UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

MARINE FISH DIVERSIFICATION ALONG THE WATER COLUMN GRADIENT: A TALE OF MANY THREADS

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

By

EMANUELL DUARTE RIBEIRO
Norman, Oklahoma
2022

MARINE FISH DIVERSIFICATION ALONG THE WATER COLUMN GRADIENT: A TALE OF MANY THREADS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Ricardo Betancur-R

Dr. Richard Lupia

Dr. Richard Broughton

Dr. Ingo Schlupp

Dr. Matt Friedman

Abstract

Decades of investigation about the evolutionary mechanisms responsible for the astonishing diversity of fishes have suggested habitat transitions along the water column as a prominent ecological divergence axis affecting morphological and lineage diversification. This process is pervasive across all major freshwater and marine fish clades, and evidence supporting its importance comes from a broad spatiotemporal spectrum. On a narrow scale, recent parallel invasions of freshwater lakes by marine three-spined stickleback populations have repeatedly triggered the evolution of two divergent types: a deep-body form living in association with the bottom (benthic) and a slender-body form found in the water column (pelagic). On the other end of the spectrum, evidence from the fossil record suggests that a significant component of the diversification of spiny-rayed fishes (acanthomorphs) has resulted from the colonization of the ecological space previously occupied by pelagic species that became extinct during the Cretaceous-Paleogene (K–Pg) mass extinction. Nevertheless, despite important progress made on this front, further investigation of the processes associated with benthic-pelagic transitions in a robust phylogenetic framework is necessary for bridging evidence from both ends of the evolutionary continuum.

The overarching goal of my dissertation was to produce comprehensive phylogenomic trees and to use an array of phylogenetic comparative methods (PCMs) to investigate the drivers and the effects of the adoption of habitat regimes in two evolutionarily intriguing and economically important ray-finned fish groups: the family Lutjanidae (ca. 130 species), including snappers and fusiliers; and the series Carangaria (ca. 1090 species), including flatfishes, jacks, marlins, barracudas, remoras, and allies. Additionally, my dissertation provides new evidence supporting the (recently challenged) single evolutionary origin of flatfishes, and used phylogenetically-informed genotype-to-phenotype (PhyloG2P) approaches to identify the molecular mechanisms underlying their asymmetrical development–a key adaptation linked to transitions from pelagic to benthic habitats.

In the first chapter of my dissertation, I applied an integrative approach to quantify the scope and strength of convergent evolution in pelagic lutjanids. As part of a large collaborative effort, I collected genome-wide DNA sequence data for approximately 80% of the total number of species in the group and aggregated data layers for body shape, habitat occupancy, geographic distribution, and paleontological information. The results

show that multiple independent colonizations of the water column by ancestrally benthic lutjanid lineages are persistently associated with the convergent evolution of slender bodies and furcate caudal fins. Lineage diversification and transition dynamics vary asymmetrically between habitats, with benthic lineages diversifying faster and colonizing midwater habitats more often than the reverse. These findings demonstrate that convergent phenotypes are ubiquitous among pelagic lutjanids, likely resulting from the strong locomotion constraints imposed by the viscosity of water and drag flow.

For my second chapter, I assessed how transitions along the water column triggered by ecological opportunity in the wake of the Cretaceous-Paleogene (K-Pg) mass extinction shaped diversification dynamics in Carangaria. The results suggest that a remarkable proportion of Carangaria's morphological variation originated in tandem with a marked incidence of habitat shifts along the benthic-pelagic axis during the Paleocene. Likewise, lineage diversification analyses show that species accumulation rates vary as a function of time, with peaks reached early in the clade's history. These results indicate that all major lineages and body plans in Carangaria originated in an early burst of evolution shortly after the K-Pg mass extinction, allowing the occupation of newly released ecological niches along the water column gradient.

In my third chapter, I explored the evolutionary origins of the flatfish asymmetric body plan—an extreme evolutionary innovation responsible for the successful colonization and diversification in benthic habitats by carangarian lineages. I analyzed three independent genome-scale phylogenetic datasets, showing that lineage-specific variation in base composition (i.e., base compositional non-stationarity) of exonic markers deeply affects the ability of commonly-used phylogenetic models to resolve flatfishes as monophyletic. This problem can be alleviated, however, by using models that accommodate lineage-specific variation in base composition. The implementation of such models provides decisive support for the single-origin hypothesis of the asymmetric body plan. I also applied models of codon substitution using complete genomes to investigate the molecular bases of the complex phenotypic modifications in flatfishes. The findings corroborate that the flatfish cranial asymmetry results from a complex interaction between thyroid hormones (THs) and other important developmental pathways (e.g., WNT, BMP, RA).

In summary, my analyses demonstrate that ecological divergence along the benthic-pelagic axis greatly impacts rates of morphological and lineage diversification in different clades. Transitions are ubiquitous in time and space but are largely affected by periods of enhanced ecological opportunities, such as the wake of mass extinction events. This process is correlated with the recurrent evolution of convergent body plans but also with some of the most extreme morphological and physiological innovations among vertebrates, such as the internal body temperature regulation (endothermy) in marlins, billfishes, and swordfishes that allows the exploitation of a broad thermal niche, and the flatfish asymmetric body-plan

responsible for their successful colonization of the sea bottom. These new insights represent an important step towards understanding the dynamics of animal biodiversity origination in the aquatic realm, ultimately providing a reference to compare against other groups of organisms diversifying along habitat gradients.

To Alice Duarte, a survivor who
believed education could change lives

Contents

List of Figures

Chapter 1

Evolutionary determinism and convergence associated with water-column transitions in marine fishes

Published in *Proceedings of the National Academy of Sciences of the United States of America* (https://doi.org/10.1073/pnas.2006511117)

Melissa Rincon-Sandoval*, Emanuell Duarte-Ribeiro*, Aaron M. Davis, Aintzane Santaquiteria, Lily C. Hughes, Carole C. Baldwin, Luisángely Soto-Torres, Arturo Acero P., H. J. Walker Jr., Kent E. Carpenter, Marcus Sheaves, Guillermo Ortí, Dahiana Arcila, Ricardo Betancur-R*
(*)*equal contribution*

1.1 Abstract

Repeatable, convergent outcomes are prima facie evidence for determinism in evolutionary processes. Among fishes, well-known examples include microevolutionary habitat transitions into the water column, where freshwater populations of different species (e.g., sticklebacks, cichlids, whitefishes) diverge towards slender-bodied pelagic forms and deep-bodied benthic forms. But the consequences of such processes at deeper macroevolutionary scales in the marine environment are less clear. We applied a phylogenomics-based integrative, comparative approach to test hypotheses about the scope and strength of convergence in a marine fish clade with a worldwide distribution (snappers and fusiliers, family Lutjanidae) featuring multiple water-column transitions over the past 45 million years. We collected

genome-wide exon data for 111 (approximately 80%) species in the group and aggregated data layers for body shape, habitat occupancy, geographic distribution, and paleontological and geological information. We also implemented novel approaches using genomic subsets to account for phylogenetic uncertainty in comparative analyses. Our results show independent incursions into the water column by ancestral benthic lineages in all major oceanic basins. These evolutionary transitions are persistently associated with convergent phenotypes, where deep-bodied benthic forms with truncated caudal fins repeatedly evolve into slender midwater species with furcate caudal fins. Lineage diversification and transition dynamics vary asymmetrically between habitats, with benthic lineages diversifying faster and colonizing midwater habitats more often than the reverse. Convergent ecological and functional phenotypes along the benthic-pelagic axis is pervasive among different lineages and across vastly different evolutionary scales, achieving predictable high-fitness solutions for similar environmental challenges, ultimately demonstrating strong determinism in body shape evolution.

