long and slender=0; moderate length and slender=1; moderate length and moderate width=2; short and broad, with its posterior edge situated internal to an anterolateral flange on haemal spine of first caudal vertebra=3; short, slender, and sometimes only a short rudiment that is difficult to determine in radiographs=4.

45) Epineurals. Present=0; absent=1.

46) Neural spine of first abdominal vertebra, length and shape. Long and slender=0; long and relatively broad=1; moderate length and slender=2; short and slender=3; short and broad=4.

47) Abdominal vertebrae parapophyses. Present on at least some of more

posterior centra=0; essentially absent or very small=1.

48) Neural and haemal spines, thickness. Most neural and haemal spines robust and at least moderately wide=0; most neural and haemal spines (except sometimes for first two neurals and haemals) exceptionally slender=1.

Dorsal fin

49) Number of dorsal-fin spines. 0=0; 1=1; 2=2; 3=3; 4=4; 5=5; 6=6; 7=7; 8=8; 9=9; 10=10; 11=11; 12=12; 13=13; 14=14.

50) Number of supernumerary dorsal spines (or soft rays if dorsal spines absent). 0=0; 1=1; 2=2.

51) Dorsal-fin spines (or more anterior soft rays if dorsal spines absent), elongation. Not greatly elongated and filamentous distally=0; all but first two or three dorsal spines elongated and filamentous=1.

52) Number of soft dorsal rays. 0 to10=0; 11 to 20=1; 21 to 30=2; 31 to 40=3; 41 to 50=4; 51to 60=5.

53) First dorsal spine (or first soft ray if dorsal spines absent/?). Visible externally, protruding through skin=0; not apparent externally, specialized as bony cap rotating on pterygiophore beneath

skin=1; reduced to nubbin on surface of pterygiophore just in front of base of well-developed second spine=2.

54) First dorsal spine, when visible externally. Relatively long, at least one-half length of second spine (or of first soft ray if only one spine present)=0; relatively short, less than one-half length of second spine (or about one-half if second spine is relatively short)=1; not applicable= –.

55) First dorsal-fin pterygiophore, position of ventral shaft relative to interneural spaces (in most specimens; if about equally frequent in several spaces recorded as polymorphic). In preneural space, often to rear of skull=0; in first interneural space (or over shorten open neural spine of first vertebra)=1; in second interneural space=2; in third interneural space=3; in fourth interneural space=4.

56) Number of anterior dorsal-fin pterygiophores present in preneural space, when preneural space is occupied. Only the first spiny-dorsal pterygiophore=0; both the first and second spiny-dorsal pterygiophores=1; not applicable= –.

57) Dorsal fin pterygiophores, number in tenth interneural space. One=0; two=1; three=2.

58) Specialized locking mechanism of first dorsal-fin spine. No specialized mechanism involving a deep indentation in dorsal region of first pterygiophore (proximal radial) and a median flange that is ribbed or roughened around which deeply concave base of first dorsal spine can rotate and lock=0; moderately specialized mechanism present, with a slightly to moderately deep indetation (maximum depth of indentation at level of base of first dorsal spine, and not well below it), and roughened or somewhat ribbed median flange=1; highly specialized mechanism, with an exceptionally deep (to well below level of base of first dorsal spine) indentation and a strongly ribbed median flange=2; not applicable when dorsal spines absent= –.

59) Number of dorsal-fin spines and median pterygial flanges involved when specialized locking mechanism present. One (first spine only)=0; three (first, second, and third spines)=1; not applicable= –.

60) Shape of anterodistal region of first pterygiophore (proximal radial) when most specialized first dorsal spine locking mechanism present (i.e., Acanthuridae). Anterodistal region only moderately if at all posterolaterally expanded and not encompassing basal region of first dorsal spine=0; prominently expanded posterolaterally and encompassing basal region of reduced first dorsal spine=1; not applicable= –.

61) Distal ends of dorsal- and anal-fin pterygiophores. Not forming a median truss around most of body=0; forming a median truss around most of body=1.

62) Dorsal- and anal-fin rays segmentation. Dorsal and anal rays segmented=0; Unsegmented=1.

63) Dorsal, anal, caudal, pectoral, and pelvic rays, spinules. Rays of these fins without spinules=0; small spinules present on most fins laterally along all or most of their lengths=1.

64) Distal ends of dorsal- and anal-fin pterygiophores. Not exceptionally expanded laterally=0; prominently expanded laterally=1.

65) First two haemal spines, orientation to one another. Parallel or divergent to one another=0; middle regions curved toward one another, at least in adults=1.

Anal fin

66) Number of anal spines. 0=0; 1=1; 2=2; 3=3; 4=4; 5=5; 6=6; 7=7; 8=8.

67) Number of supernumerary anal spines (or soft rays if anal spines absent). 0=0; 1=1; 2=2.

68) Number of soft anal rays. 0 to10=0; 11 to 20=1; 21 to 30=2; 31 to 40=3; 41 to 50=4; 51 to 60=5.

69) First anal spine (or first ray if anal spines absent). Visible externally, protruding through skin=0; not apparent externally, specialized as bony cap rotating on pterygiophore beneath skin=1.

70) First anal spine, when visible externally. Relatively long, at least one-half length of second spine=0; relatively short, less than one-half length of second spine (or about one-half if second spine is relatively short)=1; not applicable= –.

71) Anal fin pterygiophores, number in first interhaemal space. One=0; two=1; three=2; four or five=3; six or seven=4.

72) Specialized locking mechanism of first anal-fin spine. No specialized mechanism involving a deep indentation in ventral region of first pterygiophore (proximal radial) and no median flange that is ribbed or roughened around which deeply concave base of first anal spine can rotate and lock=0; moderately specialized mechanism present, with a slightly to moderately deep indentation (maximum depth of indentation at level of base of first anal spine and not well above it), and roughened or somewhat ribbed median flange=1; highly specialized mechanism, with an exceptionally deep indentation (to well above level of base of first anal spine) and a strongly ribbed median flange=2; not applicable= –.

73) Shape of anterodistal region of first proximal radial (pterygiophore) when most specialized first anal spine locking mechanism is present (i.e., Acanthuridae). Anterodistal region only slightly if at all posterolaterally expanded and not encompassing basal region of first anal spine=0; moderately to prominently expanded posterolaterally and partially encompassing basal region of reduced first anal spine=1; not applicable= –.

74) First anal-fin pterygiophore, orientation. Relatively vertical=0; slightly to moderately oblique=1; highly oblique=2.

75) Distal end of first anal-fin pterygiophore. Not greatly prolonged or positioned anteriorly=0; greatly prolonged or positioned anteriorly=1.

Caudal fin and caudal skeleton

76) Caudal fin, number of principal rays. Seventeen=0; sixteen=1; fifteen=2; fourteen=3; thirteen=4; twelve=5.

77) Caudal fin, number of procurrent rays (largest number either below or above). Three or more=0; one or two=1; none=2.

78) Caudal fin shape. Rounded=0; truncate=1; somewhat concave=2; deeply and decidedly concave or forked=3.

79) Hypurostegy. None, or only slight overlapping=0; moderate overlapping=1; extensive overlapping, with hypural plate nearly fully covered by proximal ends of caudal-fin rays=2.

80) Caudal peduncle depth. Relatively deep (ca. 6 to 12 times in SL, or between ca. 8 to 20% of SL)=0; moderate depth (ca. 13 to 17 times in SL, or between ca. 6 and 8% of SL)=1; relatively slender (ca. 18-21 or more times in SL, or between ca. 4 and 6% or less of SL)=2.

81) Hypurals one to four, number of separate elements in adults. Four=0; two, with hypurals 1+2 consolidated to one another and 3+4 consolidated to one another, and one or both of these plates fused to urostylar centrum=1; two, with hypurals 1+2 and 3+4 consolidated as in preceding state but with neither of these plates fused to urostylar centrum=2; one, with hypurals 1+4 consolidated with one another and fused to urostylar centrum=3; one, with hypurals 1+4 consolidated with one another but not fused to urostylar centrum=4; three, with hypurals 1 and 2 separate and hypurals 3+4 consolidated with one another, and none these three elements fused to urostylar centrum=5.

82) Hypural five. Hypural 5 and the anterior uroneural remain autogenous=0; hypural five remains separate from hypural four but at least its anterior end fuses to the embracing uroneural pair which in turn fuses to the urostyle=1.

83) Epurals, number of separate elements in adults. Three=0; two=1; one=2.

84) Parhypural foramen, posterior margin. Formed by anteroventral margin of first hypural=0; formed by a dorsally positioned flange of bone from parhypural that separates foramen from first hypural=1.

85) Hypurapophysis. Terminates as a more or less sharp pointed to blunt or knob-like process=0; terminates in a T-shaped expansion oriented about 45 degrees to horizontal=1; terminates as a T-shaped expansion oriented about horizontally=2.

86) Uroneurals (including stegural), number (exclusive of nubbins). Two pairs=0; one pair (halves may be fused together)=1; none=2.

87) Uroneurals, length of first pair. Long, extending posteriorly well beyond region of urostylar centrum, reaching to above uppermost hypural or second uroneural=0; short, not extending prominently beyond region of urostylar centrum and not reaching to above uppermost hypural=1.

88) Neural spine of second preural centrum (NPU2), length. Long=0; short=1.

89) Autogenous haemal spines on centra anterior to terminal centrum and parhypural. One (on HPU2)=0; two (on HPU2 and 3)=1; three (on HPU2, 3, and 4)=2; four (on HPU2, 3, 4, and 5)=3; none=4.

Pectoral girdle and pectoral fin

90) Postcleithrum, number of elements. Two=0; one=1,

91) Postcleithrum, shape. Relatively elongate and shaft-like=0; a large plate, greatly expanded posteriorly=1.

92) Postcleithrum, ventral extension. Distal end of postcleithrum not in contact with first anal-fin pterygiophore=0; distal end of postcleithrum contacting or in close association with anterior extension of first anal-fin pterygiophore=1.

93) Pectoral fin situated in about middle of body or lower=0; situated high on body=1.

94) Coracoid, postcoracoid process. Present=0; absent=1.

95) Supracleithrum, sensory canal. Present, with posttemporal canal joining main trunk lateral line through a short bony canal at dorsal end of supracleithrum=0; absent, with main trunk lateral line canal communicating directly with posttemporal=1.

Pelvic girdle and pelvic fin

96) Number of pelvic-fin spines. None (at least in adults; entire fin in some species may become rudimentary with increasing specimen size)=0; one (on outer side of fin rays)=1; spine and rays consolidated into an operculum ani in adults=2.

97) Number of soft pelvic-fin rays. None (at least in adults)=0; one=1; two=2; three=3; four=4; five=5; six=6.

98) Pelvis, posterior process (ishiac). Posterior extension of short to moderate length or essentially absent=0; a long posterior extension making contact and sutured to anterior extension of first anal-fin pterygiophore=1.

99) Subpelvic keel (anterior iliac process from anteroventral end of pelvis). Anterior process prominent, and area above it distinctly concave=0; anterior process only poorly developed, and area above it only slightly if at all concave=1; no anterior prong-like process and no distinct concave area along anteroventral region of pelvis=2.

100) Pelvis (basipterygium) greatest depth to length ratio (length from anterior end of either anterodorsal ascending pubic process or from anteroventral iliac process to posterior end of pelvis, including any posterior ishiac process, whichever measurement is longer). 56–65% or more=0; 46–55%=1; 36–45%=2; 26–35%=3; 21–25%=4; 16–20%=5; 11–15%=6; 6–10%=7.

Scales

101) Scales. Ctenoid=0; spinoid/spinous (with upright or posterior edge non-articulated spinules or large granulations or hillocks)=1; cycloid=2; apparently without scales=3.

102) Greatly enlarged specialized scales on caudal peduncle. Absent=0; present either as fixed plates or a folding spine=1.

103) Greatly enlarged specialized scales on caudal peduncle, structure when present. Fixed plates=0; folding spine in a groove of variable depths and widths=1; not applicable= –.

104) Greatly enlarged specialized scales on caudal peduncle, when present as fixed plates. One or two fixed plates=0; three or more fixed plates=1; not applicable=–.

105) Tubercular scale plates on snout. Absent=0; large (ca. one-half pupil diameter, or about one-fourth to one-fifth orbit diameter) tubercular scale plates present on the snout between eye and upper jaw=1.

106) Somewhat enlarged scales (ca. twice or more as large as the surrounding small body scales) on the body between the posterior regions of the soft dorsal and anal fin bases and anterior caudal peduncle. Absent=0; present=1; not applicable when scaleless= –.

Miscellaneous

107) Anus, position (in fossils inferred as just in front of anterodistal end of first anal-fin pterygiophore). Anus not much displaced anteriorly=0; anus moderately displaced anteriorly=1; anus displaced far forward=2.

Table S2. Ages and distribution from each fossil used in this study.

Fossil	**Age (Ma)**	**Distribution**	**References**
Eozanclus brevirostris	48.5–50.5	Monte Bolca; late Ypresian	(Agassiz 1835)
Angiolinia mirabilis	48.5–50.5	Monte Bolca; late Ypresian	(Carnevale and Tyler 2024)
Massalongius gazolai	48.5–50.5	Monte Bolca; late Ypresian	(Carnevale et al. 2014)
Gazolaichthys vestenanovae	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990; Tyler 2005)
Eonaso deani	23–28.5	Oliver Nugent quarry, Antigua and Barbuda; Oligocene	(Hussakof 1907)
Tauichthys padremenini	48.5–50.5	Monte Bolca; late Ypresian	(Tyler 1999)
Tauichthys aspesae	48.5–50.5	Monte Bolca; late Ypresian	(Tyler and Bannikov 2000)
Pesciaraichthys punctatus	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Frigosorbiniae baldwinae	48.5–50.5	Monte Bolca; late Ypresian	(Sorbini and Tyler 1998)
Protozebrasoma bloti	48.5–50.5	Monte Bolca; late Ypresian	(Sorbini and Tyler 1998)
Arambourgthurus scombrurus	34	Istehbanat, Iran; late Eocene	(Carnevale and Tyler 2018)
Sorbinithurus sorbinii	48.5–50.5	Monte Bolca; late Ypresian	(Tyler 1999)
Marosichthys huismani	16–20.4	Patoenoeang Asoe, Tonasa Formation, Sulawesi, Indonesia; Burdigalian	(de Beaufort 1926)
Glarithurus friedmani	30	Landesplattenberg slate quarry, Matt Formation, Canton Glarus, Switzerland; Rupelian	(Tyler and Micklich 2011)
Caprovesposus parvus	38.4–42.1	Kuma Horizon (Kuma Formation), Gumista River near Sukumi, Abkhazia, Georgia; middle-late Eocene	(Danil'chenko 1960)
Padovathurus gaudryi	48.5–50.5	Monte Bolca; late Ypresian	(Carnevale et al. 2014)
Proacanthurus bonatoi	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Proacanthurus elongatus	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Proacanthurus ovalis	48.5–50.5	Monte Bolca; late Ypresian	(Agassiz 1838)
Proacanthurus tenuis	48.5–50.5	Monte Bolca; late Ypresian	(Agassiz 1838)
Eorandallius rectifrons	48.5–50.5	Monte Bolca; late Ypresian	(Agassiz 1838)

Eorandallius elegans	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Tylerichthys nuchalis	48.5–50.5	Monte Bolca; late Ypresian	(Agassiz 1838)
Tylerichthys milani	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Metaspisurus emmanueli	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Lehmanichthys lessiniensis	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Acanthuroides massalongoi	48.5–50.5	Monte Bolca; late Ypresian	(Blot and Tyler 1990)
Luvarus necopinatus	55.8	Uylya-Kushlyuk, Danatinsk Formation, Turkmenistan; basal Eocene	(Bannikov and Tyler 1995)
Avitoluvarus eocaenicus	38.4–42.1	Left bank of Pshekha River, about 0.5 km from the Gorny Luch farmstead, Apsheronsk District, Krasnodar Region, Kuma Formation, Georgia; upper part of the Middle Eocene	(Bannikov and Tyler 2001)
Avitoluvarus dianae	55.8	Uylya-Kushlyuk, Danatinsk Formation, Turkmenistan; basal Eocene	(Bannikov and Tyler 1995)
Avitoluvarus mariannae	55.8	Uylya-Kushlyuk, Danatinsk Formation, Turkmenistan; basal Eocene	(Bannikov and Tyler 1995)
Kushlukia permira	55.8	Uylya-Kushlyuk, Danatinsk Formation, Turkmenistan; basal Eocene	(Bannikov and Tyler 1995)

Total-evidence dating analyses and phylogenetic uncertainty

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 (Ronquist et al. 2012). To account for topological uncertainty, we assembled largely independent subsets (randomly subsampled from the expanded matrix with genes only), each with enough genes to overcome sampling error. We divided the complete dataset into 20 (50 loci x13 + 49 loci x7), 10 (99 loci x7 + 100 loci x3), and 5 (199 loci x3 + 198 loci x2) gene subsets. To maintain the same number of species for each subset, all subsets overlapped in nine anchor genes. We ran ML trees using *raxml_ng* for each subset and found high levels of topological discordance, particularly for trees estimated with fewer genes (Fig. S8). We thus used the trees inferred with five subsets of 198-199 markers each for downstream analyses.

Each genomic subset was combined with the morphological dataset with fossil and extant taxa to include a total of 112 taxa. The ages of each fossil used to estimate divergence times are provided in Table S2. First, we performed two null tests using one of the subsets. First, using a secondary calibration for stem Acanthuriformes (68–75 Ma) with *Chaetodon striatus* as the outgroup (see Table S3 for details); and second, omitting this root calibration. For these analyses, the age for crown acanthuriformes was 73.8 Ma and 80.8 Ma respectively. Based on these null results and also other large-scale phylogenetic studies that have estimated a younger age for the group (e.g., Near et al. 2013; Betancur-R. et al. 2017; Ghezelayagh et al. 2022; Table S3), we decided to use instead a secondary calibration for crown Acanthuriformes (55.8–64 Ma). We also constrained families to be monophyletic. We estimated speciation/extinction priors ("fit.bd" function in *phytools*) based on a well-sampled tree for this group (Rabosky et al. 2018) but after rescaling its root (crown Acanthuriformes) from 77 Ma to 60 Ma based on other previous age estimates (see Table S3). After independently estimating phylogenies based on datasets with morphology only and combining molecules with morphology, we noticed that there were two fossils, †*Gazolaichthys vestenanovae* and †*Padovathurus gaudryi*, that had incongruent placements in the resulting trees. In the combined matrix these fossils appear as the sister group of Zanclidae+Acanthuridae (Scheme 1, Fig. S7), while on the morphology-only dataset (Fig. S6) and a previous study (Siqueira et al. 2019) their placement was sister to Acanthuridae (Scheme 2). To address this ambiguity, we chose to include these two schemes as constraints in our dating analyses, conducting a total of 10 analyses by running MrBayes for each scheme on every subset. Each analysis was run with eight independent runs and four Monte Carlo Markov chains (MCMCs) for over 350 million generations each, sampling every 10,000 generations. We used a sample probability of 0.94 and a relaxed clock model with the clock rate prior following a log normal distribution and independent gamma rate (IGR). The first 10% of trees sampled were discarded as relative burn-in and convergence of the MCMC was verified using the estimated sample size (ESS) criterion for each parameter in TRACER v. 1.7 (Rambaut et al. 2018). After more than 14 months of total runtime, we found that 9 (of the 10) analyses reached convergence where ESS values were close to or above 200 (Table S4). Because these analyses ran for over a year, we removed the Subset 2 based on the Scheme 2 for all downstream analyses. We sampled ~2000 trees for Scheme 1 and ~2500 trees for Scheme 2 evenly distributed along the posterior distribution from each subset to have a total of 10000 trees. For each scheme independently, we inferred a Maximum Clade Credibility tree (MCC tree) in TreeAnnotator v. 2.7.5 (Drummond and Rambaut 2007). To obtain a posterior distribution (PD) of trees for phylogenetic comparative analyses, depending on the type of analysis and their computational time, we sampled either 100 or 4 trees from each subset, resulting in a total of 500 and 20 trees for each scheme.

Table S3. Ages of Acanthuriformes (crown and stem) estimated by previous studies.

Study	**Crown Acanthuriformes**	**Acanthuriformes + Chaetodontidae**	**Acanthuriformes + Ephippidae**	**Acanthuriformes + (Chaetodontidae+ Ephippidae)**	**Acanthuridae**	**Comments**
Alfaro et al. (2018)		68 (63–72)			31 (16–52)*	No Ephippidae, Luvaridae, Zanclidae. One species per genus
Betancur-R et al. (2013)	55*	70	100			
Betancur-R et al. (2017)	64			90	50	
Ghezelayagh et al. (2022)	59 (55–65)	69 (62–74)	76 (72–82)			
Hughes et al. (2018)		88 (77–90)*	90 (89–95)			
Near et al. (2012)	64	75			60	No Ephippidae
Near et al. (2013)	61	68			51	No Ephippidae
Rabosky et al. (2018)	77.8*		96.1		57.7	
Siqueira et al. (2019)					80.9 (67.5–95.6)*	Using Zanclidae and Luvaridae as outgroup
Sorenson et al. (2013)					54 (51–62)	Using Zanclidae and Luvaridae as outgroup

* = outliers

Table S4. MrBayes runs and convergence statistics for each subset and scheme (Scheme 1 and Scheme 2).

Type of subset	Number of generations for each run	Number of generations combined	Estimated sample size (ESS) values
Null test, Subset1, No constraints, no root	265,930,000	1,914,720,000	>204 ESS
Null test, Subset1, No constraints, Second calibration at the root (stem Acanthuriformes)	209,560,000	1,508,880,000	>123 ESS
Subset1, Scheme1, With constraints, Secondary calibration at the root (crown Acanthuriformes)	400,000,000	2,880,000,000	>166 ESS
Subset2, Scheme1, With constraints, Secondary calibration at the root (crown Acanthuriformes)	347,310,000	2,500,640,000	>203 ESS
Subset3, Scheme1, With constraints, Secondary calibration at the root (crown Acanthuriformes)	379,300,000	2,739,960,000	>257 ESS
Subset4, Scheme1, With constraints, Secondary calibration at the root (crown Acanthuriformes)	285,900,000	2,778,480,000	>220 ESS
Subset5, Scheme1, With constraints, Secondary calibration at the root (crown Acanthuriformes)	333,710,000	2,402,720,000	>168 ESS
Subset1, Scheme2, With constraints, Secondary calibration at the root (crown Acanthuriformes)	400,000,000	2,880,000,000	>234 ESS
Subset2, Scheme2, With constraints, Secondary calibration at the root (crown Acanthuriformes)	150,420,000	1,083,040,000	>58 ESS
Subset3, Scheme2, With constraints, Secondary calibration at the root (crown Acanthuriformes)	389,420,000	2,803,840,000	>151 ESS
Subset4, Scheme2, With constraints, Secondary calibration at the root (crown Acanthuriformes)	349,050,000	2,513,200,000	>117 ESS
Subset5, Scheme2, With constraints, Secondary calibration at the root (crown Acanthuriformes)	360,500,000	2,595,560,000	>162 ESS

Sequencing and assembly of chromosome-level genome

We generated a phased chromosome-level genome and transcriptome for *Acanthurus chirurgus*. We obtained flash-frozen muscle tissues from an individual caught using a hand net in the Florida Keys (24°59.564 N, 80°25.753 W), US, by Phillip Rauch on the 29th of January of 2023. The voucher specimen is deposited at Scripps Institution of Oceanography (SIO), collection number SIO 24-10. Subsequently, we outsourced the DNA and RNA extractions, library preparations, sequencing, assembly, and annotation to Cantata Bio LLC.

-DNA extraction, PacBio library, and sequencing: High molecular weight (HMW) DNA was extracted using the Qiagen Blood and Cell Culture DNA Kit following the manufacturer's protocol. DNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20kb) for PacBio Sequel was constructed using SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) and the manufacturer recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio) and loaded onto PacBio Sequel II. Sequencing was performed on PacBio Sequel II 8M SMRT cells. PacBio CCS reads were used as an input to Hifiasm v. 0.15.4-r347 with default parameters. Blast results of the Hifiasm output assembly against the nt database were used as input for blobtools2 v. 1.1.1 and scaffolds identified as possible contamination were removed from the assembly. Finally, purge_dups3 v. 1.2.5 was used to remove haplotigs and contig overlaps. This process yielded one *de novo* assembly for each haplotype (Cheng et al. 2022).

-***Dovetail Omni-C library preparation and sequencing:*** For each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde in the nucleus. Fixed chromatin was digested with DNase I and then extracted, chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed, and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce ~30x sequence coverage (Lieberman-aiden et al. 2009).

-***Scaffolding the assembly with Omni-C HiRise:*** The *de novo* assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail OmniC library sequences were aligned to the draft input assembly using bwa (https://github.com/lh3/bwa). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

-***Transcriptome sequencing (RNA Seq):*** Total RNA extraction was done using the QIAGEN RNeasy Plus Kit following manufacturer protocols. Total RNA was quantified using Qubit RNA Assay and TapeStation 4200. Prior to library prep, we performed DNase treatment followed by AMPure bead clean up and QIAGEN FastSelect HMR rRNA depletion. Library preparation was done with the NEBNext Ultra II RNA Library Prep Kit following manufacturer protocols. Then these libraries were run on the NovaSeq6000 platform in 2 x 150 bp configuration.

-***Assembly statistics:*** Contiguity statistics of scaffolded assembly per haplotype was computed using Quast (Gurevich et al. 2013). The completeness of the genome was estimated in Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.0.0 (Simão et al. 2015) using the single-copy orthologs for ray-finned fishes database (actinopterygii_odb9). Scaffolds were assigned chromosomal numbers based on their length as no other surgeonfish genome was available at the time of assembly.

-***Annotation:*** Repeat families found in the genome assembly of *A. chirurgus* were identified *de novo* and classified using the software package RepeatModeler v. 2.0.1 (Flynn et al. 2020).

RepeatModeler depends on the programs RECON v. 1.08 (Bao and Eddy 2002) and RepeatScout v. 1.0.6 (Price et al. 2005) for the *de novo* identification of repeats within the genome. The custom repeat library obtained from RepeatModeler were used to discover, identify, and mask the repeats in the assembly file using RepeatMasker v. 4.1.0 (Smit et al. 2013–2015). Coding sequences from *Acanthochromis polyacanthus*, *Chaetodon austriacus*, and *Chelmon rostratus* were used to train the initial ab initio model for *A. chirurgus* using the AUGUSTUS v. 2.5.5 software (Stanke et al. 2006). Six rounds of prediction optimisation were done with the software package provided by AUGUSTUS. The same coding sequences were also used to train a separate *ab initio* model for *Acanthurus chirurgus* using SNAP (v. 2006-07-28; https://github.com/ KorfLab/SNAP). RNAseq reads were mapped onto the genome using the STAR v. 2.7 aligner software (Dobin et al. 2013) and intron hints generated with the bam2hints tools within AUGUSTUS. MAKER2 (Holt and Yandell 2011), SNAP and AUGUSTUS (with intron-exon boundary hints provided from RNA-Seq) were then used to predict for genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database (Bateman 2019) were downloaded and used in conjunction with the protein sequences from *A. polyacanthus*, *C. austriacus*, and *C. rostratus* to generate peptide evidence in the MAKER pipeline. Only genes that were predicted by both SNAP and AUGUSTUS software were retained in the final gene sets. To help assess the quality of the gene prediction, AED scores were generated for each of the predicted genes as part of the MAKER2 pipeline. Genes were further characterized for their putative function by performing a BLAST search of the peptide sequences against the UniProt database. tRNAs were predicted using the software tRNAscan-SE v. 2.05 (Chan and Lowe 2019).

Supplementary Results

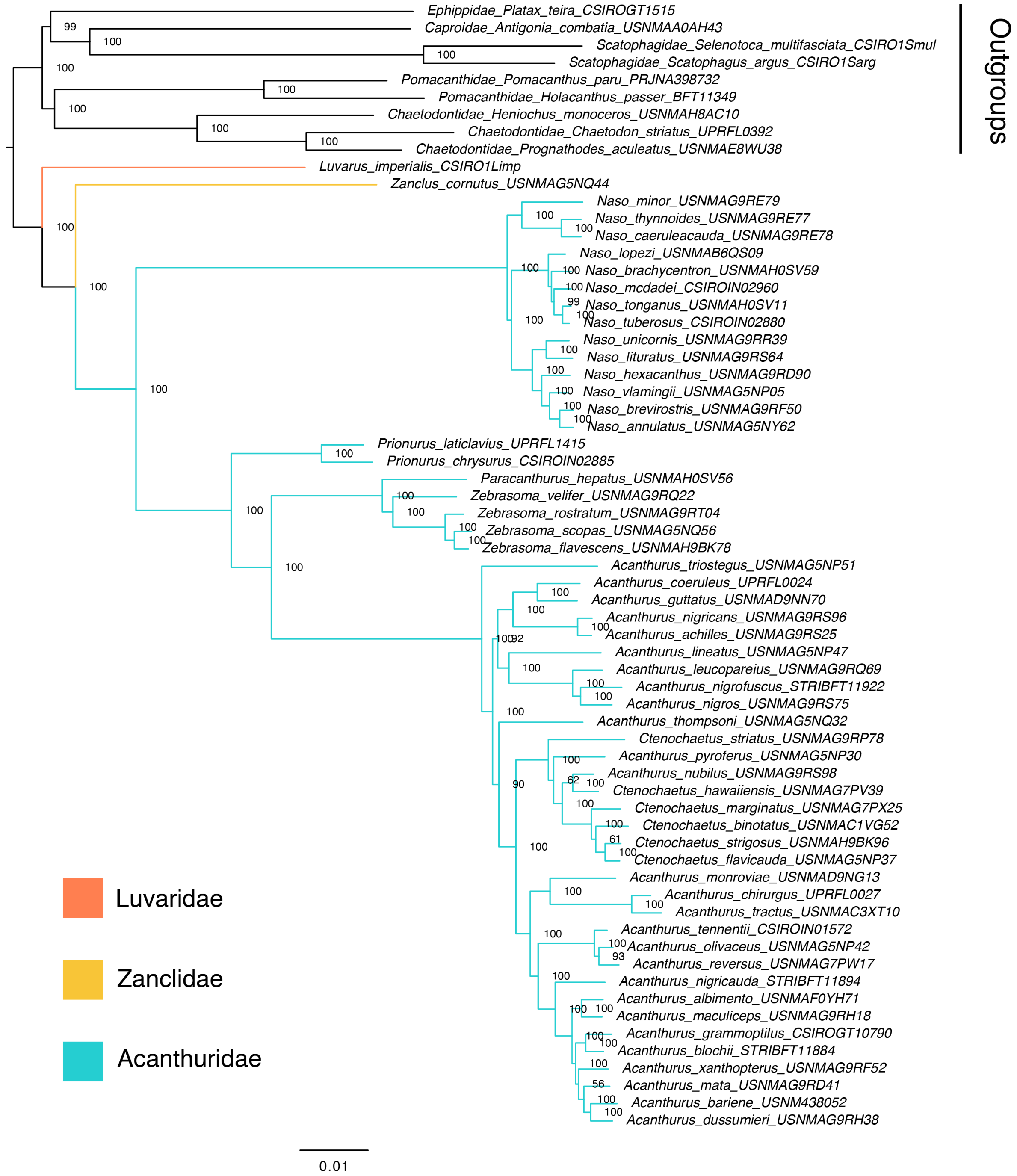


Figure S1. Phylogeny of Acanthuriformes based on concatenation analysis of the reduced matrix comprised by 998 loci and 66 species (56 Acanthuriformes, 9 outgroups). Phylogenetic tree inferred with RAxML using the best fit partition scheme identified with PartitionFinder for all newly sequenced taxa. Colors indicate families; nodal values indicate bootstrap support.

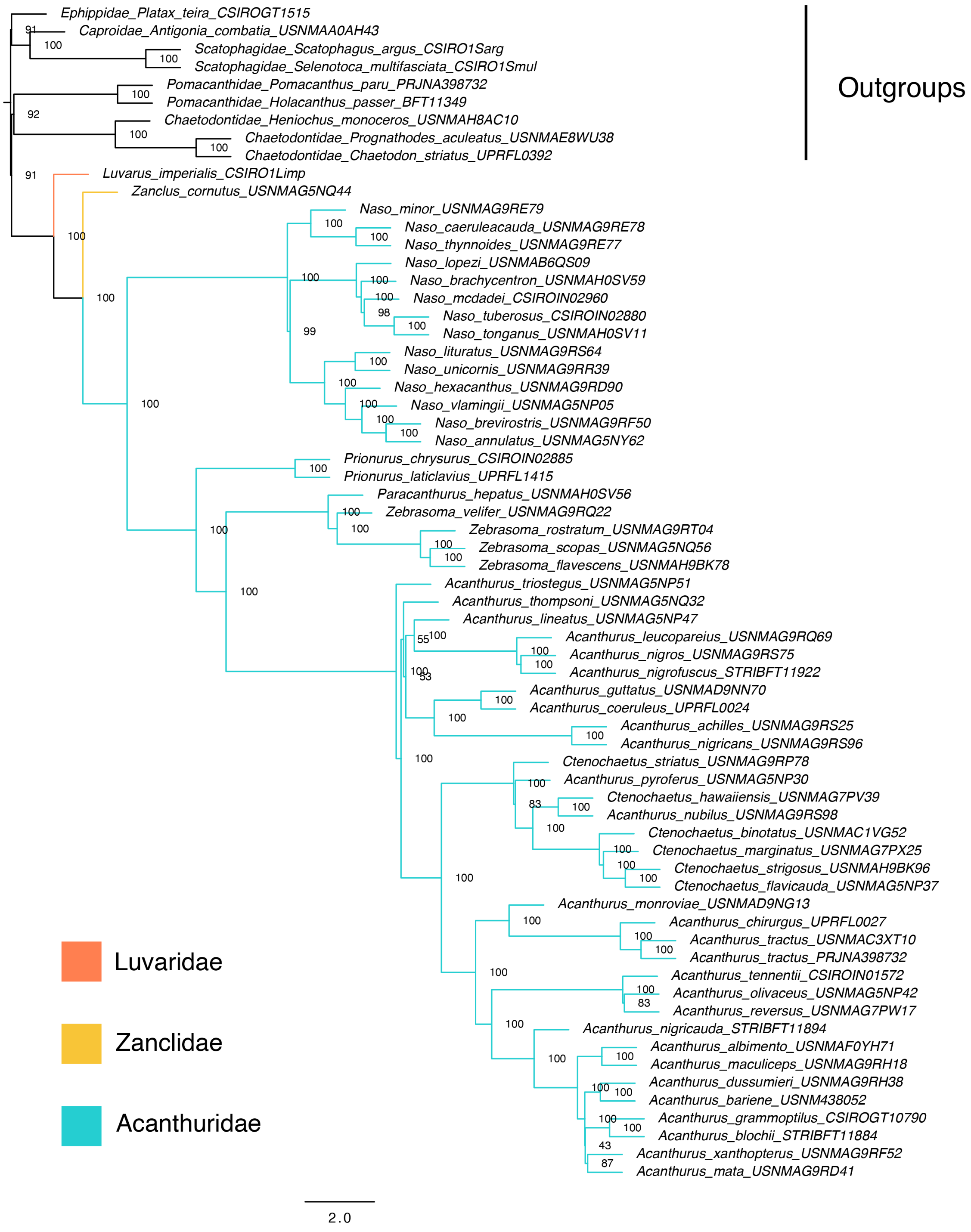


Figure S2. Phylogeny of Acanthuriformes based on multi-species coalescent analysis of the reduced matrix comprised by 998 loci and 66 species (56 Acanthuriformes, 9 outgroups). Phylogenetic tree inferred with ASTRAL-III for all newly sequenced taxa. Colors indicate families; nodal values indication bootstrap support.