1.2 Significance

Body shape is a strong predictor of habitat occupation in fishes and has been shown to change rapidly at micro-evolutionary scales in well-studied freshwater systems such as sticklebacks and cichlids. Deep-bodied forms tend to occur in benthic habitats, while pelagic species typically have slender hydrodynamic body plans. Recurrent evolution of this pattern across distantly related groups suggests that limited sets of high-fitness solutions exist as a result of environmental constraints. We demonstrate that similar constraints operate at deeper evolutionary scales. Lutjanids are primarily benthic dwellers that repeatedly transitioned into midwater habitats in all major oceans throughout their 45-million-year history. Midwater species strongly converge in body shape emphasizing the deterministic nature of the evolutionary process molding form and function along the benthic-pelagic axis.

1.3 Introduction

A question of central interest in biology is whether evolutionary outcomes can be predictable and thoroughly governed by the laws of nature or contingent on a sequence of unpredictable historical events, such as rare environmental catastrophes, which may be sensitive to circumstances inherent to particular evolutionary paths (Losos et al., 1998). Reconciling this conundrum may depend largely upon the scope and strength of evolutionary convergence—the process whereby natural selection tends to produce a limited set of high-fitness solutions when confronted with similar challenges imposed by the environment (i.e., the adaptive landscape). Convergence ranks among the most conspicuous features in biodiversity and

the general mechanisms by which the physical constants of nature constrain morphological outcomes have been recognized for decades (Agrawal, 2017; Losos, 2011; Wake et al., 2011). Nevertheless, the deterministic nature of the processes leading to convergent evolution is still contentious (Blount et al., 2018; Burns and Sidlauskas, 2019).

An emblematic example of evolutionary convergence comes from the aquatic environment, where distantly related pelagic lineages tend to evolve similar body plans. Based on these observations, G. McGhee hypothesized that there are limited ways to build a fast-swimming aquatic organism, which is why dolphins, swordfish, sharks, and ichthyosaurs all present streamlined fusiform bodies—a nontrivial adaptation to the locomotion constraints imposed by the viscosity of water and drag flow (McGhee, 2011). In ray-finned fishes (Actinopterygii), the evolution of fusiform body plans also has a strong adaptive basis and is frequently associated with the invasion of the water column by primarily benthic lineages. Body elongation has been recognized as the primary axis of diversification in fishes (Burress et al., 2017; Hulsey et al., 2013; Ribeiro et al., 2018; Tavera et al., 2018), and evidence supporting this deterministic process comes from a broad spatio-temporal spectrum. At a narrow scale, post-Pleistocene parallel invasions of freshwater lakes by marine three-spined stickleback populations have repeatedly triggered the evolution of two divergent phenotypes, a deep-body form associated with more benthic habitats and a slender-body form that occurs in the water column (Walker, 1997; Willacker et al., 2010). Quantitative assessments at microevolutionary scales have documented similar cases of resource partitioning on sympatric populations of lake cichlids and whitefishes, among others (Clabaut et al., 2007; Cooper et al., 2010; Hulsey et al., 2013; Lu and Bernatchez, 2011; Rundle et al., 2000). At the other end of the evolutionary spectrum, evidence from the fossil record shows a significant component of the Paleogene spiny-rayed teleost (acanthomorph) radiation that colonized areas of the morphospace previously occupied by incumbent pelagic species that became extinct during the Cretaceous-Paleogene (K–Pg) mass extinction (Friedman, 2010; Ribeiro et al., 2018; Sallan et al., 2018).

Here, we assess the role of convergent evolution associated with transitions along the benthic-pelagic axis in a clade of marine fishes—the snappers and fusiliers in the family Lutjanidae—that bridges both ends of the evolutionary continuum. Previous phylogenetic studies on snappers suggest that benthic clades of different ages have repeatedly transitioned into the water column, and these transitions are seemingly associated with modifications in feeding ecology and body elongation (Frédérich and Santini, 2017). Based on these observations, we first test the hypothesis that independent incursions into the water column in this group are constrained to a narrow portion of the adaptive landscape, representing a remarkable case of convergent evolution across fish diversity. Second, given the widespread distribution of this family and the potential temporal range of habitat transitions in the clade, we hypothesize that these evolutionary transitions have occurred independently within all

major oceanic basins.

Community ecology studies have demonstrated that marine biodiversity is higher in benthic than pelagic environments (Angel, 1993; Gray, 1997), suggesting that the adoption of the midwater lifestyle by lutjanids may have resulted in an evolutionary ratchet, where the acquisition of specialized traits are selectively advantageous in the short term, but in the long term can create an evolutionary trap due to lowered speciation or elevated extinction rates (Agnarsson et al., 2006; Goldberg et al., 2010). This hypothesis makes two predictions: (i) habitat transitions from benthic to midwater systems are expected to be unidirectional or asymmetric, and (ii) midwater lineages diversify slower than their benthic counterparts, which stems from the fact that the microhabitat homogeneity of pelagic systems provides fewer opportunities for speciation than the niche-diverse benthic environment.

To address these questions using rigorous quantitative approaches, we estimated a set of taxonomically rich time trees for lutjanids based on genome-wide data and used an integrative comparative dataset that includes morphological and ecological data layers in combination with geographic distribution data. By conducting a suite of phylogenetic comparative analyses, we examined the temporal and geographic scope of evolutionary convergence among midwater snapper and fusilier lineages. These analyses show that repeated habitat transitions from bottom to midwater systems are linked strongly to patterns of evolutionary convergence in body shape and also are associated with asymmetric habitat transitions and slower rates of lineage diversification. These transitions took place independently within all major biogeographic regions. Taken together, our findings ultimately reinforce the deterministic nature of evolution as a consequence of the similar use of the niche space along the benthic-pelagic axis.

1.4 Materials and Methods

Taxonomic sampling and genomic data

Extended Materials and Methods are reported in the Appendix A. Our genomic sampling includes 85 newly sequenced species of snappers and fusiliers from specimens deposited in multiple fish collections. To further expand the taxonomic scope, we retrieved sequences for 26 additional ingroup species from GenBank. Our combined dataset contains 111 species (Appendix A). We also used fourteen grunt species (Haemulidae) as outgroups. High quality DNA extractions were sent to Arbor Biosciences for a targeted enrichment and sequencing. Our target capture probes are based on a set of 1,104 single-copy exons optimized for ray-finned fish phylogenetics (Hughes et al., 2018, 2020). We also included 15 legacy exons into the probe set. After performing standard procedures for sequence quality control and assembly, we aligned exons by taking into account their reading frames.

Accounting for missing data in phylogenomic inference

We assembled two main data matrices: (i) an expanded matrix with all genes and taxa, including GenBank sequences, and (ii) a reduced matrix obtained with the MARE (matrix reduction) package (Meyer et al., 2011). For each matrix, we determined the best-fitting partitioning schemes and nucleotide substitution models for both genes and codon positions using PartitionFinder2 (Lanfear et al., 2017). We also assembled 13 additional subsets by manually subsampling the expanded matrix (see details below). For all datasets, we estimated ML trees in RAxML v8.2.4 (Stamatakis, 2006; Stamatakis et al., 2008) using the partition output obtained with PartitionFinder2. Species trees were then inferred with ASTRAL-II v4.7.12 (Mirarab and Warnow, 2015) using individual RAxML-based gene trees as input.