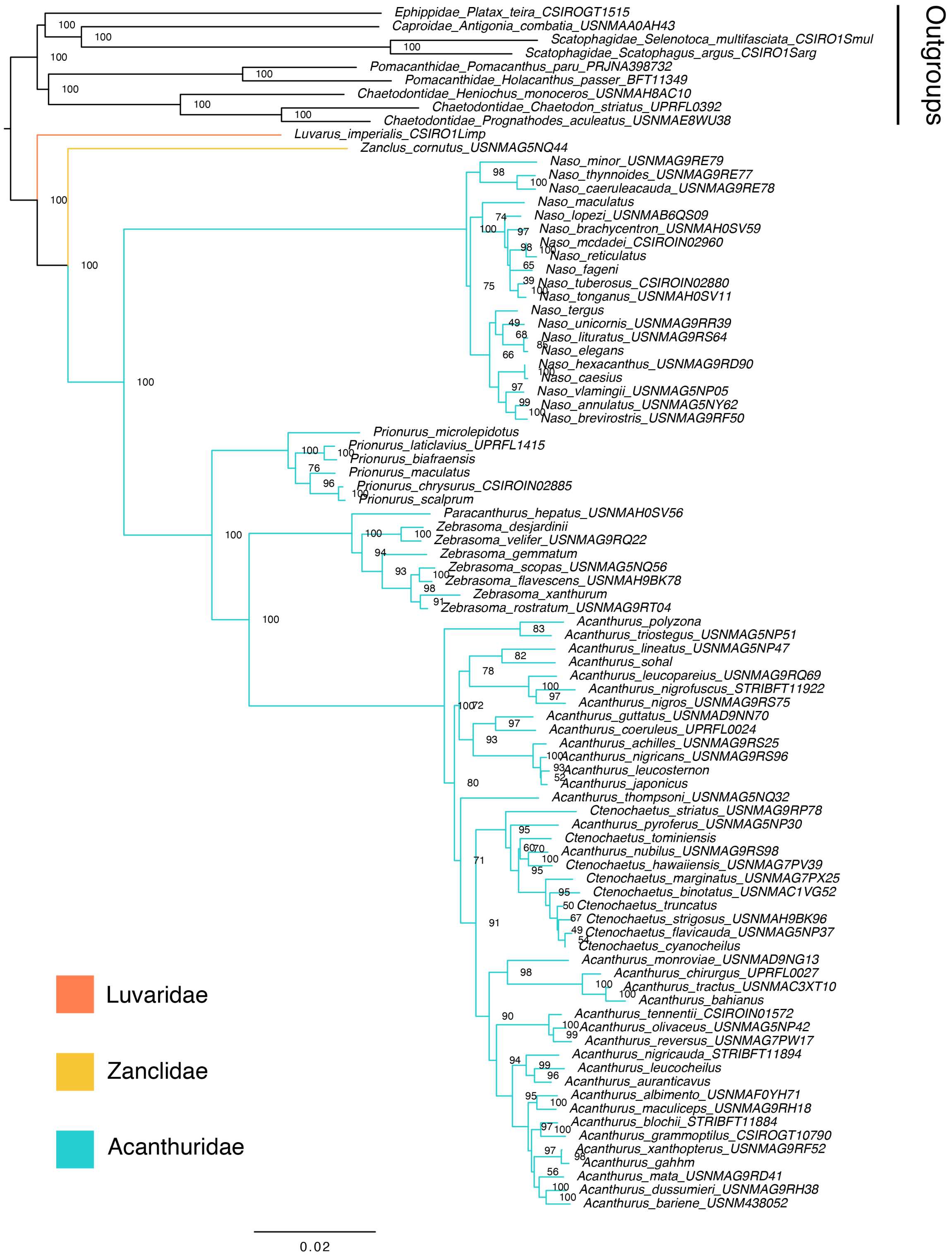


Figure S3. Phylogeny of Acanthuriformes based on concatenation analysis of the expanded matrix comprised by 999 loci and 89 species (80 Acanthuriformes, 9 outgroups). Phylogenetic tree inferred with RAxML using the best fit partition scheme identified with PartitionFinder for all newly sequenced taxa. Colors indicate families; nodal values indicate bootstrap support.

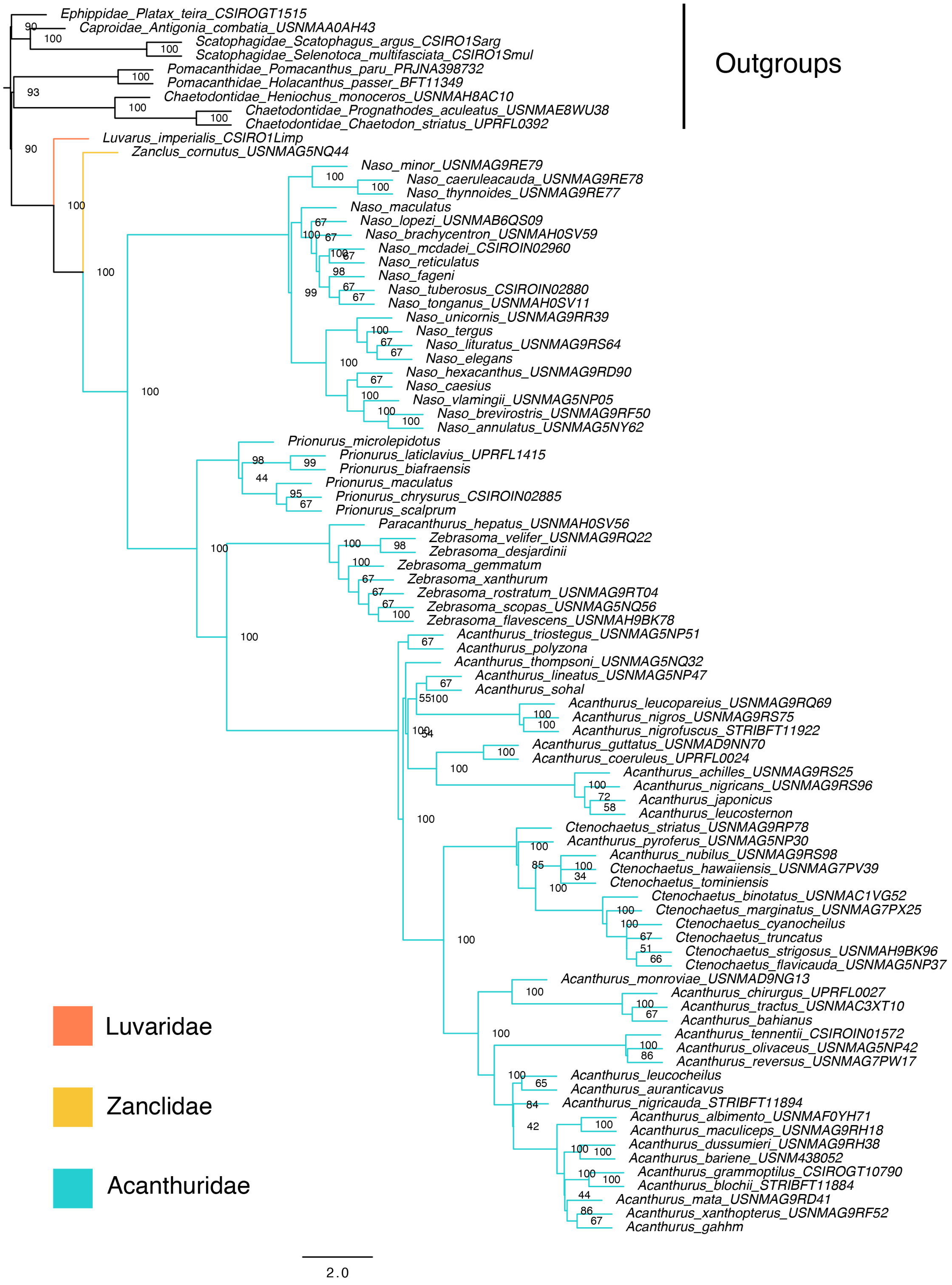


Figure S4. Phylogeny of Acanthuriformes based on multi-species coalescent analysis of the expanded matrix comprised by 999 loci and 89 species (80 Acanthuriformes, 9 outgroups). Phylogenetic tree inferred with ASTRAL-III for all newly sequenced taxa. Colors indicate families. Nodal values indicate bootstrap support.

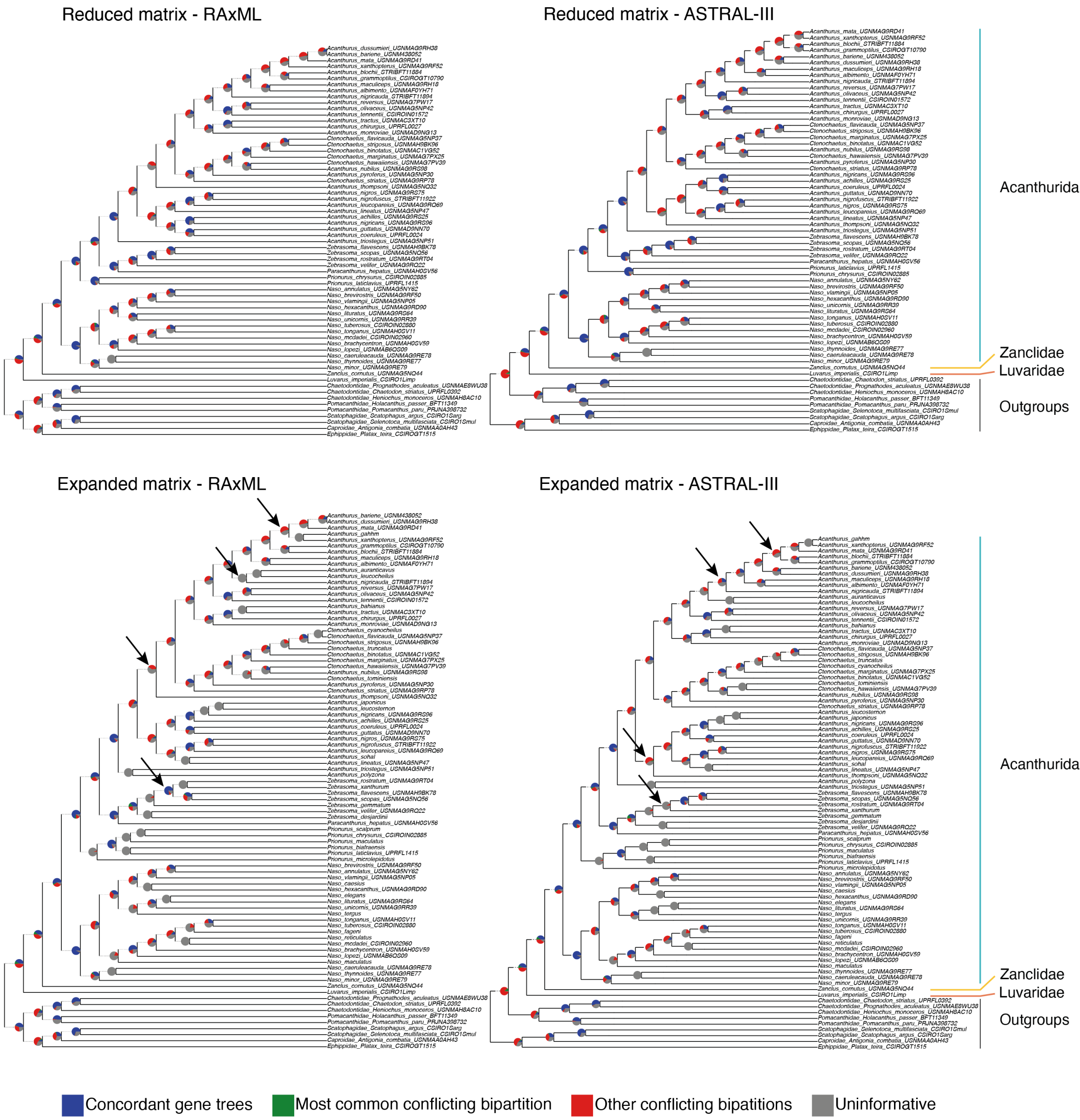


Figure S5. Discordance among the phylogenetic trees inferred in RAxML and ASTRAL-III based on reduced and expanded matrices, and their corresponding gene trees. Pie charts at each node indicate the proportion of gene trees that support the illustrated clade (blue), the proportion that supports the most common conflicting bipartition (green), the proportion that supports other conflicting bipartitions (red), and the proportion that are uninformative for each branch with less 50% bootstrap support, representing gene trees with no information (gray). Arrows indicate the more unstable nodes among RAxML and ASTRAL-III trees with the expanded matrix.

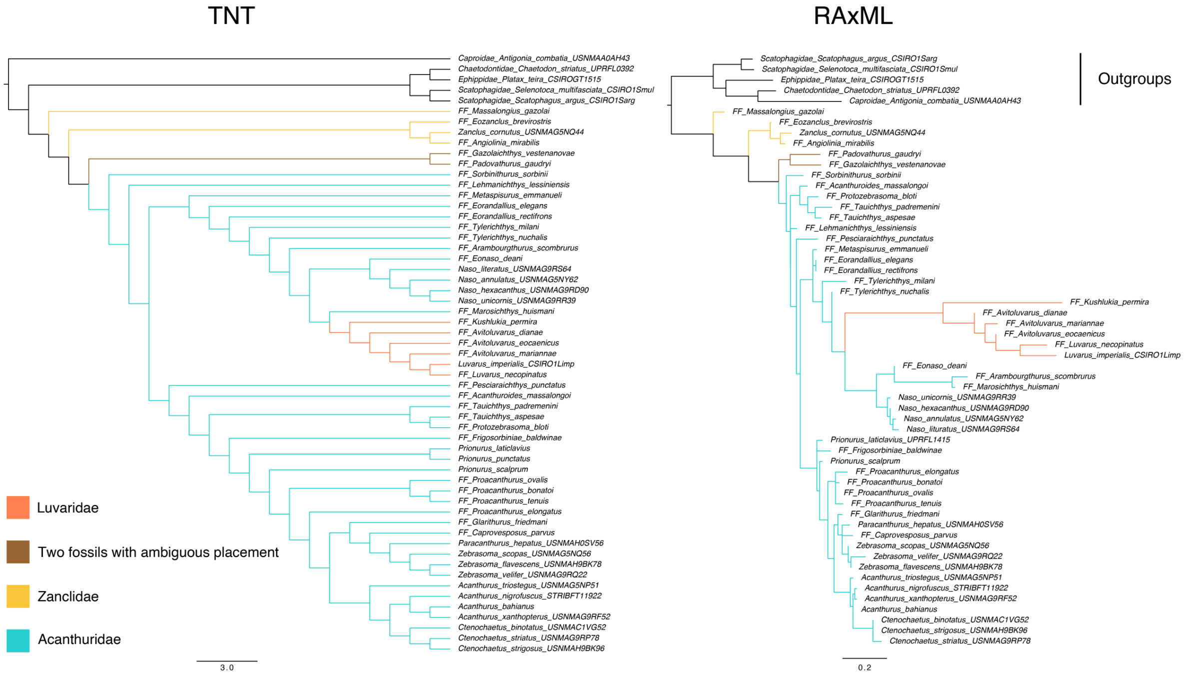


Figure S6. Phylogenies of Acanthuriformes based on the morphological matrix comprised by 107 characters and 56 species (32 fossil + 19 extant Acanthuriformes, 5 outgroups). The phylogeny on the left was estimated using TNT (parsimony), while the one on the right was inferred with RAxML (maximum likelihood, MK model). Colors indicate different families. Note that the brown color represents two fossils with ambiguous placements in combined molecular and morphological analyses (see Materials and Methods and Figs. S7, S9, and S10).

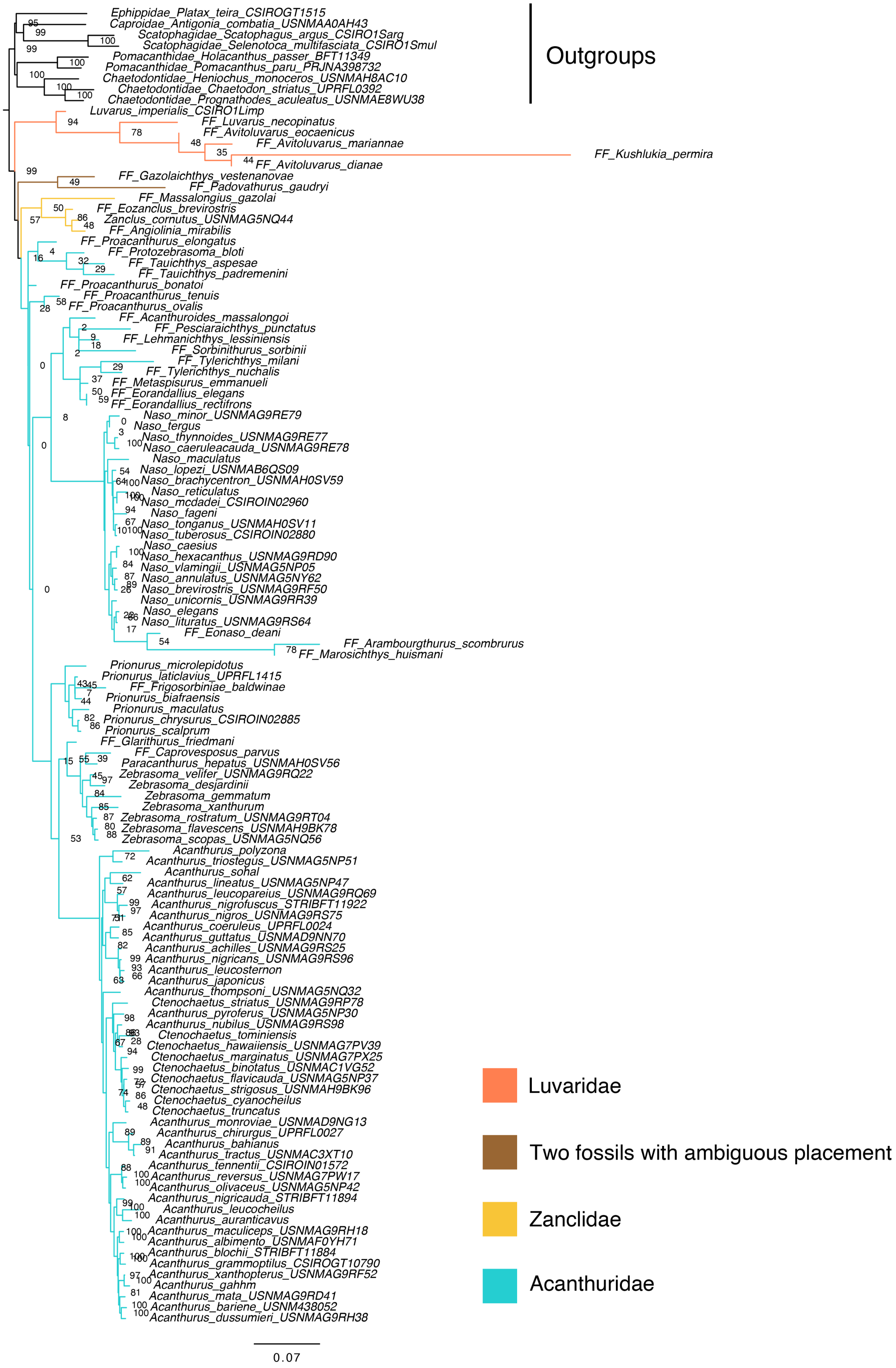

Ephippidae_Platax_teira_CSIROGT1515
Caproidae_Antigonia_combatia_USNMAA0AH43
Scatophagidae_Scatophagus_argus_CSIRO1Sarg
Scatophagidae_Selenotoca_multifasciata_CSIRO1Smul
Pomacanthidae_Holacanthus_passer_BFT11349
Pomacanthidae_Pomacanthus_paru_PRJNA398732
Chaetodontidae_Heniochus_monoceros_USNMAH8AC10
Chaetodontidae_Chaetodon_striatus_UPRFL0392
Chaetodontidae_Prognathodes_aculeatus_USNMAE8WU38
Outgroups
Luvarus_imperialis_CSIRO1Limp
FF_Luvarus_necopinatus
FF_Avitoluvarus_eocaenicus
FF_Avitoluvarus_mariannae
FF_Kushlukia_permira
FF_Avitoluvarus_dianae
FF_Gazolaichthys_vestenanovae
FF_Padovathurus_gaudryi
FF_Massalongius_gazolai
FF_Eozanclus_brevirostris
Zanclus_cornutus_USNMAG5NQ44
FF_Angiolinia_mirabilis
FF_Proacanthurus_elongatus
FF_Protozebrasoma_bloti
FF_Tauichthys_aspesae
FF_Tauichthys_padremenini
FF_Proacanthurus_bonatoi
FF_Proacanthurus_tenuis
FF_Proacanthurus_ovalis
FF_Acanthuroides_massalongoi
FF_Pesciaraichthys_punctatus
FF_Lehmanichthys_lessiniensis
FF_Sorbinithurus_sorbinii
FF_Tylerichthys_milani
FF_Tylerichthys_nuchalis
FF_Metaspisurus_emmanueli
FF_Eorandallius_elegans
FF_Eorandallius_rectifrons
Naso_minor_USNMAG9RE79
Naso_tergus
Naso_thynnoides_USNMAG9RE77
Naso_caeruleacauda_USNMAG9RE78
Naso_maculatus
Naso_lopezi_USNMAB6QS09
Naso_brachycentron_USNMAH0SV59
Naso_reticulatus
Naso_mcdadei_CSIROIN02960
Naso_fageni
Naso_tonganus_USNMAH0SV11
Naso_tuberosus_CSIROIN02880
Naso_caesius
Naso_hexacanthus_USNMAG9RD90
Naso_vlamingii_USNMAG5NP05
Naso_annulatus_USNMAG5NY62
Naso_brevirostris_USNMAG9RF50
Naso_unicornis_USNMAG9RR39
Naso_elegans
Naso_lituratus_USNMAG9RS64
FF_Eonaso_deani
FF_Arambourgthurus_scombrurus
FF_Marosichthys_huismani
Prionurus_microlepidotus
Prionurus_laticlavius_UPRFL1415
FF_Frigosorbiniae_baldwinae
Prionurus_biafraensis
Prionurus_maculatus
Prionurus_chrysurus_CSIROIN02885
Prionurus_scalprum
FF_Glarithurus_friedmani
FF_Caprovesposus_parvus
Paracanthurus_hepatus_USNMAH0SV56
Zebrasoma_velifer_USNMAG9RQ22
Zebrasoma_desjardinii
Zebrasoma_gemmatum
Zebrasoma_xanthurum
Zebrasoma_rostratum_USNMAG9RT04
Zebrasoma_flavescens_USNMAH9BK78
Zebrasoma_scopas_USNMAG5NQ56
Acanthurus_polyzona
Acanthurus_triostegus_USNMAG5NP51
Acanthurus_sohal
Acanthurus_lineatus_USNMAG5NP47
Acanthurus_leucopareius_USNMAG9RQ69
Acanthurus_nigrofuscus_STRIBFT11922
Acanthurus_nigros_USNMAG9RS75
Acanthurus_coeruleus_UPRFL0024
Acanthurus_guttatus_USNMAD9NN70
Acanthurus_achilles_USNMAG9RS25
Acanthurus_nigricans_USNMAG9RS96
Acanthurus_leucosternon
Acanthurus_japonicus
Acanthurus_thompsoni_USNMAG5NQ32
Ctenochaetus_striatus_USNMAG9RP78
Acanthurus_pyroferus_USNMAG5NP30
Acanthurus_nubilus_USNMAG9RS98
Ctenochaetus_tominiensis
Ctenochaetus_hawaiiensis_USNMAG7PV39
Ctenochaetus_marginatus_USNMAG7PX25
Ctenochaetus_binotatus_USNMAC1VG52
Ctenochaetus_flavicauda_USNMAG5NP37
Ctenochaetus_strigosus_USNMAH9BK96
Ctenochaetus_cyanocheilus
Ctenochaetus_truncatus
Acanthurus_monroviae_USNMAD9NG13
Acanthurus_chirurgus_UPRFL0027
Acanthurus_bahianus
Acanthurus_tractus_USNMAC3XT10
Acanthurus_tennentii_CSIROIN01572
Acanthurus_reversus_USNMAG7PW17
Acanthurus_olivaceus_USNMAG5NP42
Acanthurus_nigricauda_STRIBFT11894
Acanthurus_leucocheilus
Acanthurus_auranticavus
Acanthurus_maculiceps_USNMAG9RH18
Acanthurus_albimento_USNMAF0YH71
Acanthurus_blochii_STRIBFT11884
Acanthurus_grammoptilus_CSIROGT10790
Acanthurus_xanthopterus_USNMAG9RF52
Acanthurus_gahhm
Acanthurus_mata_USNMAG9RD41
Acanthurus_bariene_USNM438052
Acanthurus_dussumieri_USNMAG9RH38
Luvaridae
Two fossils with ambiguous placement
Zanclidae
Acanthuridae
0.07

Figure S7. Phylogeny of Acanthuriformes based on a concatenation analysis of the combined matrix comprised by 999 loci, 107 morphological characters and 121 species (32 fossils + 80 extant Acanthuriformes, 9 outgroups). Phylogenetic tree was inferred with RAxML using the MK model for the morphological partition, and the MULTIGAMMA model for the DNA sequences using five by-gene and by-codon partitions. Colors indicate families. Nodal values indicate bootstrap support.

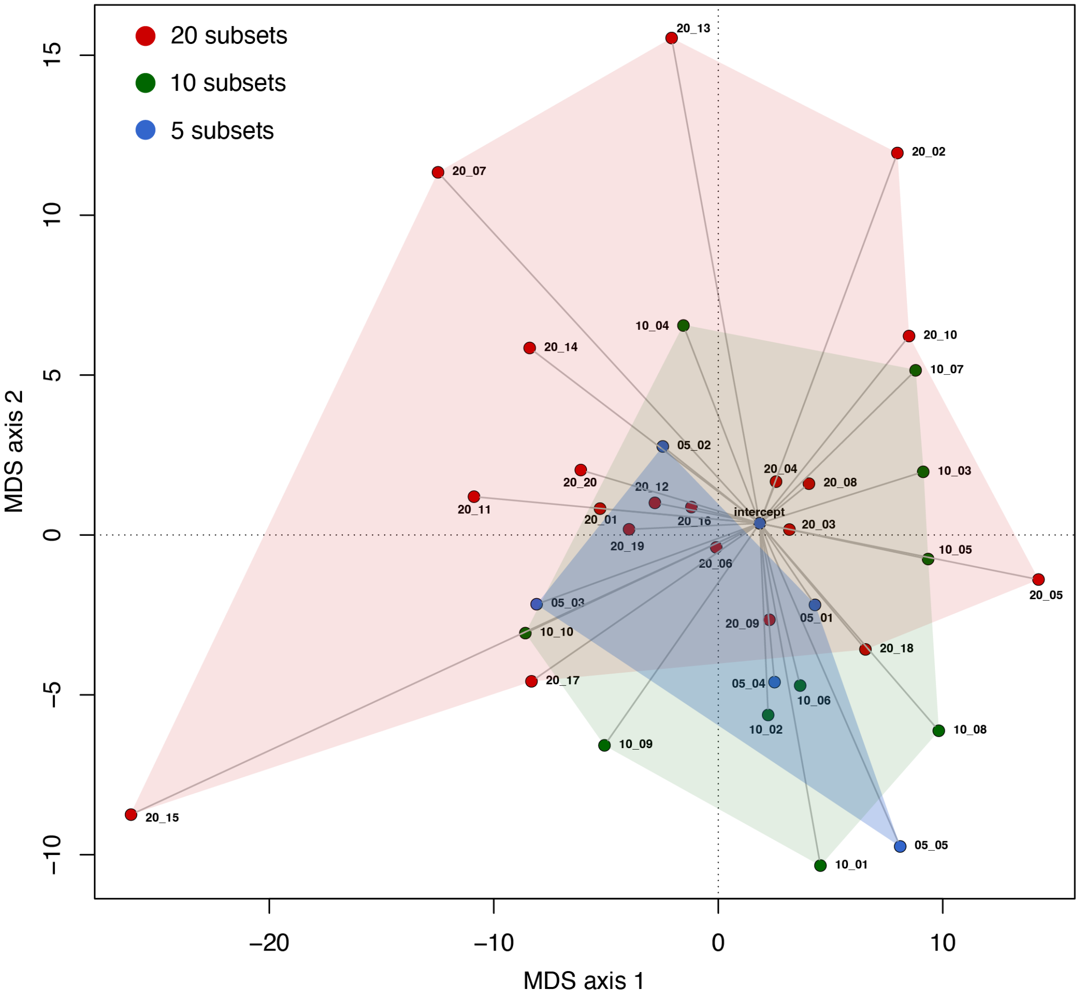


Figure S8. Tree space plot for trees inferred in RAxML based on different subsets of the expanded matrix. This matrix was divided into 20 (50 loci x13 + 49 loci x7), 10 (99 loci x7 + 100 loci x3), and 5 (199 loci x3 + 198 loci x2) subsets. Colors represent the number of subsets. Labels next to the circles indicate the type of dataset and the subset number.

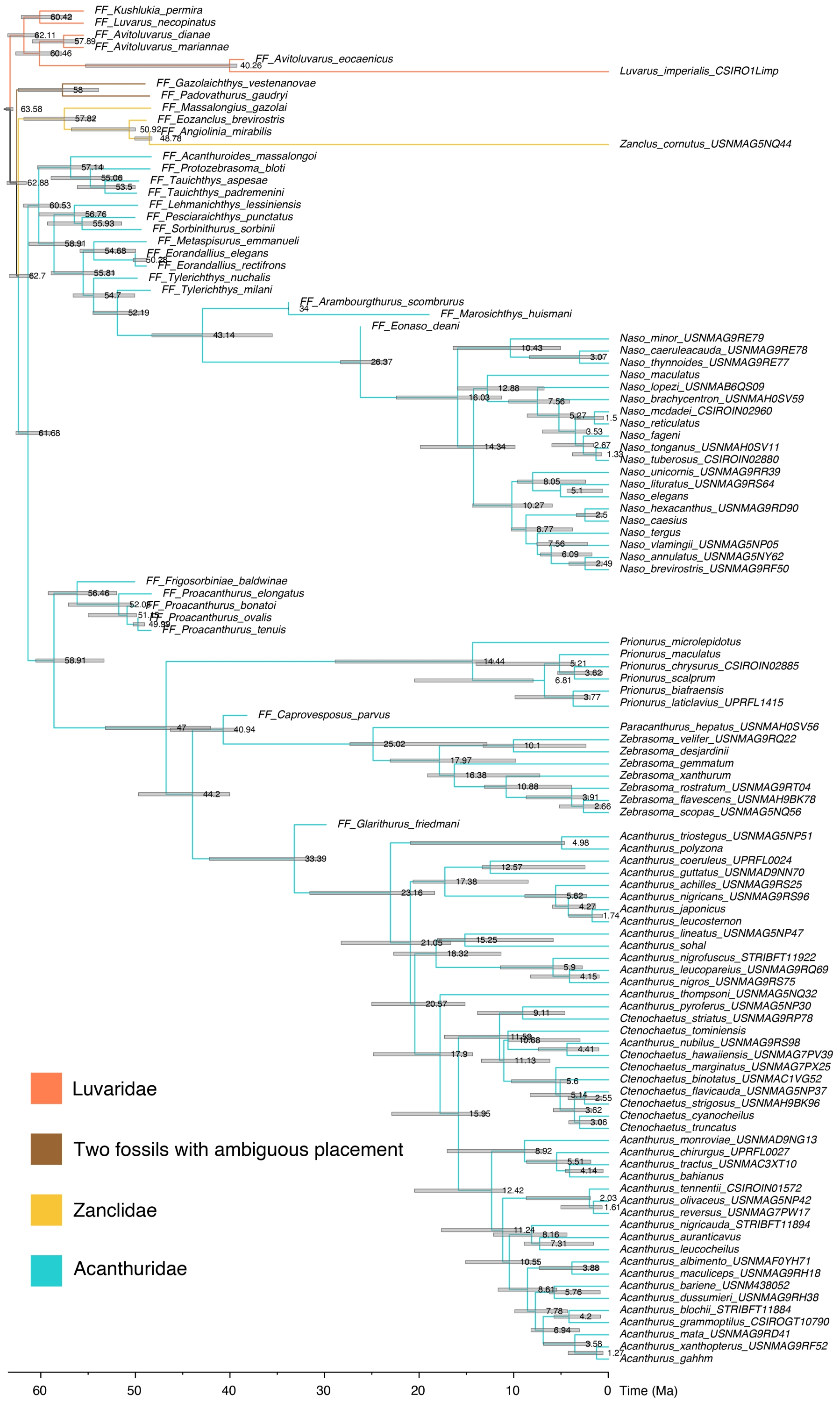


Figure S9. Maximum clade credibility (MCC) tree of Acanthuriformes based on Scheme 1. Time-calibrated phylogenetic tree using a total-evidence framework based on Bayesian inference of 999 exons and 112 species (32 fossil, 80 extant). MCC tree generated from 10,000 trees evenly selected from the posterior distribution of five subsets. These analyses took over a year to reach convergence. Colors indicate families. Posterior probability ages are given for each node. Ma: millions of years.

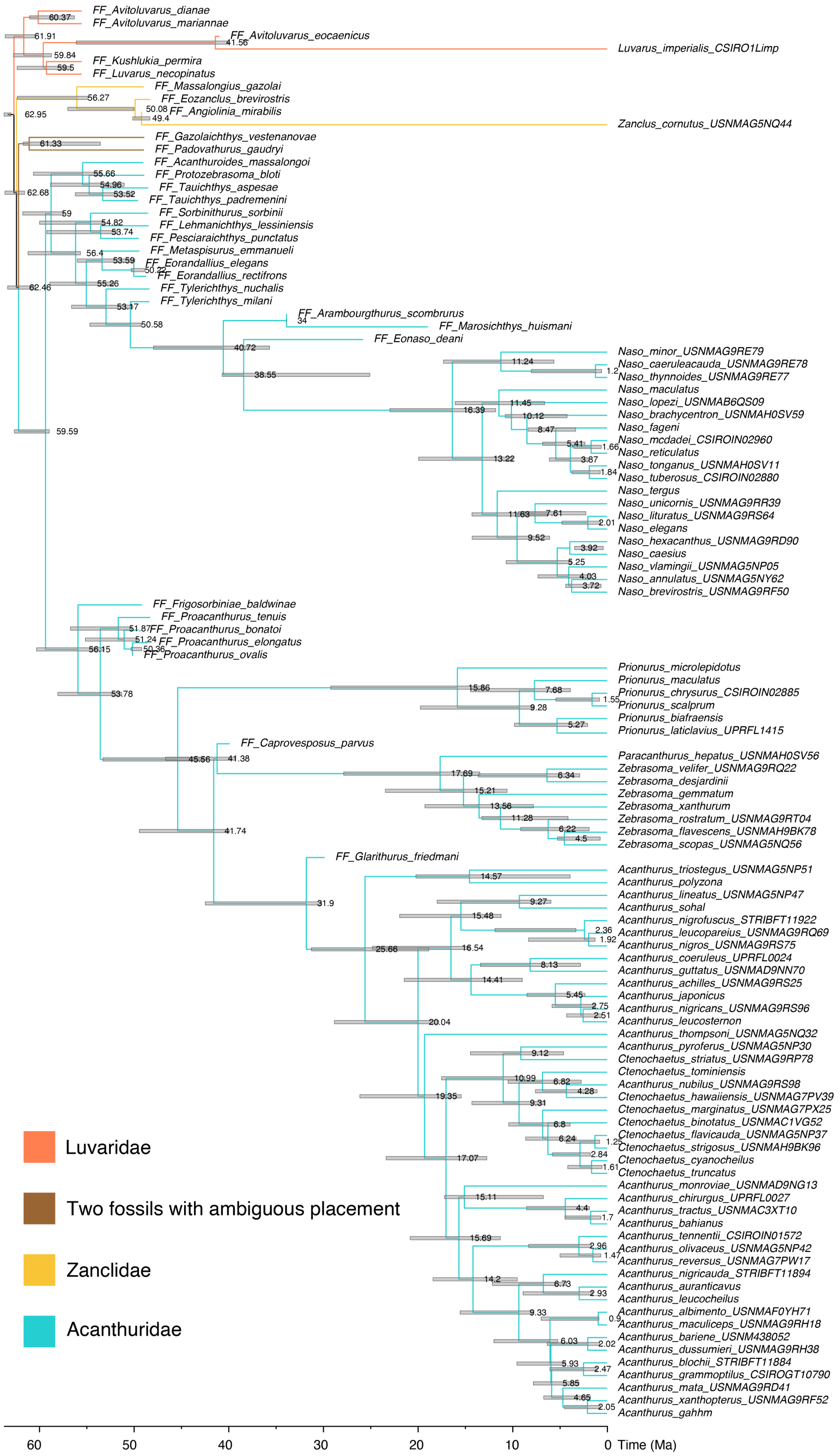


Figure S10. Maximum clade credibility (MCC) tree of Acanthuriformes based on Scheme 2. Time-calibrated phylogenetic tree using a total-evidence framework based on Bayesian inference of 999 exons and 112 species (32 fossil, 80 extant). MCC tree generated from 10,000 trees evenly selected from the posterior distribution of five subsets. These analyses took over a year to reach convergence. Colors indicate families. Posterior probability ages are given for each node. Ma: millions of years.