Accounting for topological and temporal uncertainty

Our genome-scale exonic dataset provides an opportunity to account for tree topology and divergence time uncertainties in phylogenetic comparative methods by analyzing alternative gene sets. The goal was to build a number of largely independent subsets (subsampled from the expanded matrix), each with a sufficient number of genes to overcome sampling error by capturing our knowledge of the phylogeny of the group in the best possible manner. We assembled thirteen largely independent subsets (seven with 89 loci and six with 90 loci), all of which overlap in only four genes thereby maintaining the same set of species across subsets. As input topologies for phylogenetic dating in MCMCTree (see below), we inferred a total of 28 phylogenetic trees using both RAxML and ASTRAL-II. Two trees were estimated using the complete expanded matrix, including a 'master tree' based on the RAxML topology; the remaining 26 trees were obtained with the 13 subsets subsampled from this matrix. While most downstream comparative analyses used the 28 trees, some were computationally demanding and therefore were based on the 'master tree' only (indicated whenever applicable).

Phylogenetic dating

We conducted divergence time estimations using the MCMCTree package as implemented in the program PAML v4.9a (Yang, 2007), which can handle genome-scale datasets in a Bayesian framework (dos Reis Yang, 2019). Because MCMCTree running time depends more on the number of partitions defined rather than the number of genes included (dos Reis Yang, 2019), all 28 subsets used only two partitions (1st+2nd and 3rd codon positions). We applied seven calibration points, two based on fossils with uniform distributions and five based on a geological event with flat-tailed Cauchy distributions (Appendix A, Materials

and Methods).

Reconstruction of ancestral habitats and ancestral ranges

The habitat occupancy dataset (Appendix A) was compiled by aggregating information from a wide range of sources, including FishBase (Froese and Pauly, 2019), the primary literature (Allen, 1985; Andrews et al., 2014; Carpenter, 2001; Dunlap et al., 2016; Haight et al., 1993; Kuiter and Tonozuka, 2001; Lloyd, 2006; Newman, 2009; Robertson and Allen, 2015; Robertson and Van Tassell, 2016; Taquet and Diringer, 2013; White et al., 2013), and by consulting experts. To account for cases with uncertain habitat occupancy, we implemented ancestral character reconstructions that take into account tip-state ambiguity based on stochastic character mapping analyses (SIMMAP Revell, 2012), as implemented in the R package phytools (Revell, 2012). We explicitly tested the relative fit of two models of discrete state evolution for binary states: equal (ER) or symmetric rates, and all rates different (ARD) or asymmetric rates. The best fit model was then used to identify the best possible explanation concerning the distribution of habitat transition events throughout the evolutionary history of snappers and fusiliers using SIMMAP. We also classified species according to their geographical ranges. We built a presence/absence matrix of species considering six recognized marine biogeographic regions (Appendix A): West-Indian Ocean (WIO), Central Indo-Pacific (CIP), Central Pacific (CP), Tropical Eastern Pacific (TEP), Western Atlantic (WA), and Eastern Atlantic (EA). Ancestral area reconstructions were performed using the R package BioGeoBEARS (Matzke, 2013). Using the 'master tree' as the input phylogeny, 12 different biogeographic models were tested. We analyzed each model using three time-slices according to different geological events (see Appendix A for details on models and matrices used for BioGeoBEARS). For simplicity, we summarized ancestral ranges into three major ocean realms by merging EA and WA into the Atlantic, WIO, CIP, and CP into the Indo-Pacific, and leaving the TEP as originally coded (Figure 1.1).

Geometric morphometrics on body shape

The laterally compressed body plan of snappers and fusiliers makes this group well suited for the summarization of morphological diversity using two-dimensional geometric morphometric approaches. We assembled a specimen imagery dataset from museum collections or curated images retrieved from online repositories. To account for intraspecific variation, our dataset includes 1-4 individuals from each of the 111 species (total 413 individuals; mean 3.72 individuals per species; Appendix A). We generated three alternative datasets based on digitized landmarks: (i) A full-body and fin shape dataset; (ii) a body-only dataset; and (iii) a fins-only dataset (see Appendix A, Figure A.1). For each dataset, we performed Procrustes superimposition, calculated species-average coordinates, and conducted both stan-

dard (PCA) and phylogenetically-corrected (pPCA) principal component analyses (Adams and Collyer, 2019; Revell, 2009). Finally, we determined the number of meaningful PC axes using the broken-stick model (Jackson, 1993; Peres-Neto et al., 2005), which minimizes loss of signal while avoiding noise from less relevant axes.

Convergence analyses

To assess the scale and nature of convergence among taxa exhibiting similar habitat regimes, we ran a set of recently proposed multivariate phylogenetic comparative methods for each of the three alternative morphological datasets (full-body shape, body-only, and fins-only). We first tested the relative fit of a range of evolutionary models using the package mvMORPH (Clavel et al., 2015). These include a single-rate Brownian Motion (BM) model, a single-regime Orstein-Uhlenbeck (OU) model, and multi-selective regime BM (BMM) and OU (OUM) models. We also tested for correlation between habitat occupation and the four most relevant PC axes using the threshold model, which assesses the association between a discrete trait and a continuous character that co-vary according to an underlying, unobserved trait called liability (Felsenstein et al., 2012). We explicitly tested for convergent evolution using the C1-C4 distance-based metrics implemented in convevol (ran using the 'master tree'), as well as the Wheatsheaf index implemented in the R package Windex (Arbuckle et al., 2014). Finally, we used other data-driven approaches, as implemented in the R package ℓ1ou v1.42 (Khabbazian et al., 2016) and SURFACE v0.4 (Ingram and Mahler, 2013), to estimate the optimal number of selective regimes under an Ornstein-Uhlenbeck process applied to the least absolute shrinkage and selection operator (LASSO).

State-dependent diversification

We evaluated the influence of habitat type (benthic vs. midwater dwellers) on lineage diversification dynamics using state-dependent speciation and extinction (SSE) approaches (Maddison et al., 2007). We first used BiSSE, as implemented in the R package diversitree (Fitzjohn, 2012), to estimate habitat-dependent rates of diversification in a Bayesian framework. We also applied HiSSE (Hidden State Speciation and Extinction), an SSE approach that accounts for 'hidden' states, by testing the relative fit of a set of alternative branching models while accounting for hidden states. Finally, we used the nonparametric FiSSE approach, which has shown to be robust to phylogenetic pseudo-replication and model misspecification (Rabosky and Goldberg, 2017). See Appendix A for details.

1.5 Results

Phylogenomic inference and tree uncertainty in comparative analysis

Extended results are reported in the Appendix A. Using exon capture approaches (Hughes et al., 2018, 2020), we assembled two main phylogenomic data matrices: (a) an expanded matrix that includes all genes and taxa sequenced for this study, with the addition of GenBank sequences aimed at increasing taxonomic coverage for downstream comparative analyses (1,115 exons and 474,132 nucleotide sites for 111 out of ca. 136 species; 37% data missingness); and (b) a reduced matrix obtained with a matrix reduction algorithm, used to assess the sensitivity of phylogenetic results to missing data (1,047 exons and 448,410 nucleotide sites for 84 species; 16% data missingness). We conducted phylogenomic analyses using maximum likelihood (ML) and coalescent-based approaches. Inferred trees were resolved with strong support and are largely congruent among tested approaches and with results from previous studies (Alfaro et al., 2018; Betancur-R et al., 2013; Frédérich and Santini, 2017; Hughes et al., 2018; Rabosky et al., 2018). All analyses invariably resolved seven major clades (Figure 1.1; Appendix A, Figure A.2-A.5), confirming that the family Lutjanidae, as traditionally defined, is non-monophyletic with fusiliers (Caesionidae) deeply nested within the broader snapper clade. The relationships estimated with the expanded matrix were highly consistent with those based on the reduced matrix, providing a robust phylogenomic framework for downstream comparative analyses.

In addition to expanded and reduced datasets, we also analyzed 13 randomly assembled subsets derived from the expanded matrix, each with a sufficient number of genes to overcome sampling error (Appendix A). Resulting trees reflect uncertainty in divergence times and phylogenetic relationships, an approach that is fundamentally different 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. We estimated a total of 28 trees that include all taxa using both concatenation-based maximum likelihood (RAxML) and coalescent-based (ASTRAL-II) approaches applied to the expanded matrix and its 13 subsets. Divergence-time estimates, using the 28 input topologies and seven calibration points (Appendix A), were in reasonably good agreement with those from previous multi-locus studies for the family (Appendix A, Figure A.7-A.8). Divergence time estimations date the age of crown lutjanids to the middle Eocene (46 Ma, 95% HPD: 40-49 Ma), and the stem age close to the Cretaceous–Paleogene (K-Pg) boundary, around 64 Ma.

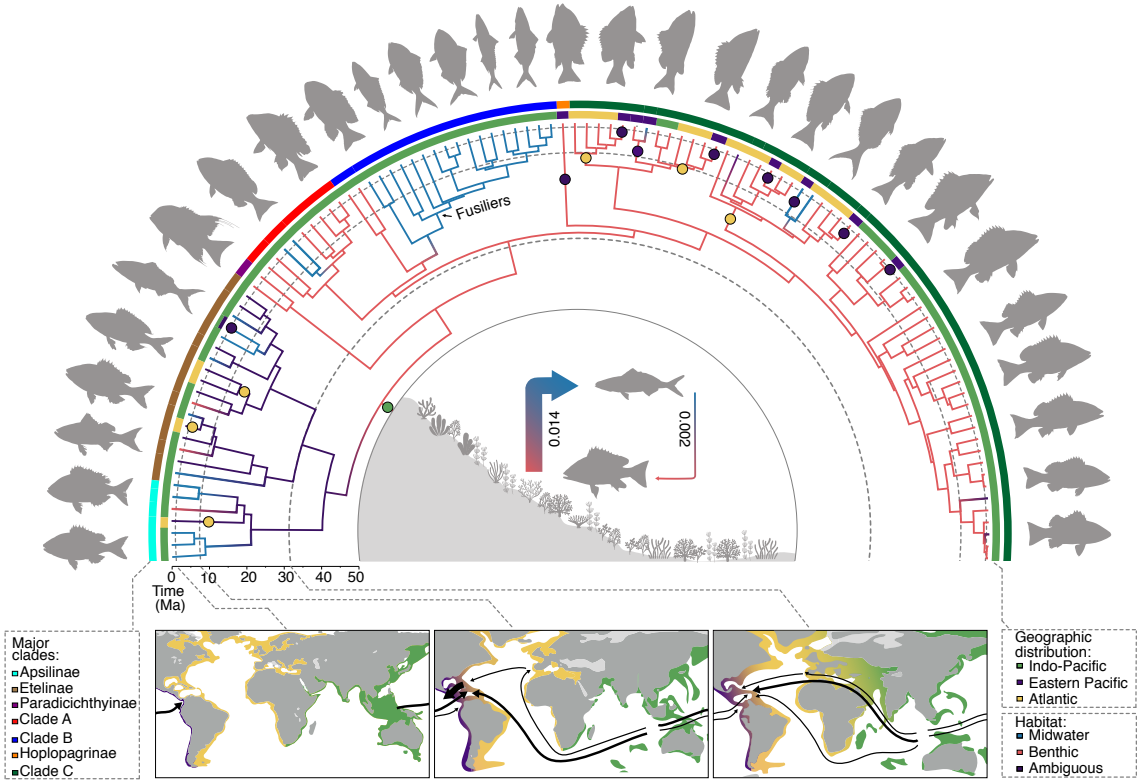

Figure 1.1: Phylogeny, habitat transitions, and biogeography of snappers and fusiliers. The tree shown is derived from a concatenation-based maximum-likelihood analysis of 1,115 exons, with node ages estimated from a time-calibrated analysis using seven calibration points in MCMCTree. The habitat reconstructions for benthic and midwater lineages, shown as colored branches in the tree, account for phylogenetic uncertainty (28 trees) and habitat coding ambiguity (13 tips with uncertain or multistate habitats). The color gradients along branches denote habitat transitions; the purple branches indicate lineages with ambiguous habitats based on reconstructions using alternative coding schemes (Appendix A, Figures A.8–A.11). The colored circles indicate colonization events (inferred with BioGeoBEARS; see also Appendix A, Figures A.8 and A.12–A.14) of the Atlantic (yellow circles) and the tropical eastern Pacific (purple circles) from Indo-Pacific lineages (center of origin; green circle). The arrows in the maps depict reconstructed colonization routes by different lineages in three time slices: 50 to 12 Ma (mean, 31 Ma), before the closure of Tethys Seaway; 12 to 2.8 Ma (mean, 7.4 Ma), after closure of Tethys Seaway and before the closure of the Isthmus of Panama; and 2.8 Ma to present (mean, 1.4 Ma), after the closure of the Isthmus of Panama. The thickness of the arrows is proportional to the number of lineages that colonized via each route; for some lineages, colonization routes are uncertain, and thus all alternative routes are depicted. The arrows in the central panel show the transition rates between benthic and pelagic habitats, as estimated with HiSSE (see also Appendix A).

The geography of habitat transitions

To test the hypothesis that evolutionary transitions have occurred independently within all major oceanic basins, we performed ancestral habitat and ancestral area reconstructions. To infer the history of habitat transitions in Lutjanidae, we first assigned species into two major habitat categories (benthic and midwater dwellers) and accounted for uncertainty in habitat coding. Benthic habitats represent the most likely ancestral condition for Lutjanidae (Figure 1.1). Although an initial assessment of the fit of two alternative models of discrete state evolution (equal rates or ER and all rates different or ARD) failed to reject the null ER model (see details in Appendix A), ancestral state reconstructions using both models produced similar results, identifying an average of 14.01 ($\pm$0.5) transitions between the two habitat categories. Midwater lutjanid lineages originated from benthic lineages 9.01 ($\pm$0.1) times on average throughout their evolutionary history, which is roughly twice as frequent as midwater-to-benthic transitions (mean 4.5$\pm$0.4). In line with these results, HiSSE analyses support a model that accounts for habitat dependent diversification (HiSSE benthic; see below) and asymmetric transition rates (mean q = 0.016 and 0.008 for benthic-to-midwater and midwater-to-benthic transitions, respectively; Figures 1.1,1.3f; Appendix A, Figure A.21b).