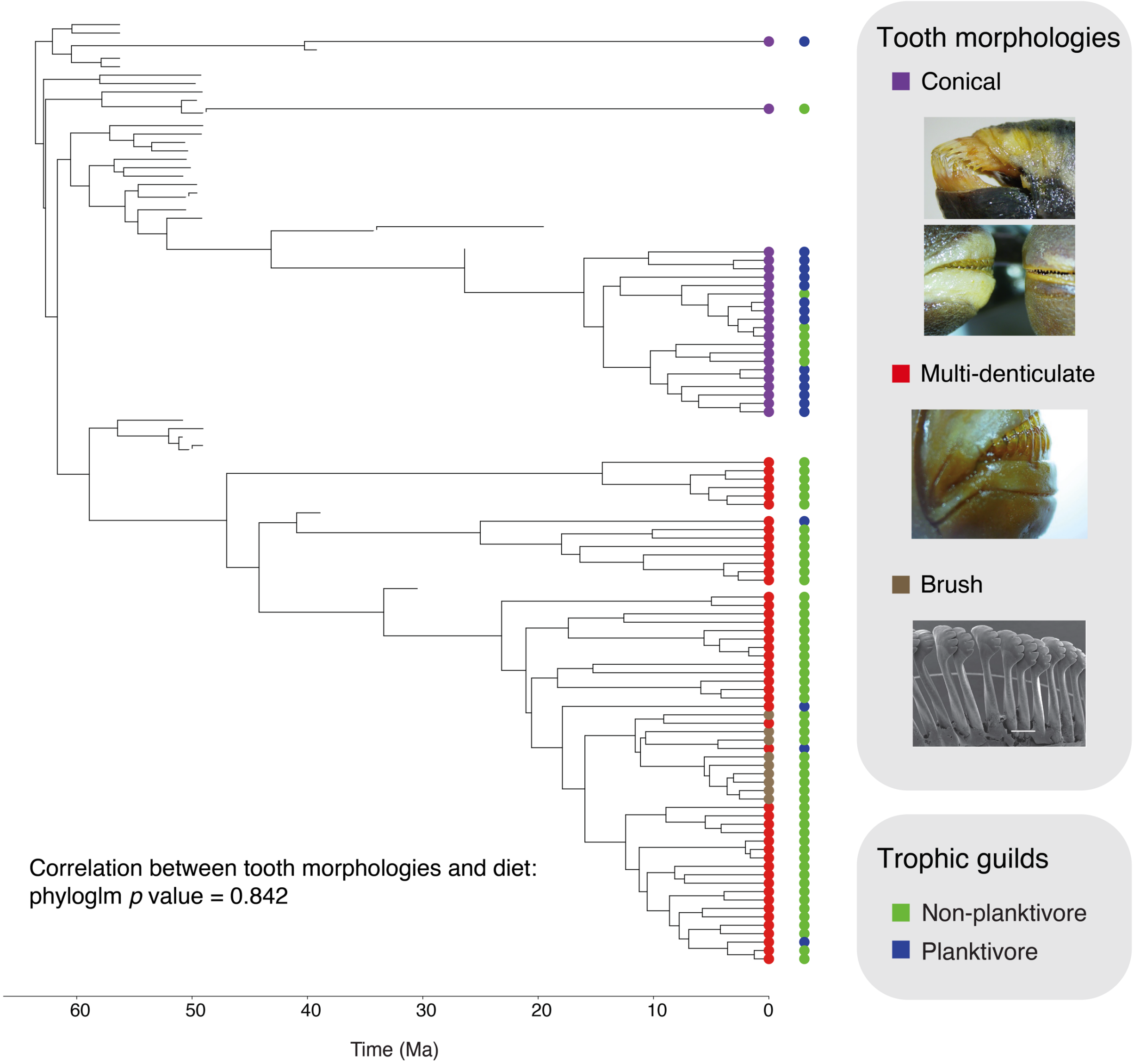


Figure S11. Relationship between tooth morphology and diet in extant species only. The pies are color-coded according to different tooth morphologies and diets. Images depicting teeth are included for each tooth morphology category: conical (*Zanclus* and *Naso*), multi-denticulate (*Acanthurus*), and brush (*Ctenochaetus*). Most photographs were taken by A. Santaquiteria; the SEM image of the *Ctenochaetus* specimen, was sourced from Bellwood et al. (2014). Ma: millions of years.

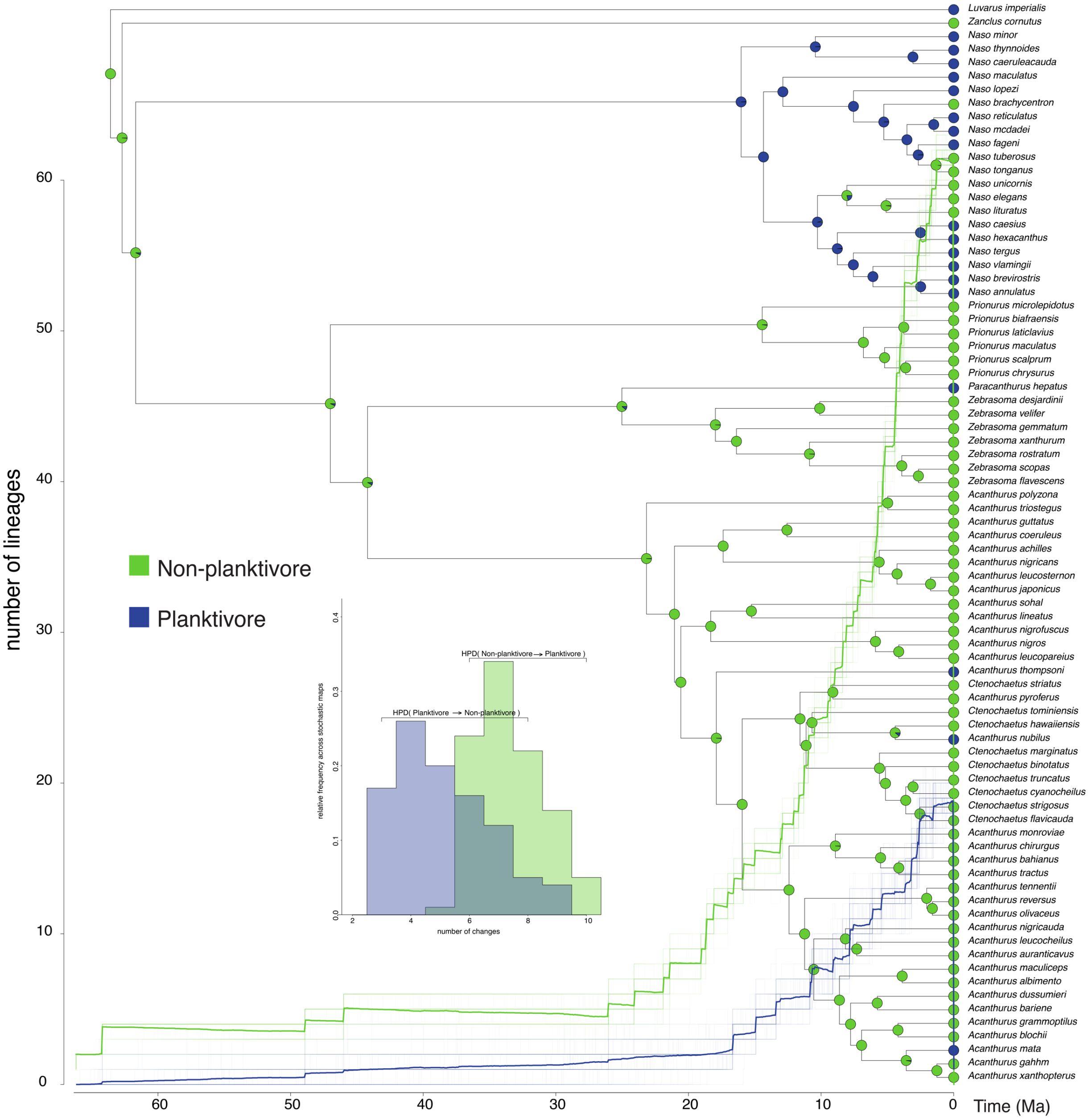


Figure S12. Ancestral diet reconstruction of extant acanthuriforms based on MCC tree of Scheme 1. SIMMAP analyses based on the best-fit model, asymmetric rates. The lineage through time (LTT) plot illustrates the number of non-planktivore and planktivore lineages over time (the line solid depicts the mean values from 100 simulations). Colors indicate each trophic guild. Ma: millions of years.

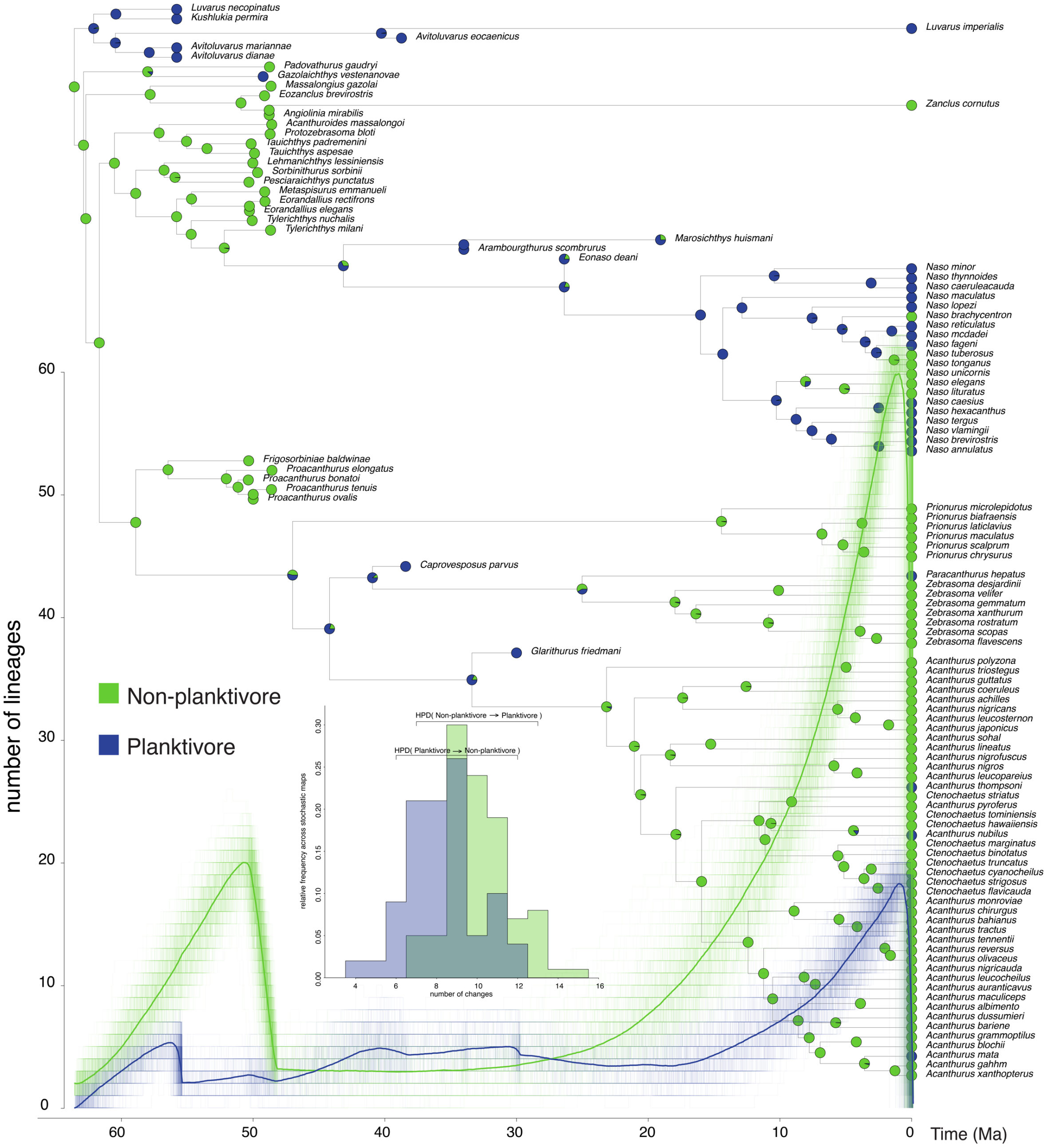


Figure S13. Ancestral diet reconstruction of both extant and fossil acanthuriforms based on MCC tree of Scheme 1. SIMMAP analyses based on the best-fit model, asymmetric rates. The lineage through time (LTT) plot illustrates the number of non-planktivore and planktivore lineages across time (the solid line depicts the mean values from 500 trees). Colors indicate each trophic guild. Ma: millions of years.

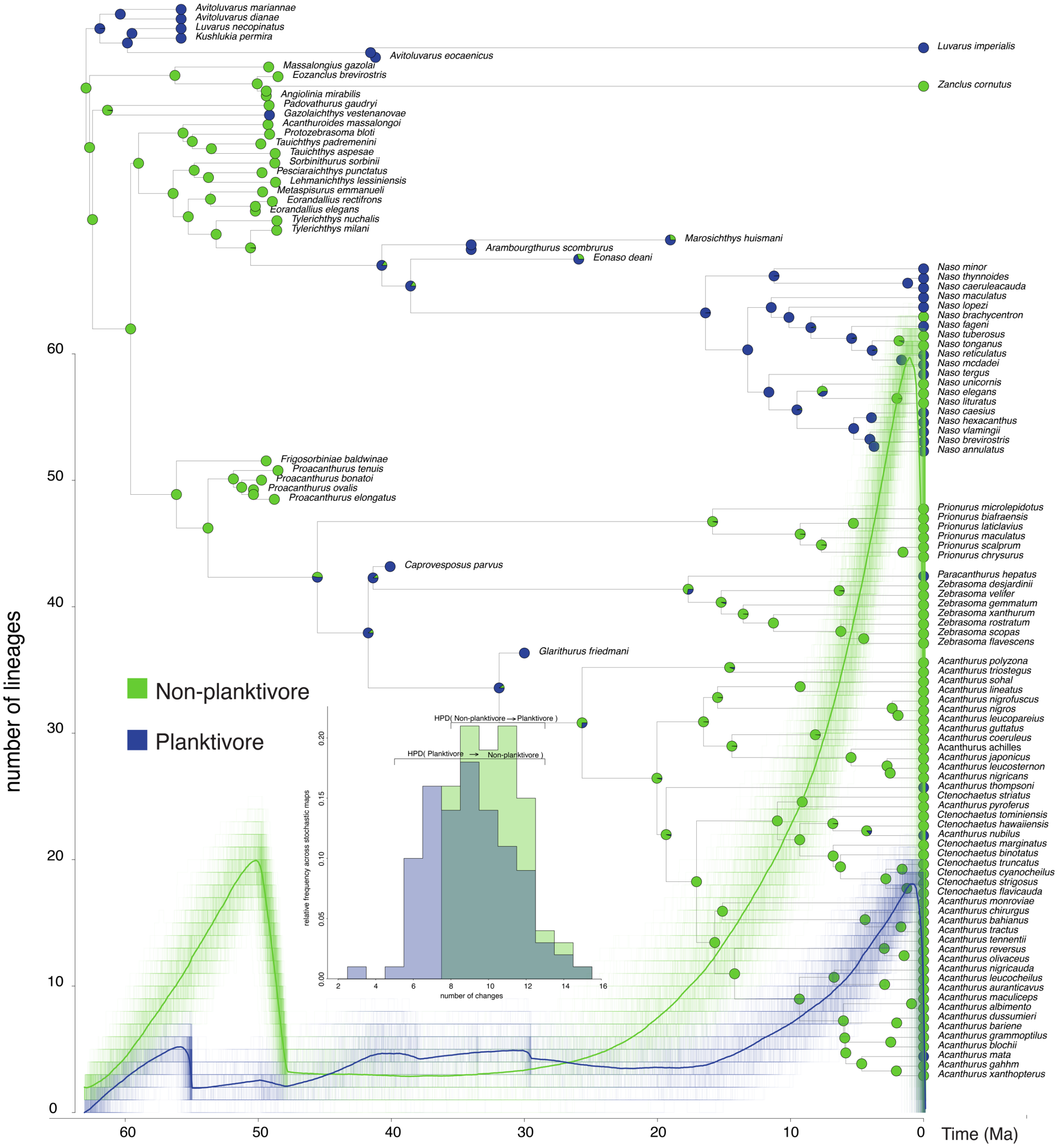


Figure S14. Ancestral diet reconstruction of both extant and fossil acanthuriforms based on MCC tree of Scheme 2. SIMMAP analyses based on the best-fit model, asymmetric rates. The lineage through time (LTT) plot illustrates the number of non-planktivore and planktivore lineages across time (the solid line depicts the mean values from 500 trees). Colors indicate each trophic guild. Ma: millions of years.

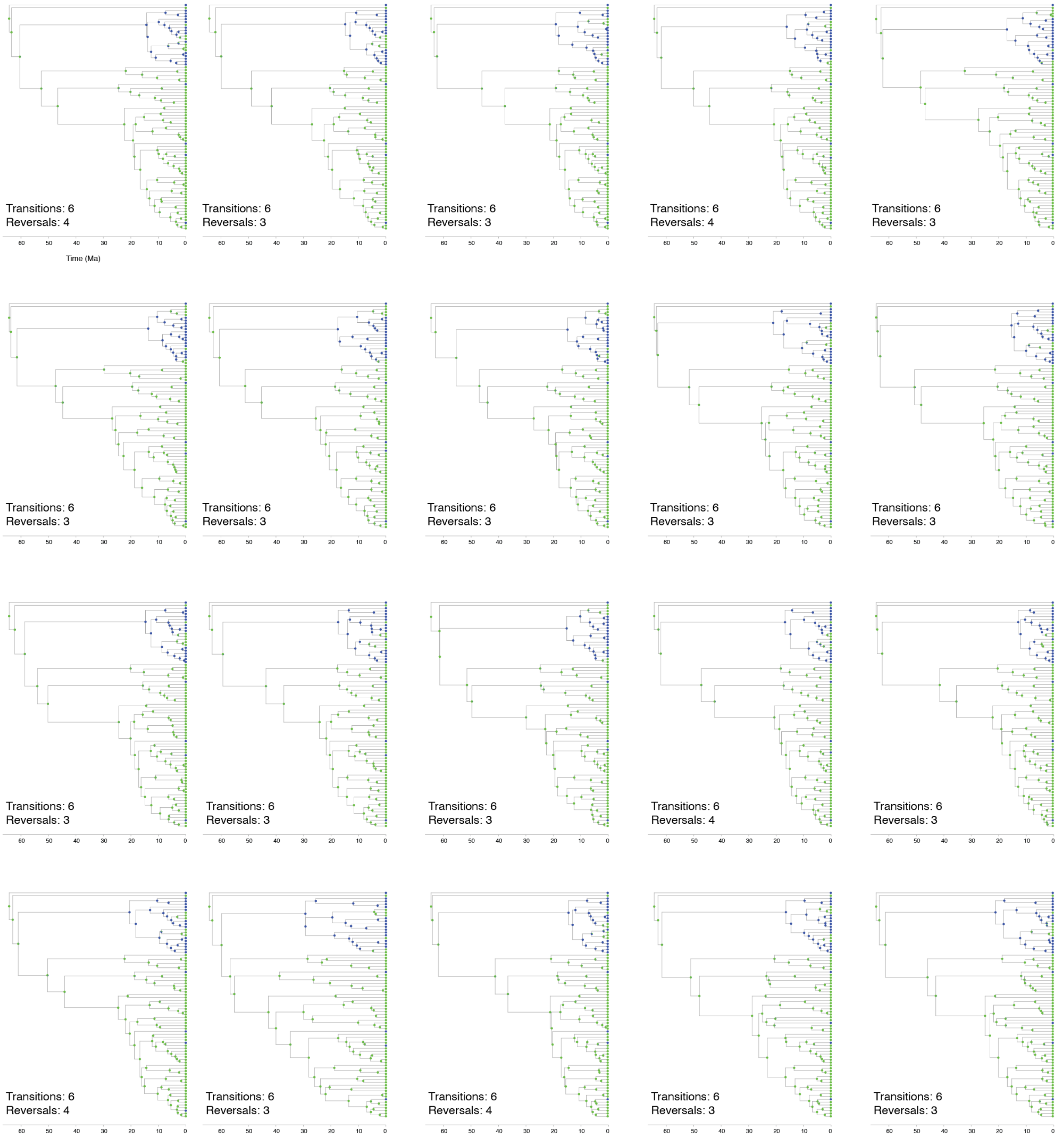


Figure S15. Uncertainty in ancestral diet reconstruction analyses using only extant species. SIMMAP analyses based on the best-fit model, asymmetric rates for a total of 20 trees evenly sampled across the 500 trees from Scheme 1, excluding fossils. Pies are color-coded based on trophic guild, with non-planktivores represented in green and planktivores in blue. At the bottom left of each panel, counts of transitions to planktivory and reversals to non-planktivory are provided. Ma: millions of years.

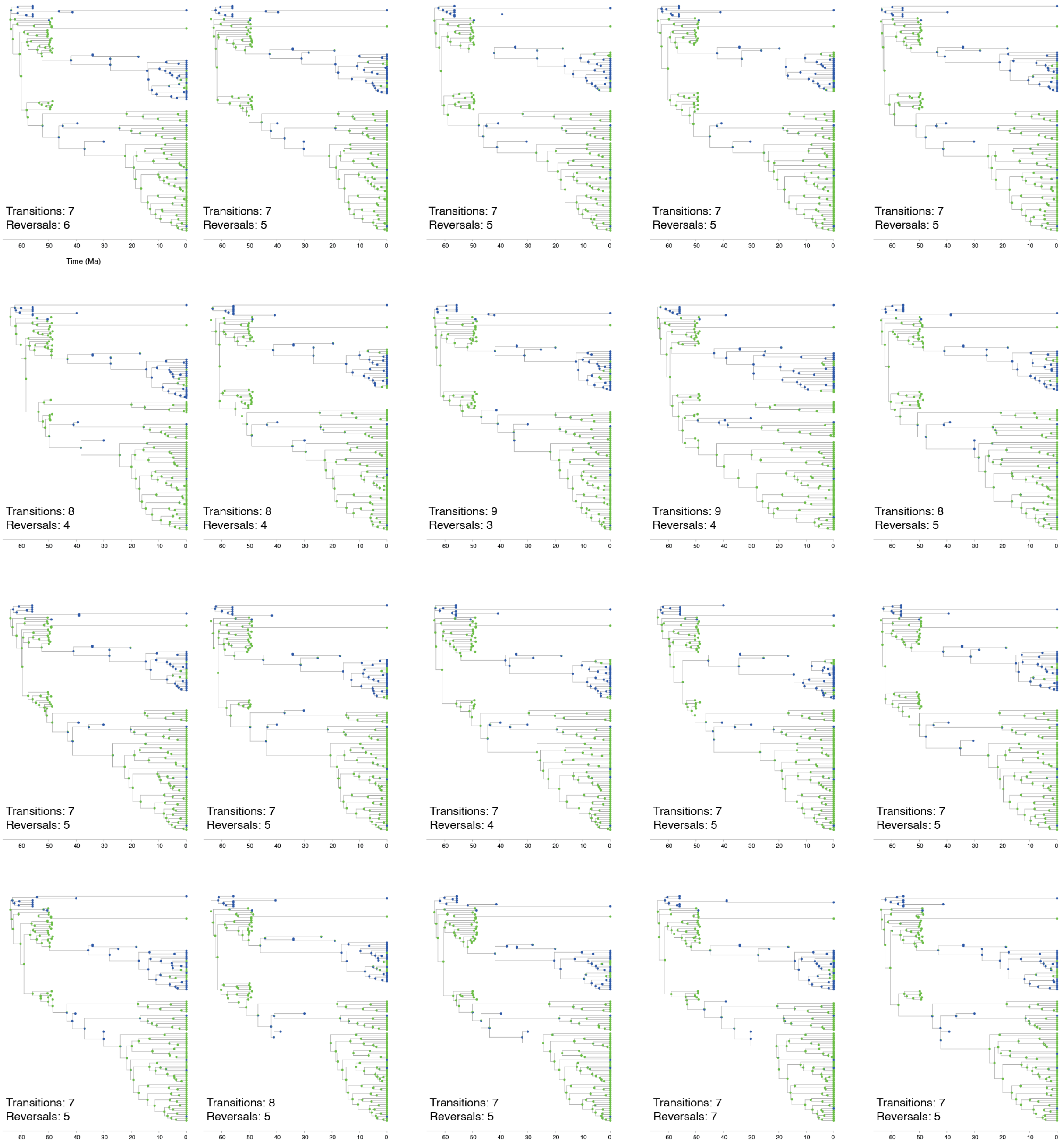


Figure S16. Uncertainty in ancestral diet reconstruction analyses. SIMMAP analyses based on the best-fit model, asymmetric rates for a total of 20 trees evenly sampled across the 500 trees from Scheme 1. Pies are color-coded based on trophic guild, with non-planktivores represented in green and planktivores in blue. At the bottom left of each panel, counts of transitions to planktivory and reversals to non-planktivory are provided. Ma: millions of years.

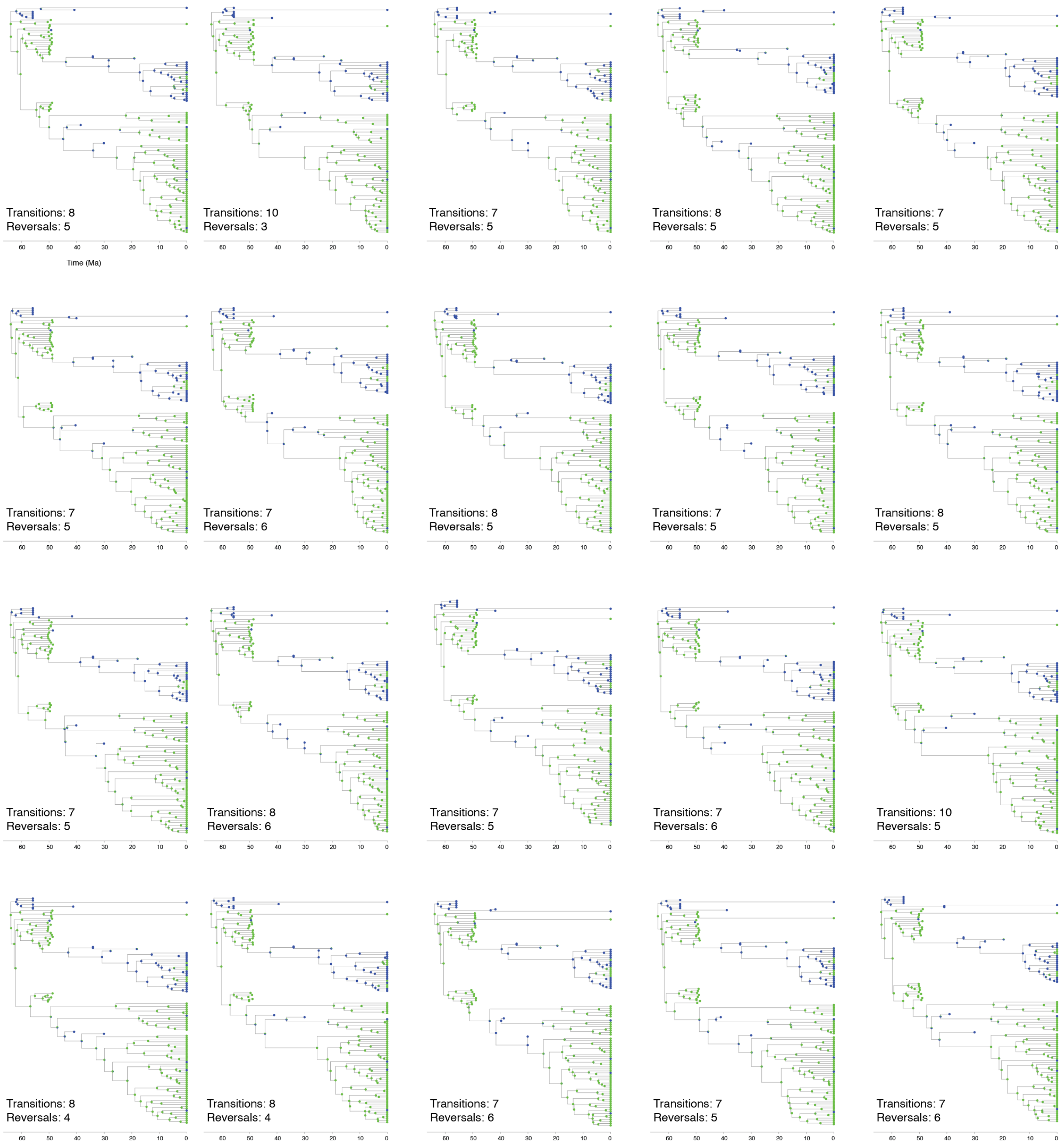


Figure S17. Uncertainty in ancestral diet reconstruction analyses. SIMMAP analyses based on the best-fit model, asymmetric rates for a total of 20 trees evenly sampled across the 500 trees from Scheme 2. Pies are color-coded based on trophic guild, with non-planktivores represented in green and planktivores in blue. At the bottom left of each panel, counts of transitions to planktivory and reversals to non-planktivory are provided. Ma: millions of years.

Table S5. Summary statistics of the 12 biogeographic models implemented in *BioGeoBEARS* based on the MCMC tree of Scheme 1 and 2. The best-fitting model for each scheme is indicated in bold.

		Model	LnL	numparams	d	e	j	w	AICc	AICc weight
Scheme 1	*Tethys closure at 12Ma*	DEC	-304.2	2	0.047	0.011	0	1	612.4	1.10E-07
		DEC+J	-304.2	3	0.047	0.011	1.00E-05	1	614.5	3.80E-08
		DEC+W	-304.9	3	0.037	0.0087	0	0.27	616	1.90E-08
		DEC+J+W	-328.5	4	0.019	0.0074	1.00E-05	0.082	665.4	3.50E-19
		DIVALIKE	-319.4	2	0.052	0.011	0	1	642.9	2.70E-14
		DIVALIKE+J	-319.4	3	0.052	0.011	1.00E-05	1	645	9.30E-15
		DIVALIKE+W	-321.7	3	0.039	0.0083	0	0.23	649.7	9.10E-16
		DIVALIKE+J+W	-315.2	4	0.044	0.0077	1.00E-05	0.4	638.7	2.10E-13
		BAYAREALIKE	-294.7	2	0.023	0.038	0	1	593.4	0.0015
		BAYAREALIKE+J	-291	3	0.021	0.031	0.014	1	588.3	0.02
		BAYAREALIKE+W	**-287.5**	**3**	**0.015**	**0.033**	**0**	**0.25**	**581.3**	**0.64**
		BAYAREALIKE+J+W	-287.1	4	0.0075	0.026	0.0063	0.086	582.6	0.34
	Tethys closure at 16Ma	DEC	-302.7	2	0.049	0.0097	0	1	609.5	1.60E-07
		DEC+J	-302.7	3	0.049	0.0097	1.00E-05	1	611.6	5.50E-08
		DEC+W	-311.8	3	0.027	0.0043	0	0.17	629.8	6.30E-12
		DEC+J+W	-300.7	4	0.038	0.0098	1.00E-05	0.3	609.8	1.30E-07
		DIVALIKE	-318.1	2	0.055	0.0094	0	1	640.4	3.10E-14
		DIVALIKE+J	-318.1	3	0.055	0.0094	1.00E-05	1	642.5	1.10E-14
		DIVALIKE+W	-349.2	3	0.021	0.0033	0	0.056	704.5	3.70E-28
		DIVALIKE+J+W	-312.7	4	0.045	0.0072	1.00E-05	0.37	633.7	8.70E-13
		BAYAREALIKE	-292.8	2	0.025	0.038	0	1	589.7	0.0032
		BAYAREALIKE+J	-289.6	3	0.022	0.032	0.013	1	585.4	0.027
		BAYAREALIKE+W	**-286**	**3**	**0.017**	**0.033**	**0**	**0.3**	**578.3**	**0.94**
		BAYAREALIKE+J+W	-288.5	4	0.0067	0.025	0.0083	0.058	585.5	0.026
Scheme 2	*Tethys closure at 12Ma*	DEC	-305.5	2	0.049	0.0098	0	1	615.1	2.00E-10
		DEC+J	-305.5	3	0.049	0.0098	1.00E-05	1	617.2	6.80E-11
		DEC+W	-306.2	3	0.037	0.0089	0	0.28	618.5	3.50E-11
		DEC+J+W	-306.8	4	0.036	0.0081	1.00E-05	0.27	622.1	6.00E-12
		DIVALIKE	-322.6	2	0.054	0.0098	0	1	649.2	7.60E-18
		DIVALIKE+J	-322.5	3	0.056	0.0095	1.00E-05	1	651.3	2.70E-18
		DIVALIKE+W	-326	3	0.037	0.006	0	0.23	658.3	8.20E-20
		DIVALIKE+J+W	-318.8	4	0.044	0.0064	1.00E-05	0.37	646	3.70E-17
		BAYAREALIKE	-288.1	2	0.025	0.042	0	1	580.3	0.007
		BAYAREALIKE+J	-285.4	3	0.022	0.034	0.012	1	577.1	0.035
		BAYAREALIKE+W	**-282.1**	**3**	**0.018**	**0.036**	**0**	**0.31**	**570.5**	**0.95**
		BAYAREALIKE+J+W	-286.7	4	0.0051	0.025	0.0054	0.036	581.7	0.0034
	Tethys closure at 16Ma	DEC	-303.1	2	0.051	0.0092	0	1	610.3	4.30E-10
		DEC+J	-303.1	3	0.051	0.0092	1.00E-05	1	612.5	1.50E-10
		DEC+W	-298.6	3	0.046	0.0076	0	0.43	603.5	1.30E-08
		DEC+J+W	-298.6	4	0.044	0.0076	1.00E-05	0.43	605.5	4.90E-09
		DIVALIKE	-320.7	2	0.065	0.0094	0	1	645.5	1.00E-17
		DIVALIKE+J	-320.1	3	0.058	0.0081	1.00E-05	1	646.4	6.30E-18
		DIVALIKE+W	-315.5	3	0.051	0.0075	0	0.39	637.2	6.40E-16
		DIVALIKE+J+W	-337.6	4	0.028	0.0063	1.00E-05	0.13	683.6	5.40E-26
		BAYAREALIKE	-286.2	2	0.026	0.042	0	1	576.5	0.0099
		BAYAREALIKE+J	-283.8	3	0.023	0.033	0.013	1	573.8	0.038
		BAYAREALIKE+W	**-280.5**	**3**	**0.019**	**0.036**	**0**	**0.31**	**567.3**	**0.95**
		BAYAREALIKE+J+W	-287	4	0.0048	0.027	0.0057	0.028	582.4	0.0005

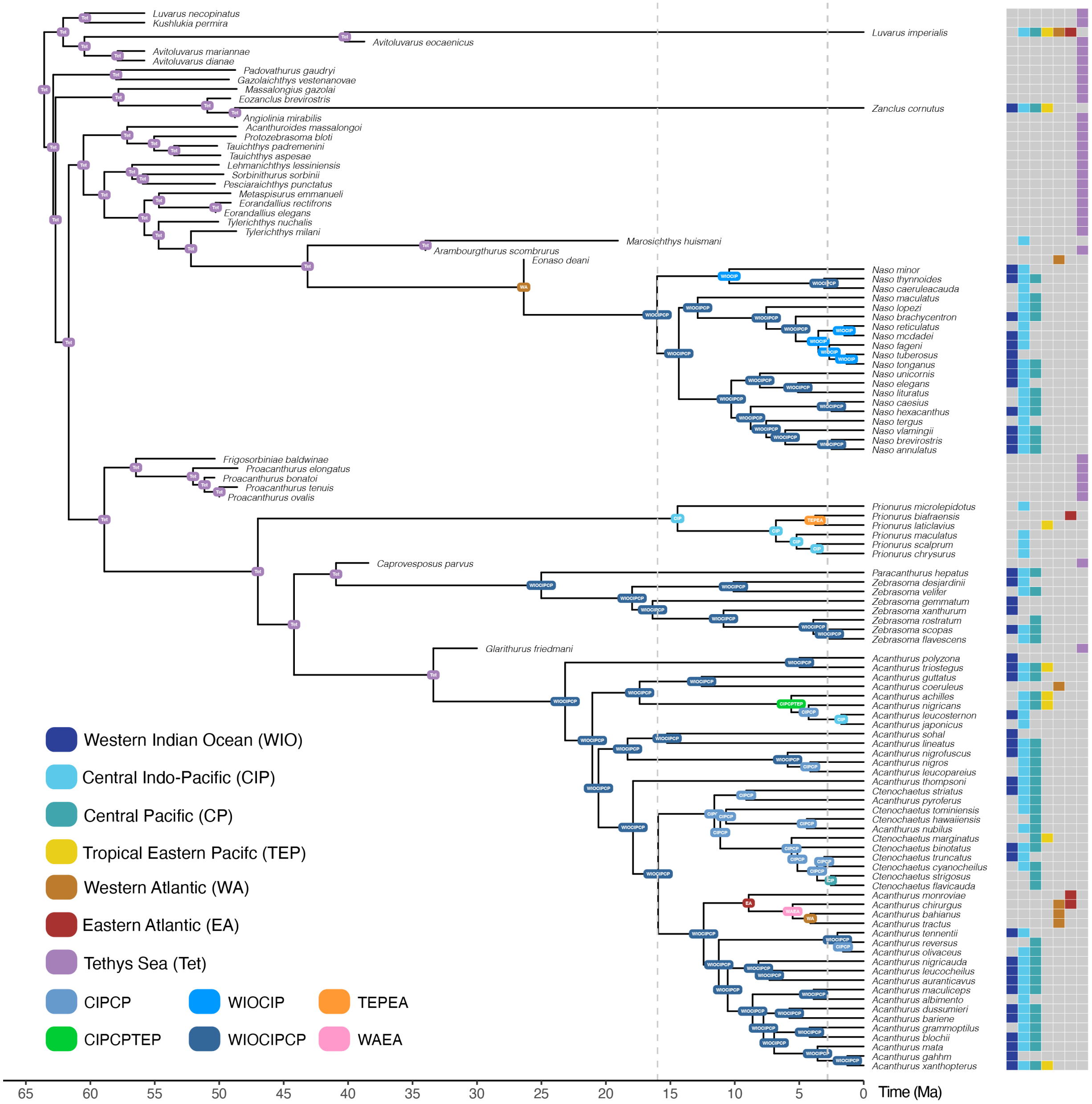


Figure S18. Ancestral range estimations for Acanthuriformes using the best-supported biogeographic model (BAYAREALIKE+w) applied to 20 trees subsampled across the five subsets from Scheme 1 and using the **MCC tree of Scheme 1** as a fixed topology in *BioGeoBEARS*. Note that this analysis assumes that the **final closure of the Tethys Seaway occurred at 16 Ma**. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. Ma: millions of years.