To reconstruct ancestral areas, we built a presence-absence matrix of species distribution using alternative biogeographic schemes (Appendix A). Inferences of ancestral ranges using BioGeoBEARS (Matzke, 2013) indicate an Indo-Pacific Ocean origin for lutjanids, with subsequent independent colonization events of the New World via multiple routes (see Appendix A Figure A.9-A.11, and Supplementary Results for an expanded account on the biogeography). By merging results of ancestral habitat and ancestral range reconstructions we find support for the independent and recurrent invasion of the water column by benthic lineages at least once within each of the three major oceanic basins (Figure 1.1). While some basins feature more transitions than others (e.g., Indo-Pacific vs eastern Pacific; Figure 1.1), habitat transitions in lutjanids are clearly widespread. Ancestral habitat reconstructions for benthic and midwater lineages are shown as colored branches in the tree (reconstructed from 28 trees subjected to SIMMAP). Nodal pies indicate the most likely ancestral area based on BioGeoBEARS where SIMMAP analyses identify a habitat transition. Arrow values in the central panel show the transitions rates between benthic and pelagic habitats estimated with HiSSE.

Ecomorphological convergence

To test whether invasions of the water column were associated with a set of convergent high-fitness solutions, we assembled a specimen imagery database and built three alternative datasets based on digitized landmarks: (i) a full-body shape dataset; (ii) a body-only dataset; and (iii) a fins-only dataset (Appendix A, Figure A.1). Traitgram-informed morphospaces

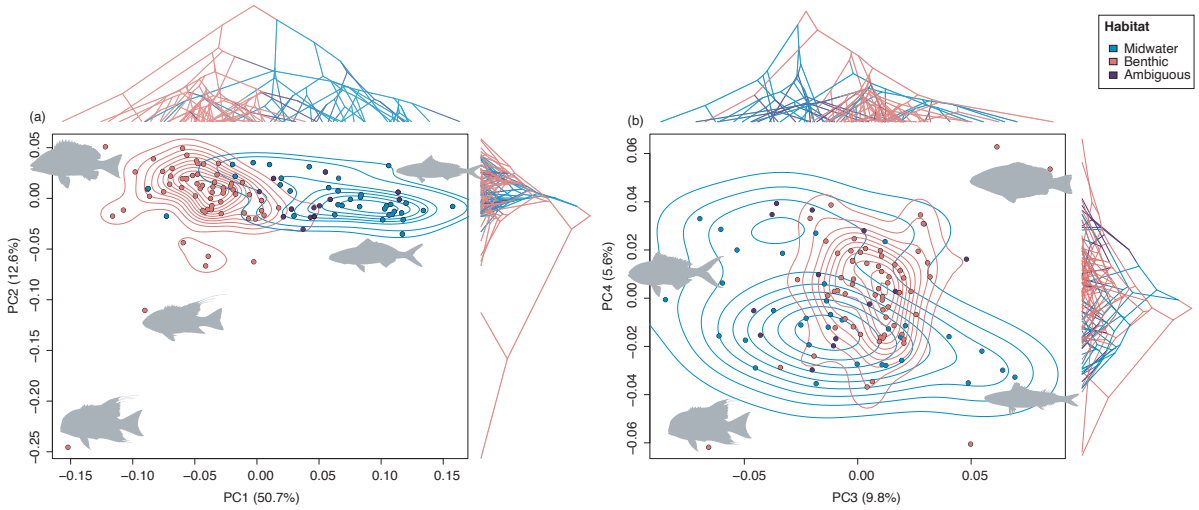

Figure 1.2: The traitgram-informed morphospaces for lutjanids illustrating ecomorphological partitioning and convergence across benthic and midwater line- ages, as estimated using the full-body dataset. The contour lines represent the two-dimensional density distributions of the species in each habitat state. The traitgrams overlain along PC axes depict the phylogeny in Figure 1.1, including the ancestral habitat reconstructions estimated with SIMMAP (A, PC1 versus PC2; B, PC3 versus PC4). The color gradients along branches denote habitat transitions; the purple branches and data points indicate lineages with ambiguous habitats based on alternative coding schemes. The branches shifting from red to blue along PC1 extremes highlight convergent evolution in midwater lineages. The parenthetical values indicate the total variance explained by each PC axis. Downloaded

(Figure 1.2) show that different lutjanid midwater lineages independently evolved slender bodies and furcated caudal fins, an indication of strong ecologically-driven evolutionary convergences. This pattern is further confirmed using the threshold model (Felsenstein, 2005), where the full-body shape dataset reveals substantial correlation between PC1 and the two habitat states ($r^2 = 0.70$). The remaining three PC axes (PC2-4) summarize further relevant aspects in fin-shape variation and ornamentation. We detected the same pattern for the body-only ($r^2 = 0.66$) and fins-only ($r^2 = 0.68$) datasets, where only PC1 exhibits significant correlations. We also found an extensive overlap between benthic and midwater species at the lower PC axes, reflecting lower correlations between the PC2-PC4 and habitat occupancy data ($r^2 = 0.04$-0.20). These results suggest that ecomorphological convergence is less clearly associated with PC2-PC4 axes than it is to the main PC1 axis (Appendix A, Figure A.12).

To further assess the nature and strength of convergence, we compared the relative fit of a set of models of trait evolution in a multivariate framework using mvMORPH. While results of the full-body and fins-only datasets tend to show split support for the two multi-selective-regime models (OUM and BMM; see Methods), analyses based on the body-only dataset unequivocally favor the OUM model with distinct selective regimes corresponding to different habitat categories (Figures 1.3a,b; Appendix A, Figure A.13). This result provides

evidence for strongly constrained adaptive optima among independent lineages with similar habitat preferences along the benthic-pelagic axis. To gauge the strength of convergence associated with the invasion of the water column, we used the convevol distance-based metrics, calculated as the difference in distance between species tips and the maximum distance between those taxa through their evolutionary history (C1-C4; Appendix A). The C1-C4 statistics were all significant for the three alternative morphometric datasets, with midwater lineages shortening about half of their phenotypic distance by subsequent convergent evolution (C1 = 42-48%; Appendix A). Likewise, results using the Wheatsheaf index (w = 1.2-1.35; Appendix A, Figure A.14) identified significantly stronger convergence in midwater species than would be expected from a random distribution of trait values simulated under a Brownian Motion model (BM) across the tree ($p < 0.01$). All w values were similar, and the confidence interval overlapped among the three alternative morphometric datasets, suggesting that both body shape and fin morphologies have a similar strength in convergent evolution. To further validate these results, we calculated w using benthic species as focal clades. In this case, w was significantly smaller than values simulated under BM in all three morphometric datasets (w = 0.82-1.03; $p > 0.99$). This suggests that morphological diversity is high among benthic dwellers, whereas strong convergent evolution is mostly restricted to midwater lutjanids.