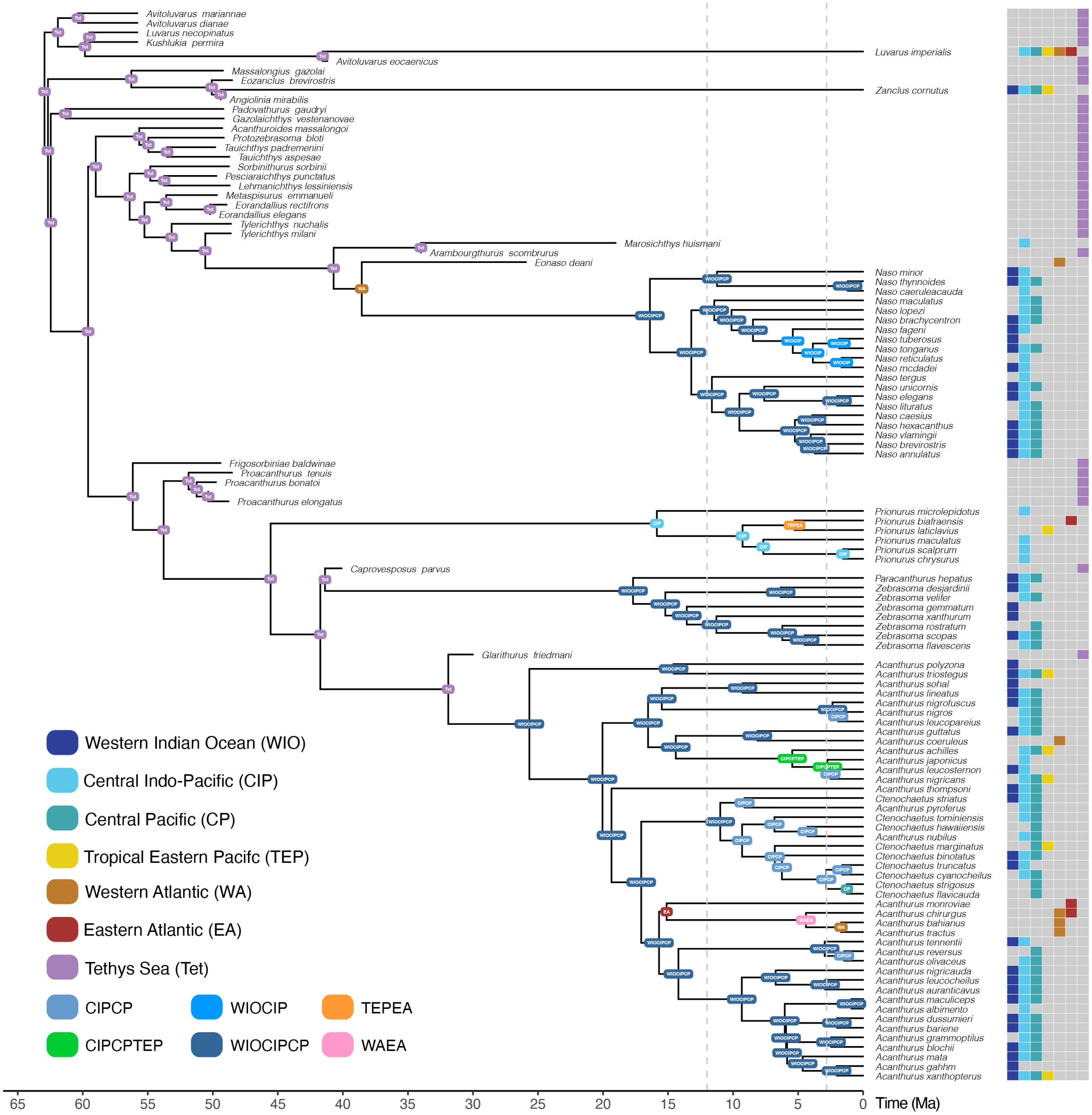


Figure S19. Ancestral range estimations for Acanthuriformes using the best-supported biogeographic model (BAYAREALIKE+w) applied to 20 trees subsampled across the five subsets from Scheme 2 and using the **MCC tree of Scheme 2** as a fixed topology in *BioGeoBEARS*. Note that this analysis assumes that the **final closure of the Tethys Seaway occurred at 12 Ma**. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. Ma: millions of years.

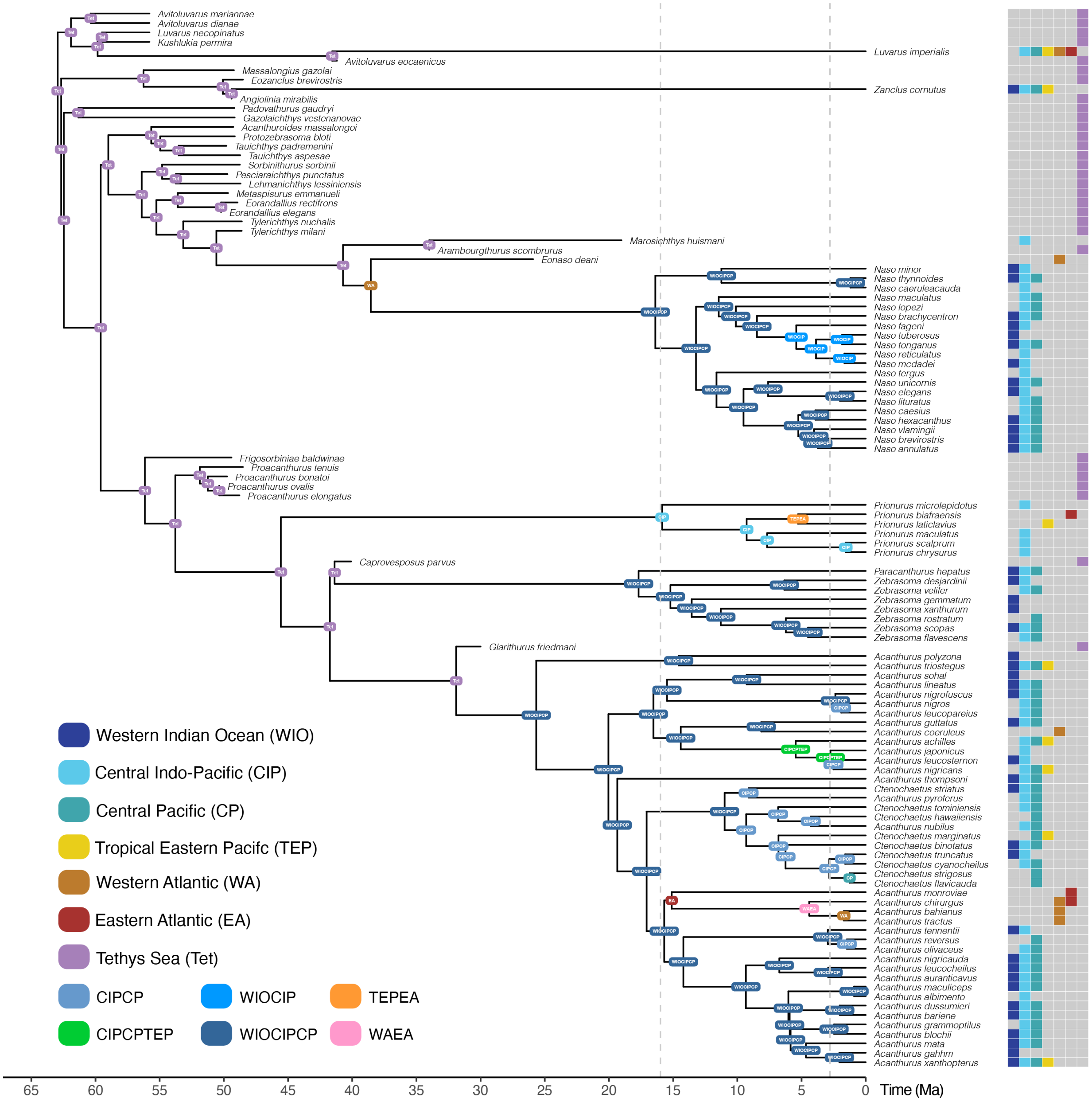


Figure S20. Ancestral range estimations for Acanthuriformes using the best-supported biogeographic model (BAYAREALIKE+w) applied to 20 trees subsampled across the five subsets from Scheme 2 and using the **MCC tree of Scheme 2** as a fixed topology in *BioGeoBEARS*. Note that this analysis assumes that the **final closure of the Tethys Seaway occurred at 16 Ma**. Boxes at nodes and tips are color-coded by area, or areas with the highest ML probability. Ma: millions of years.

Table S6. HiSSE alternative models of lineage diversification and model fitting results trophic guilds based on the MCC tree of each scheme and excluding fossils.

	Model	Number of hidden states	lnL	AIC	AICc	AIC weight
Scheme 1	Null BiSSE	0	-303.033	614.066	614.599	0.015
	Full BiSSE	0	-302.995	617.991	619.141	0.002
	Full HiSSE	2	-293.419	612.838	618.353	0.028
	CID-2	2	-295.954	605.908	607.463	0.910
	CID-4	4	-294.990	611.980	615.862	0.044
Scheme 2	Null BiSSE	0	-298.158	604.316	604.850	0.053
	Full BiSSE	0	-297.905	607.810	608.961	0.009
	Full HiSSE	2	-289.339	604.677	610.192	0.044
	CID-2	2	-292.380	598.760	600.316	0.845
	CID-4	4	-291.212	604.423	608.306	0.050

lnL = log likelihood; AIC = Akaike Information Criterion.

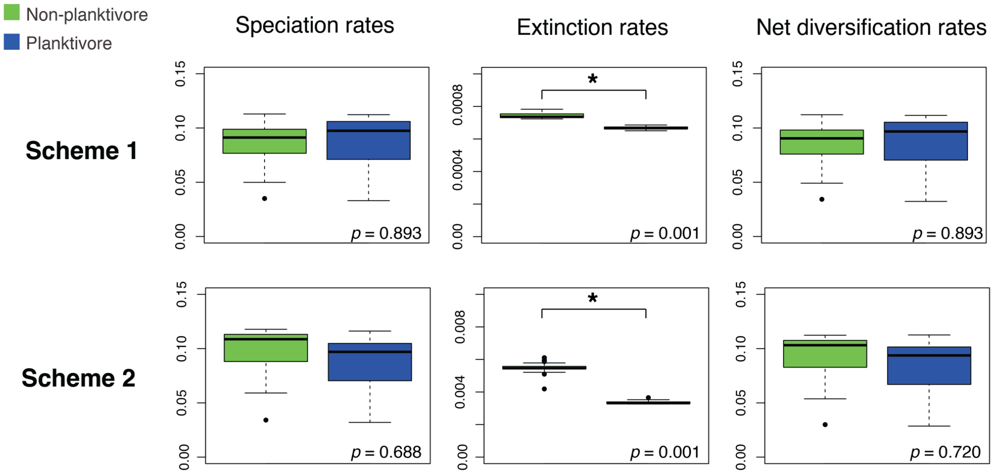


Figure S21. Box plots of model averaged tip rates estimated with HiSSE for speciation, extinction, and net diversification rates in extant non-planktivore and planktivore lineages. Phylogenetic ANOVA significance values are also shown at the bottom right of each panel. First row of panels shows the results obtained based on the MCC tree of Scheme 1 and the second row based on the MCC tree of Scheme 2, both excluding fossils.

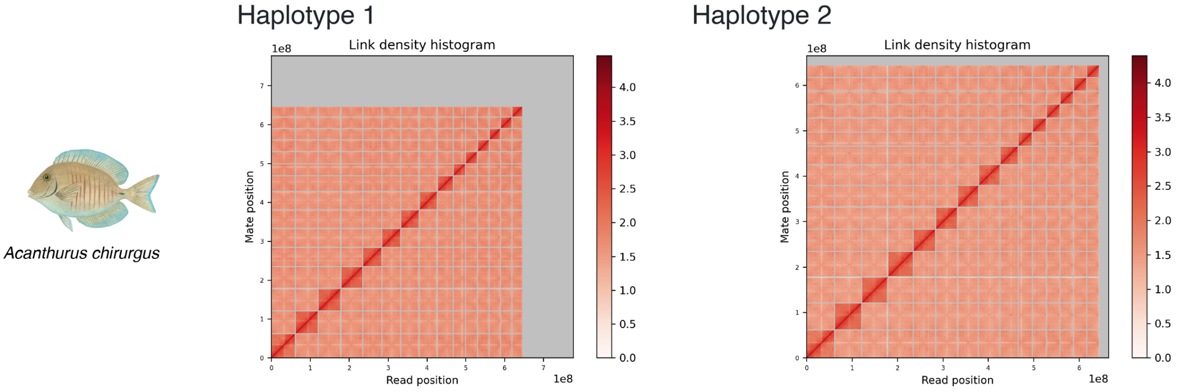
Table S7. HiSSE results across the MCC tree and the 20 trees subsampled across the five subsets from Schemes 1 and 2. Transition rates between the two trophic guilds, non-planktivore and planktivore, are displayed for each tree. Statistical results estimating the relationships between diet and the different tip-associated rates (diversification, speciation, and extinction) are illustrated with p-values calculated for each tree using phylogenetic ANOVA. Significance p-values are highlighted in red.

		Transition rates		p-values between rates and diet		
	Tree	**from non-planktivore to planktivore**	**from planktivore to non-planktivore**	**Diversication rates**	**Speciation rate**	**Extinction rate**
Scheme 1	MCC	0.010	0.026	0.893	0.893	0.001
	1	0.011	0.032	0.954	0.803	0.001
	2	0.010	0.030	0.982	0.929	0.001
	3	0.010	0.025	0.837	0.799	0.001
	4	0.011	0.033	0.949	0.838	0.001
	5	0.010	0.032	0.933	0.910	0.001
	6	0.009	0.033	0.402	0.527	0.424
	7	0.010	0.026	0.900	0.489	0.001
	8	0.010	0.034	0.951	0.972	0.731
	9	0.010	0.026	0.746	0.745	0.757
	10	0.010	0.031	0.995	0.964	0.001
	11	0.010	0.031	0.981	0.996	0.001
	12	0.009	0.023	0.703	0.700	0.001
	13	0.009	0.032	0.767	0.856	0.602
	14	0.011	0.032	0.886	0.832	0.001
	15	0.010	0.028	0.933	0.906	0.530
	16	0.010	0.027	0.894	0.884	0.001
	17	0.007	0.013	0.001	0.001	0.001
	18	0.011	0.034	0.983	0.887	0.001
	19	0.009	0.029	0.878	0.926	0.893
	20	0.010	0.028	0.963	0.962	0.895
Scheme 2	MCC	0.010	0.026	0.772	0.688	0.001
	1	0.011	0.030	0.808	0.836	0.001
	2	0.010	0.022	0.878	0.885	0.001
	3	0.009	0.025	0.927	0.916	0.898
	4	0.009	0.032	0.993	0.809	0.158
	5	0.009	0.022	0.828	0.833	0.001
	6	0.009	0.022	0.893	0.800	0.001
	7	0.011	0.024	0.757	0.707	0.052
	8	0.011	0.030	0.925	0.928	0.968
	9	0.008	0.018	0.704	0.683	0.001
	10	0.009	0.022	0.848	0.857	0.613
	11	0.009	0.024	0.903	0.778	0.518
	12	0.011	0.028	0.890	0.821	0.006
	13	0.009	0.022	0.733	0.714	0.001
	14	0.010	0.029	0.897	0.832	0.001
	15	0.010	0.035	0.984	0.988	0.001
	16	0.008	0.020	0.800	0.741	0.001
	17	0.009	0.020	0.706	0.647	0.001
	18	0.010	0.030	0.874	0.955	0.001
	19	0.009	0.025	0.992	0.913	0.001
	20	0.011	0.037	0.955	0.958	0.487

Table S8. Climatic independent and dependent model results. Statistical results (AIC and AIC weights) from the correlation between diet and the global average and tropical paleo-climatic curves. Analyses were conducted for the MCC tree and for the 500 posterior trees from both schemes. The best-fitting model for each scheme is indicated in bold.

		Temperature Curve	Clim fit AIC	BM fit AIC	EB fit AIC	Lambda fit AIC	sigma	beta	Clim fit AICw	BM fit AICw	EB Fit AICw	Lambda Fit AICw
Scheme 1	*MCC tree*	Scotese Global Avg.	222.948	231.953	233.953	222.486	1.000	2.987	**0.440**	0.005	0.002	**0.554**
		Scotese Tropical	222.590	231.953	233.953	222.486	1.000	2.940	**0.484**	0.004	0.002	**0.510**
	500 trees	Scotese Global Avg.	257.766	263.505	265.503	256.533	1.000	-0.094	**0.340**	0.019	0.007	**0.633**
		Scotese Tropical	259.961	263.505	265.503	256.533	1.000	-0.094	0.147	0.025	0.009	**0.818**
Scheme 2	*MCC tree*	Scotese Global Avg.	223.334	233.116	235.116	222.495	1.000	0.000	**0.395**	0.003	0.001	**0.601**
		Scotese Tropical	224.920	233.116	235.116	222.495	1.000	0.000	0.228	0.004	0.001	**0.767**
	500 trees	Scotese Global Avg.	256.964	262.390	264.389	255.716	1.000	-0.088	**0.339**	0.022	0.008	**0.630**
		Scotese Tropical	258.997	262.390	264.389	255.716	1.000	-0.088	0.156	0.029	0.011	**0.805**

BM = Brownian Motion; EB = Early Burst; Clim= Climatic model; AIC = Akaike Information Criterion; AICw = AIC weight.



Assembly	**Total Length (bp)**	**N50**	**L50**	**N90**	**L90**	**Total Length (bp)**	**N50**	**L50**	**N90**	**L90**
Input Assembly	776,875,734	23,491,650	12	35,706	1,224	664,955,783	8,334,248	10	8,334,248	25
Dovetail HiRise Assembly	776,879,634	47,485,454	8	35,706	1,185	664,959,085	29,137,030	7	29,137,030	14

Figure S22. Overview results from each haplotype of the *Acanthurus chirurgus* chromosome-level genome. Link-density histograms visualize how scaffolds interact with each other across the entire genome for each haplotype. The table represents the scaffold genome assembly statistics for each haplotype from PacBio and Hi-C sequencing. The Dovetail HiRise assembly resulted in 15 putative chromosomes.