Finally, we assessed the extent of convergent evolution without a priori habitat designations using ℓ1ou and SURFACE for the three different datasets (Appendix A, Figure A.15-A.19). Although the pattern is less clear with ℓ1ou, we also identified cases of convergence across lineages with convergent peaks between clades with similar body plans (deep or slender bodies). In all cases, the number of non-convergent (adaptive) peak shifts was higher than the number of convergent peaks (Appendix A). Simulations for ℓ1ou performed under simple OU and BM models revealed significantly greater numbers of convergent shifts (Appendix A) than would be expected by chance (Appendix A, Figure A.20). As expected, ℓ1ou results using pBIC for shift detection were, on average, more conservative for the three datasets. Univariate analyses using SURFACE were similar, but this approach identified a higher number of convergent regimes in most cases (Appendix A, Figure A.18). Taken together, our results suggest overall convergence of many lineages to multiple, shared adaptive peaks in body shape ecomorphology.

Diversification in benthic and midwater lineages

We gauged the preference for different habitat states and their effect on rates of habitat transitions (see above; Figure 1.3f) and lineage diversification (Figure 1.3e), providing a test for the prediction that the adoption of the midwater lifestyle may result in evolutionary ratchets. For 20 out of the 28 trees, model fitting comparisons supported a state-dependent

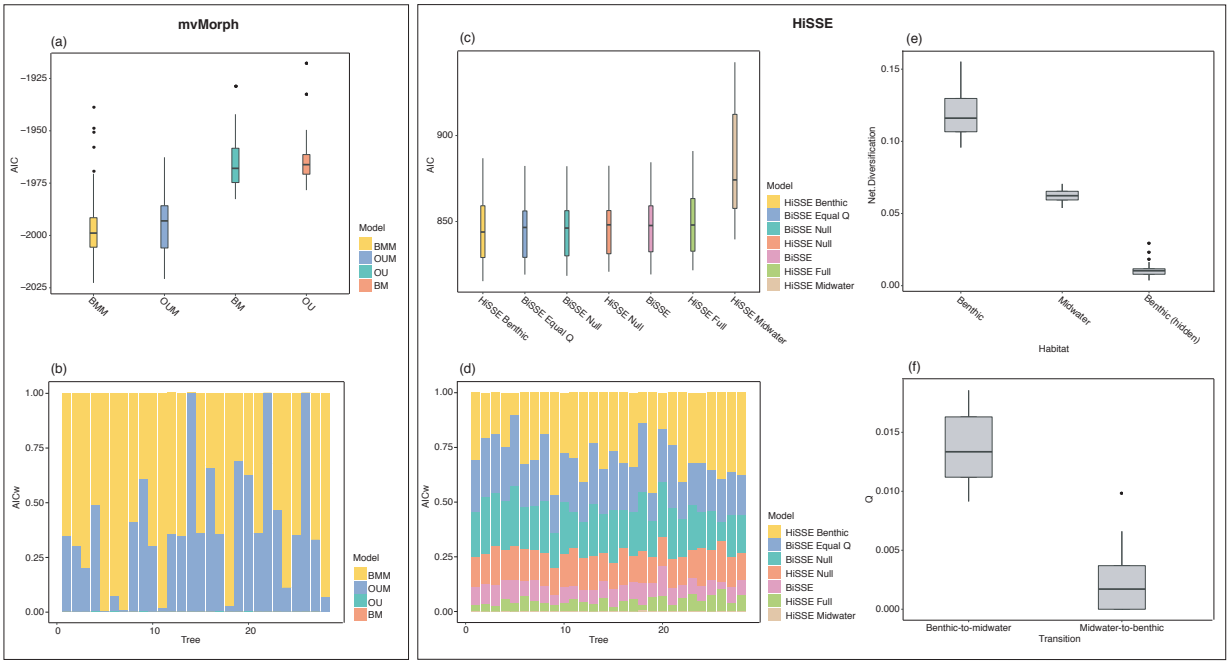

Figure 1.3: The model-fitting comparisons and lineage diversification parameters estimated by accounting for phylogenetic uncertainty (28 trees) and habitat coding ambiguity (13 tips with uncertain or multistate habitats). The comparisons of alternative models of morphological evolution using the full-body dataset: (A) distribution of the Akaike information criterion (AIC) values for the four alternative models of continuous trait evolution (BM, OU, BMM, and OUM) and (B) AIC weights (AICw) of each alternative model and tree. The comparisons for alternative models of lineage diversification: (C) distribution of AIC values for seven alternative SSE models (Appendix A) and (D) AICw for each SSE model based on each of the 28 trees. The estimated lineage diversification parameters: (E) net-diversification values for the three habitat states and (F) transition rates (Q) between benthic and midwater states.

model (Figures 1.3c,d) that incorporates a hidden state (Appendix A) associated with benthic lineages (HiSSE benthic; Appendix A, Figure A.20a). While the 'HiSSE benthic' model is not decisively favored across all trees, finding in some cases substantial support for two alternative null models, under this model net diversification rates (speciation minus extinction) are roughly two times faster in benthic lineages compared to their midwater counterparts. The results we obtained with HiSSE were consistent with those obtained using the non-parametric FiSSE and the more traditional BiSSE approach (Appendix A, Figure A.22), identifying support for habitat-dependent diversification. In agreement with our hypotheses, benthic dwellers tend to show faster rates of net diversification than midwater species, including both faster speciation and slower extinction (Appendix A).

1.6 Discussion

By implementing integrative comparative analyses in a robust phylogenomic framework we find strong evolutionary determinism in benthic-to-midwater transitions along the water column in snappers and fusiliers. While deep body plans in benthic lineages enhance maneuverability in complex habitats with crevices, like coral reefs or rocky bottoms, primarily benthic lineages that independently transitioned into midwater habitats consistently evolved elongate, fusiform bodies and furcate caudal fins, convergent adaptations that reduce hydrodynamic drag and recognizably promote increased swimming performance (Burns and Sidlauskas, 2019; Frédérich and Santini, 2017; Friedman et al., 2016; Langerhans and Reznick, 2010; Tavera et al., 2018; Velotta et al., 2018)—a strong match between form and function (Pigot et al., 2020). This deterministic process appears to be ubiquitous at both temporal and spatial scales, with transitions taking place in lutjanid lineages of different ages and within all major marine biogeographic regions. For instance, while the oldest benthic-to-midwater transition we identified was at ca. 40 Ma (Apsilinae + Etelinae clade), more recent divergences (e.g., 5 Ma) include sister species that lie at extremes of this ecological axis (e.g., *Lutjanus colorado* and *L. aratus*). Snapper and fusiliers thus bridge the gap of this recurrent ecological divergence that is well documented at shallower ends of the evolutionary continuum in model clades such as sticklebacks, cichlids, and whitefish (Clabaut et al., 2007; Cooper et al., 2010; Hulsey et al., 2013; Lu and Bernatchez, 2011; Rundle et al., 2000; Walker, 1997), and more ancient animal lineages such as sharks and aquatic tetrapods (McGhee, 2011). Convergent morphologies among pelagic species strongly suggest that lineages with independent evolutionary histories but similar habitat preferences are drawn towards similar adaptive optima.

The independent evolution of similar phenotypic traits in response to the adoption of similar habitat regimes is a well-characterized indicator of evolutionary convergence. Recurrent transitions are thus indicative of strong evolutionary determinism as a result of similar use of the niche space along the benthic–pelagic axis. The convergent morphologies among midwater species strongly suggest that lineages with independent evolutionary histories but similar habitat preferences are drawn toward similar adaptive optima. Unlike patterns observed among midwater lutjanids, benthic lineages reveal higher phenotypic diversity and weaker convergence. These differences may be the result of greater levels of niche diversity in benthic habitats (Friedman et al., 2020). Similar outcomes are observed at shallower evolutionary scales in European whitefishes (Præbel et al., 2013) and cichlids in Lakes Apoyo and Xiloá in Nicaragua (Elmer et al., 2014), where independent radiations each harbor a single elongated limnetic phenotype and a flock of more variable benthic lineages.