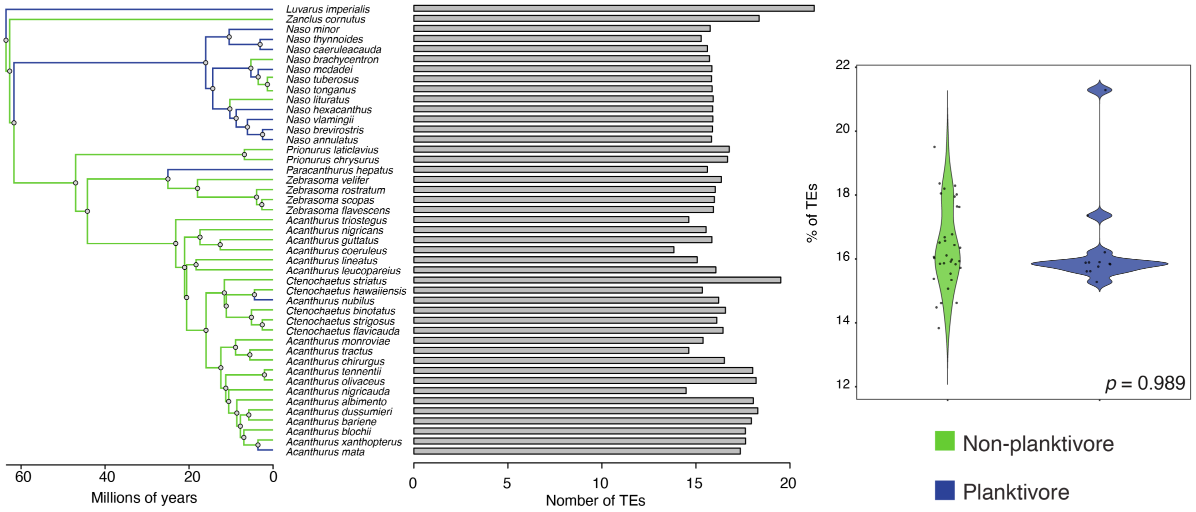


Figure S23. Correlation between the number of transportable elements (TEs) and diet. The phylogenetic tree illustrates the ancestral diet reconstruction for species with whole-genome data. Horizontal gray bars represent the number of TEs per species. Violin plots depict the distribution of the % of TEs across non-planktivore and planktivore lineages, with dots representing the raw scores for each species. Phylogenetic ANOVA significance value is shown at the bottom right of the violin plot.

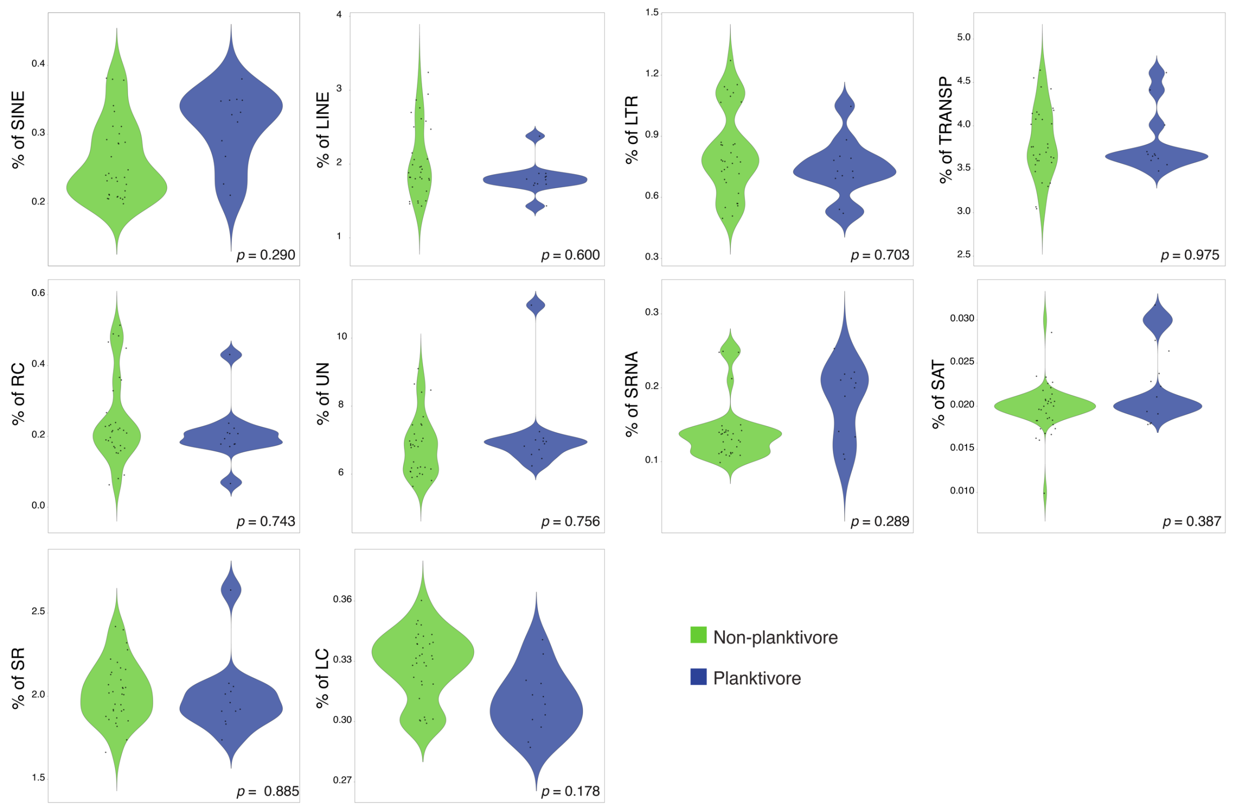


Figure S24. Correlation between the the different repetitive element categories with each trophic guild. Violin plots depict the distribution of the % of each category across non-planktivore and planktivore lineages, with dots representing the raw scores for each species. Phylogenetic ANOVA significance values are also shown at the bottom right of each panel.

Table S9. Convergently evolved and unique for each trophic guild positively selected BUSCO genes. For each gene, the name, biological functions, and molecular functions with its gene ontology (GO) terms are given.

Method		**Busco ID**	**Gene name, biological (B) and molecular (M) functions**
aBSREL	Convergently evolved in the five planktivore transition	1258at7898	**NOTCH2**; neurogenic locus notch homolog protein 2-like B: development process (skeletal development: dorso-ventral axis formation) M: calcium (GO:0005509) and metal (GO:0046872) ion binding
		1312at7898	**PTPN13**; protein tyrosine phosphatase, non-receptor type 13 B: metabolic process and cellular process M: phosphatidylinositol 3-kinase binding (GO:0043548); protein tyrosine phosphatase activity (GO:0004725)
		21851at7898	**CMTR1**; cap-specific mRNA (nucleoside-2-O-)-methyltransferase 1 B: metabolic process and cellular process M: RNA methyltransferase activity (GO:0008173); O-methyltransferase activity (GO:0008171); S-adenosylmethionine-dependent methyltransferase activity (GO:0008757)
		27418at7898	**TPCN1**; two pore segment channel 1 B: localization (transport) M: ion channel activity (GO:0005216)
		408at7898	**TRRAP**; transformation/transcription domain-associated protein B: metabolic process, biological process, and response to stimulus M: protein serine/threonine kinase activity (GO:0004674)
		423at7898	**DYNC2H1**; cytoplasmic dynein 2 heavy chain 1 B: cellular process M: cytoskeletal (minus-end-directed microtubule) motor activity, atp-dependent activity (GO:0008569); dynein intermediate chain binding (GO:0045505)
		892at7898	**UNC80**; protein unc-80 homolog B: homeostatic process M: monoatomic cation channel activity (GO:0005261).
		9220at7898	**DUF2428**; DUF2428 domain in thyroid adenoma-associated proteins (THADA) homolog isoform X1 B: metabolic process and cellular process M: unknown
BUSTED-PH	Unique for planktivore species	124972at7898	**ERG28**; ergosterol biosynthesis 28 homolog B: metabolic process M: protein-macromolecule adaptor activity (GO:0030674)
		76563at7898	**EGR2**; early growth response 2 B: biological regulation M: RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978); DNA-binding transcription factor activity, RNA polymerase II-specific (GO:0000981)
		85468at7898	**SLC25A33**; solute carrier family 25 member 33 B: cellular process (transmembrane transport) and localization M: nucleobase-containing compound transmembrane transporter activity (GO:0015932)
	Unique for non-planktivore species	89467at7898	**KsgA/Erm**; ribosomal RNA adenine methyltransferase KsgA/Erm B: metabolic process and cellular process M: N-methyltransferase activity (GO:0008170); rRNA (adenine) methyltransferase activity (GO:0016433); transcription regulator activity (GO:0140110)

Supplementary References

Agassiz L. 1835. Kritische revision der Ittiolitologia Ve- ronese abgebildenten fossilen Fische. Neues Jahrb. für Mineral. Geognosie, Geol. und Petrefaktenkunde,. 1835:290–316.

Agassiz L. 1838. Recherches Sur Les Poissons Fossiles. Tome IV (10me, 12me livraison).:1–16.

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Bannikov A.F., Tyler J.C. 1995. Phylogenetic revision fo the fish families Luvaridae and Kushlukidae (Acanthuroidei), with a new genus and two new species of Eocene luvarids. Smithson. Contrib. to Paleobiol. 81:1–45.

Bannikov A.F., Tyler J.C. 2001. A New Species of the Luvarid Fish Genus Avitoluvarus (Acanthuroidei, Perciformes) from the Eocene of the Caucasus in Southwest Russia. Proc. Biol. Soc. Washingt. 114:579–588.

Bao Z., Eddy S.R. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 12:1269–1276.

Bateman A. 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47:D506–D515.

de Beaufort L.F. 1926. On a collection of marine fishes from the Miocene of South Celebes. Geol. Onderz. den Oostelijken Oost-Indischen Arch.:117–148.

Bellwood D.R., Hoey A.S., Bellwood O., Goatley C.H.R. 2014. Evolution of long-toothed fishes and the changing nature of fish-benthos interactions on coral reefs. Nat. Commun. 5:1–6.

Betancur-R. R., Broughton R.E., Wiley E.O., Carpenter K., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton II J.C., Zhang F., Buser T., Campbell M., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Ortí G. 2013. The Tree of Life and a New Classification of Bony Fishes. PLoS Curr.:0–45.

Betancur-R. R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17.

Blot J., Tyler J.C. 1990. New genera and species of fossil surgeon fishes and their relatives (Acanthuroidei, Teleostei) from the Eocene of Monte Bolca, Italy, with application of the Blot formula to both fossil and recent forms. Stud. e Ric. sui giacimenti terziari di Bolca. 6:13–92.

Borowiec M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016.

Broughton R.E., Betancur-R. R., Li C., Arratia G., Ortí G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr.

Carnevale G., Bannikov A.F., Marramà G., Tyler J.C., Zorzin R. 2014. The Pesciara-Monte Postale Fossil-Lagerstätte: 2. Fishes and other vertebrates. Rend. della Soc. Paleontol. Ital. 4:37–63.

Carnevale G., Tyler J.C. 2018. The caudal skeleton of Arambourgthurus scombrurus (Arambourg, 1967), a Paleogene oceanic surgeonfish. Proc. Biol. Soc. Washingt. 131:101–110.

Carnevale G., Tyler J.C. 2024. A new moorish idol (teleostei, Zanclidae) from the eocene of bolca, italy. Riv. Ital. di Paleontol. e Stratigr. 130:19–33.

Chan P.P., Lowe T.M. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol. Biol. 1962:1–29.

Cheng H., Jarvis E.D., Fedrigo O., Koepfli K., Urban L., Gemmell N.J., Li H. 2022. Haplotype-resolved assembly of diploid genomes without parental data. Nat. Biotechnol.:1–13.

Danil'chenko P.G. 1960. Bony fishes of the Maikop deposits of the Caucasus. Tr. Paleontol. Instituta. 78:1–247.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol.

Flynn J.M., Hubley R., Goubert C., Rosen J., Clark A.G., Feschotte C., Smit A.F. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117:9451–9457.

Ghezelayagh A., Harrington R.C., Burress E.D., Campbell M.A., Buckner J.C., Chakrabarty P., Glass J.R., Mccraney W.T., Unmack P.J., Thacker C.E., Alfaro M.E., Friedman S.T., Ludt W.B., Cowman P.F., Friedman M., Price S.A., Dornburg A., Faircloth B.C., Wainwright P.C., Near T.J. 2022. Prolonged morphological expansion of spiny-rayed fishes following the end-Cretaceous. Nat. Ecol. Evol. 6:1211–1220.

Goloboff P.A., Catalano S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. 32:221–238.

Guiasu R.C., Winterbottom R. 1993. Osteological Evidence for the Phylogeny of Recent Genera of Surgeonfishes. merican Soc. Ichthyol. Herpetol. 1993:300–312.

Gurevich A., Saveliev V., Vyahhi N., Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.

Holt C., Yandell M. 2011. MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12.

Hughes L.C., Ortí G., Huang Y., Sun Y., Baldwin C.C., Thompson A.W., Arcila D., Betancur-R. R., Li C., Becker L., Bellora N., Zhao X., Li X., Wang M., Fang C., Xie B., Zhou Z., Huang H., Chen S., Venkatesh B., Shi Q. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. 115:6249–6254.

Hughes L.C., Ortí G., Saad H., Li C., White W.T., Baldwin C.C., Crandall K.A., Arcila D., Betancur-R R. 2020. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. Mol. Ecol. Resour. 21:816–833.

Hussakof L. 1907. Zebrasoma deani, a fossil surgeon-fish from the West Indies. Bull. Am. Museum Nat. Hist. 23:125–126.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics.

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

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:772–773.

Li C., Ortí G., Zhang G., Lu G. 2007. A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7:1–11.

Lieberman-aiden E., Berkum N.L. Van, Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. 33292:289–294.

Ludt W.B., Bernal M.A., Kenworthy E., Salas E., Chakrabarty P. 2019. Genomic, ecological, and morphological approaches to investigating species limits: A case study in modern taxonomy from Tropical Eastern Pacific surgeonfishes. Ecol. Evol. 9:4001–4012.

Minh B.Q., Hahn M.W., Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37:2727–2733.

Mok H.-K. 1977. Gut Patterns of the Acanthuridae and Zanclidae. Japan J. Ichthyol. 23:215–219.

Near T.J., Dornburg A., Eytan R.I., Keck B.P., Smith W.L., Kuhn K.L., Moore J.A., Price S.A., Burbrink F.T., Friedman M., Wainwright P.C. 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci.

Near T.J., Eytan R.I., Dornburg A., Kuhn K.L., Moore J.A., Davis M.P., Wainwright P.C., Friedman M., Smith W.L. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci.

Price A.L., Jones N.C., Pevzner P.A. 2005. De novo identification of repeat families in large genomes. Bioinformatics. 21:i351–i358.

Price M.N., Dehal P.S., Arkin A.P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One.

Putnam N.H., Connell B.O., Stites J.C., Rice B.J., Hartley P.D., Sugnet C.W., Haussler D., Rokhsar D.S. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26:342–350.

Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol.

Randall J.. 1955. A contribution to the biology of the Acanthuridae (surgeon fishes). .

Ranwez V., Douzery E.J.P., Cambon C., Chantret N., Delsuc F. 2018. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E. V., Zdobnov E.M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212.

Siqueira A.C., Bellwood D.R., Cowman P.F. 2019. Historical biogeography of herbivorous coral reef fishes: The formation of an Atlantic fauna. J. Biogeogr. 46:1611–1624.

Smit A.F.A., Hubley R., Green P. RepeatMasker Open-4.0. http://www.repeatmasker.org. .

Smith J.L.. 1966. Fishes of the sub-family Nasinae with a synopsis of the Prionurinae. Ichthyol Bull (Rhodes Univ. Grahamstown). 32::634–682.

Sorbini L., Tyler J.C. 1998. A new genus and species of Eocene surgeon fish (Acanthuridae) from Monte Bolca, Italy, with similarities to the Recent Zebrasoma. Stud. e Ric. sui giacimenti terziari di Bolca. 7:7–19.

Sorenson L., Santini F., Carnevale G., Alfaro M.E. 2013. A multi-locus timetree of surgeonfishes (Acanthuridae, Percomorpha), with revised family taxonomy. Mol. Phylogenet. Evol. 68:150–160.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics.

Stanke M., Keller O., Gunduz I., Hayes A., Waack S., Morgenstern B. 2006. AUGUSTUS: A b initio prediction of alternative transcripts. Nucleic Acids Res. 34:435–439.

Tyler J.C. 1970. Osteological Aspects of Interrelationships of Surgeon Fish Genera (Acanthuridae) Author (s): James C . Tyler Source : Proceedings of the Academy of Natural Sciences of Philadelphia , 1970 , Vol . 122 Published by : Academy of Natural Sciences Stable U. 122:87–124.

Tyler J.C. 1997. The Miocene fish Marosichthys, a putative tetraodontiform, actually a perciform surgeon fish (Acanthuridae) related to the recent Naso. Beaufortia. 47:1–10.

Tyler J.C. 1999. A new genus and species of surgeon fish (Acanthuridae) with four dorsal-fin spines from the Eocene of Monte Bolca, Italy. Stud. e Ric. sui giacimenti terziari Bolca.:257–268.

Tyler J.C. 2005. Redescription and basal phylogenetic placement of the acanthurid surgeon fish Gazolaichthys vestenanovae from the Eocene of Monte Bolca, Italy (Perciformes; Acanthuroidea). Stud. E Ric. Sui Giacimenti Terziari Di Bolca, Mus. Civ. Di Stor. Nat. Di Verona. 11:97–117.

Tyler J.C., Bannikov A.F. 2000. A new species of the surgeon fish genus Tauichthys from the Eocene of Monte Bolca, Italy (Perciformes, Acanthuridae). Boll. del Mus. Civ. di Stor. Nat. di Verona. 24:29–36.

Tyler J.C., Johnson G.D., Nakamura I., Collette B.B. 1989. Morphology of Luvarus imperialis (Luvaridae) with a phylogenetic analysis of the Acanthuroidei (Pisces). Smithson. Contrib. to Zool.

Tyler J.C., Micklich N.R. 2011. A new genus and species of surgeon fish (Perciformes, Acanthuridae) from the Oligocene of Kanton Glarus, Switzerland. Swiss J. Palaeontol. 130:203–216.

Tyler J.C., Sorbini L. 1998. On the relationships of Eonaso, an Antillean fossil surgeon fish (Acanthuridae). Stud. E Ric. Sui Giacimenti Terziari Di Bolca, Mus. Civ. Di Stor. Nat. Di Verona. 7:35–42.

Winterbottom R. 1993. Myological Evidence for the Phylogeny of Recent Genera of Surgeonfishes (Percomorpha , Acanthuridae), with Comments on the Acanthuroidei. Am. Soc. Ichthyol. Herpetol. 1993:21–39.

Winterbottom R., McLennan D.A. 1993. Cladogram versatility: Evolution and biogeography of acanthuroid fishes. Evolution (N. Y). 47:1557–1571.

Yang C., Zheng Y., Tan S., Meng G., Rao W., Yang C., Bourne D.G., O'Brien P.A., Xu J., Liao S., Chen A., Chen X., Jia X., Zhang A. bing, Liu S. 2020. Efficient COI barcoding using high throughput single-end 400 bp sequencing. BMC Genomics. 21:1–10.

Zhang C., Sayyari E., Mirarab S. 2017. ASTRAL-III: Increased scalability and impacts of contracting low support branches. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 10562 LNBI:53–75.