While the focus of this study is on convergent evolution, it is worth emphasizing the strength of evolutionary forces driving phenotypic divergence in body plans along the

benthic-pelagic axis (Burress et al., 2017; Friedman et al., 2016; Price et al., 2019; Ribeiro et al., 2018; Tavera et al., 2018). Midwater lineages with slender bodies are typically a subclade of more generalized deep-bodied benthic groups, and this ecological partition in phylogenetically nested clades has often led to taxonomic misclassifications. This explains why the midwater and planktivorous fusiliers were previously placed in their own family, Caesionidae (Carpenter, 2001, 1987, 1988, 1990, 1993; Fricke et al., 2016; Froese and Pauly, 2019; Miller and Cribb, 2007). Remarkably, some adaptive landscape analyses that detected a single adaptive shift in Lutjanidae (Appendix A, Figure A.14-A.17), identified the shift at the base of the fusilier clade—a direct quantification of the distinct morphology in this group. Similar instances are increasingly being documented in many other marine fish groups based on molecular phylogenies. A prime example includes the midwater Boga in the Caribbean, formerly listed as *Inermia vittata* in the family Emmelichthydae, and now classified as *Haemulon vittatum* in the family Haemulidae (Tavera et al., 2012). A more extreme case comprises the picarels, previously placed in Centracanthidae, a family that is polyphyletically nested within benthic porgies in the family Sparidae (Pavlidis and Mylonas, 2011; Sanciangco et al., 2016). Benthic porgy lineages have thus independently colonized the water column multiple times leading to strong, if not perfect, instances of convergent 'centracanthid' body plans. These divergences can even cross species boundaries, as demonstrated by the benthic Coney (*Cephalopholis fulva*) which is known to practice 'intergeneric hybridization' with the midwater *C. colonus* (formerly *Paranthias colonus*, Tavera et al., 2012; Sanciangco et al., 2016). In all these cases, it is recurrently the planktivorous and slender midwater subclade or species that is derived from the more generalized benthic clade, a result of speciation and adaptation by shifting dietary resources along the water column axis (Bellwood et al., 2004; Floeter et al., 2018; Lobato et al., 2014), ultimately creating taxonomic confusion.

The midwater lifestyle may be an evolutionary ratchet due to overall lower levels of diversity in these habitats, both taxonomically and morphologically, compared to the more species-rich benthic communities. For instance, relatively ancient species-poor clades of marine fishes, such as billfishes, swordfishes, and marlins, suggest slow diversification in pelagic environments (Ribeiro et al., 2018). This is, however, not necessarily the case for other pelagic fish clades (e.g., Scombriformes, Clupeiformes) or midwater lutjanid lineages. While most tests identified higher diversification rates in benthic lineages (Figure 1.1; Appendix A, Figure A.20-A.22), which are roughly twice as fast compared to the midwater counterparts (Figures 1.3e, f), HiSSE analyses failed to support a model of habitat-dependent diversification in 30% of the trees. A remarkable exception includes the fusiliers, a relatively young lutjanid subclade (16 Ma) that comprises 23 species. Fusilier species may school together with congeners and other pelagic species. For instance, the mottled fusilier (*Dipterygonotus balteatus*), the only lutjanid that has adopted an exclusive pelagic lifestyle

as an adult, is often caught together with clupeoids (herrings and anchovies). These observations suggest that midwater lutjanid species present important functional differences and elevated levels of niche partitioning, which may explain the occurrence of species-rich pelagic clades. Ultimately, however, niche partitioning in the resource-poor and homogeneous pelagic environment may result in population density declines and increased trophic specializations, mechanisms that are known to increase extinction vulnerability over long timescales. State-dependent diversification analyses provide some support for these ideas, identifying remarkably faster rates of extinction in midwater than benthic lineage.

Snappers and fusiliers exhibit strong but imperfect morphological convergence associated with habitat transitions. Whereas functional traits associated with ecological partitioning along the benthic-pelagic axis have consistently resulted in similar evolutionary outcomes, some lineages have evolved distinct non-convergent phenotypic adaptations. Exceptions include deep-bodied lineages that tend to occur higher in the water column, such as species in the genus *Macolor*. As pointed out by Hobson (1974) "Obviously many conflicting pressures have differentially affected the morphologies of the various fishes that forage on tiny organisms in the midwaters." Thus, although the slender body plan is pervasive among midwater dwellers, a limited set of alternative phenotypic solutions can meet the conditions necessary to thrive in pelagic habitats (i.e., many-to-one mapping). Outside Lutjanidae, remarkable departures from typical streamlined body shapes found in most oceanic pelagic vertebrates include the slow-swimming ocean sunfishes (*Mola*, *Ranzania*), which feature deep and laterally-compressed bodies—a clear case of historical contingency that is likely the result of developmental constraints. Although we did not examine diets in this study, a key factor that appears to trigger the invasion of the water column is the trophic adaptation to planktivory (Frédérich and Santini, 2017), likely as a result of competition for resources in benthic habitats. Morphological convergence has been reported in many groups that share specialized dietary shifts to planktivory (e.g., butterflyfishes, wrasses, angelfishes, damselfishes, and sea basses; Friedman et al., 2016; Langerhans and Reznick, 2010; Velotta et al., 2018). Ecological opportunity for the exploitation of different resources has thus repeatedly promoted morphological and behavioral adaptations associated with water-column transitions (Cooper et al., 2017; Floeter et al., 2018).

In conclusion, we find strong evidence of evolutionary convergence in major traits related to body elongation and fin morphology as a result of ecological transitions into pelagic habitats, ultimately reinforcing the deterministic role of evolution driven by similar ecological pressures. Our research shows incursions into the water column that are strongly linked to patterns of evolutionary convergence in body plans. We also have identified asymmetric habitat transitions and slower rates of lineage diversification associated with incursions into midwater habitats. The fact that these independent transitions took place in all major biogeographic regions further reinforces the deterministic nature of evolution. While conver-

gent evolution associated with the adoption of the pelagic lifestyle has governed the mode of diversification in Lutjanidae, future work should consider whether this conclusion can be generalized to support other habitat transitions along the benthic-pelagic axis as a primary mechanism of diversification in fishes.

References

Adams, D. C. and Collyer, M. L. (2019). Phylogenetic comparative methods and the evolution of morphological integration. *Annual Review ofEcology, Evolution, and Systematics*, (25):50:405.

Agnarsson, I., Avilés, L., Coddington, J. A., and Maddison, W. P. (2006). Sociality in Theridiid Spiders: Repeated Origins of an Evolutionary Dead End. *Evolution*, 60(11):2342.

Agrawal, A. A. (2017). Toward a Predictive Framework for Convergent Evolution: Integrating Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life*. *The American Naturalist*, 190(S1):S1–S12.

Alfaro, M. E., Faircloth, B. C., Harrington, R. C., Sorenson, L., Friedman, M., Thacker, C. E., Oliveros, C. H., Černý, D., and Near, T. J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary.

Allen, G.-R. (1985). *FAO species catalogue Vol.6. Snappers of the world. An annotated and illustrated catalogue of lutjanid species known to date*. Rome.

Andrews, K. R., Moriwake, V. N., Wilcox, C., Grau, E. G., Kelley, C., Pyle, R. L., and Bowen, B. W. (2014). Phylogeographic Analyses of Submesophotic Snappers Etelis coruscans and Etelis " marshi " (Family Lutjanidae) Reveal Concordant Genetic Structure across the Hawaiian Archipelago. *PLoS ONE*, 9(4).

Arbuckle, K., Bennett, C. M., and Speed, M. P. (2014). A simple measure of the strength of convergent evolution. pages 685–693.

Bellwood, D. R., Van Herwerden, L., and Konow, N. (2004). Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). *Molecular Phylogenetics and Evolution*, 33(1):140–155.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton, J. C., Zhang, F., Buser, T., Campbell, M. a., 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., Betancur-R., R., Broughton, R. E.,

Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. a., Ballesteros, J. a., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, Apr 18(APR 2013):1–45.

Blount, Z. D., Lenski, R. E., and Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science (New York, N.Y.)*, 362(6415).

Burns, M. D. and Sidlauskas, B. L. (2019). Ancient and contingent body shape diversification in a hyperdiverse continental fish radiation. *Evolution*, 73(3):569–587.

Burress, E. D., Holcomb, J. M., Tan, M., and Armbruster, J. W. (2017). Ecological diversification associated with the benthic-to-pelagic transition by North American minnows. *Journal of Evolutionary Biology*, 30(3):549–560.

Carpenter, K. (1987). Revision of the Indo-Pacific fish family Caesionidae (Lutjanoidea), with descriptions of five new species. *Indo-Pacific Fishes*, 15(56).

Carpenter, K. (2001). Caesionidae: Fusiliers. In *The Living Marine Resources of the Western Central Pacific*, pages 2919–2941. Rome, fao edition.

Carpenter, K. E. (1988). Vol.8. Fusiliers fishes of the world. In *FAO Species Catalogue*. Rome.

Carpenter, K. E. (1990). A Phylogenetic Analysis of the Caesionidae (Perciformes : Lutjanoidea). *Copeia*, (3):692–717.

Carpenter, K. E. (1993). Optimal cladistic and quantitative evolutionary classifications as illustrated by fusilier fishes (Teleostei: Caesionidae). *Syst Biol*, 42(2):142–154.

Clabaut, C., Bunje, P. M., Salzburger, W., and Meyer, A. (2007). Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, 61(3):560–578.

Clavel, J., Escarguel, G., and Merceron, G. (2015). mvmorph : an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution*, 6(11):1311–1319.

Cooper, W. J., Carter, C. B., Conith, A. J., Rice, A. N., and Westneat, M. W. (2017). The evolution of jaw protrusion mechanics is tightly coupled to bentho-pelagic divergence in damselfishes (Pomacentridae). *Journal of Experimental Biology*, 220(4):652–666.

Cooper, W. J., Parsons, K., McIntyre, A., Kern, B., McGee-Moore, A., and 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*.

Dunlap, M., Everson, A., Jayewardene, D., Makaiau, J., Mitsuyasu, M., Pautzke, S., and Walker, R. (2016). Appendix 3 Essential Fish Habitat Species Descriptions Part 1: Hawaiian Bottomfish. In *Fishery Ecosystem Plan for the Hawaii Archipielago*.

Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., and Meyer, A. (2014). Parallel evolution of nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature communications*, 5(1):1–8.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1427–1434.

Felsenstein, J., Ackerly, D. D., and Mcpeek, M. A. (2012). A comparative method for both discrete and continuous characters using the threshold model. *American Naturalist*, 179(2):154–56.

Fitzjohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3(6):1084–1092.

Floeter, S. R., Bender, M. G., Siqueira, A. C., and Cowman, P. F. (2018). Phylogenetic perspectives on reef fish functional traits. *Biological Reviews*, 93(1):131–151.

Frédérich, B. and Santini, F. (2017). Macroevolutionary analysis of the tempo of diversification in snappers and fusiliers (Percomorpha: Lutjanidae). *Belgian Journal of Zoology*, 147(1):17–35.

Fricke, R., Eschmeyer, W. N., van der Laan, R., Fricke, R., and van der Laan, R. (2016). Catalog of fishes: genera, species, references.

Friedman, M. (2010). Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society B: Biological Sciences*, 277(1688):1675–1683.

Friedman, S., Price, S., Corn, K., Larouche, O., Martinez, C., and Wainwright, P. (2020). Body shape diversification along the benthic–pelagic axis in marine fishes. *Proceedings of the Royal Society B*, 287(1931):20201053.

Friedman, S. T., Price, S. A., Hoey, A. S., and Wainwright, P. C. (2016). Ecomorphological convergence in planktivorous surgeonfishes. *Journal of Evolutionary Biology*, 29(5):965–978.

Froese, R. and Pauly, D. (2019). FishBase.

Goldberg, E. E., Kohn, J. R., Lande, R., Robertson, K. A., Smith, S. A., and Igić, B. (2010). Species selection maintains self-incompatibility. *Science*, 330(6003):493–495.

Haight, W., Parrish, J., and Hayes, T. (1993). Feeding Ecology of Deepwater Lutjanid Snappers at Penguin Bank, Hawaii. *Transactions of the American Fisheries Society*, 122(3):328–347.

Hobson, E. (1974). Feeding relationships of teleostean fishes on coral reefs in kona. *Hawaii. Fish. Bul1., US*, 72:9l5–1031.

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., Betancur, R., Li, C., Becker, L., Bellora, N., Zhao, X., Li, X., Wang, M., Fang, C., Xie, B., Zhoui, Z., Huang, H., Chen, S., Venkatesh, B., and Shi, Q. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences*, 115(24):201719358.

Hughes, L. C., Ortí, G., Saad, H., Li, C., White, W. T., Baldwin, C. C., Crandall, K. A., Arcila, D., and Betancur-R., R. (2020). Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *bioRxiv*, page 2020.02.18.949735.

Hulsey, C. D., Roberts, R. J., Loh, Y.-H. H. E., Rupp, M. F., and Streelman, J. T. (2013). Lake Malawi cichlid evolution along a benthic/limnetic axis. *Ecology and Evolution*, 3(7):2262–2272.

Ingram, T. and Mahler, D. L. (2013). SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425.

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*.

Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 7(7):811–824.

Kuiter, R. and Tonozuka, T. (2001). *Pictorial guide to Indonesian reef fishes. Part 2. Fusiliers - Dragonets, Caesionidae - Callionymidae*. Australia.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3):772–773.

Langerhans, R. B. and Reznick, D. N. (2010). Ecology and evolution of swimming performance in fishes: predicting evolution with biomechanics. *Fish locomotion: an eco-ethological perspective*, 200:248.

Lloyd, J. A. (2006). *Demography of Pristipomoides multidens in northern Australia and a comparison within the Family Lutjanidae with respect to depth*. PhD thesis.

Lobato, F. L., Barneche, D. R., Siqueira, A. C., Liedke, A. M., Lindner, A., Pie, M. R., Bellwood, D. R., and Floeter, S. R. (2014). Diet and diversification in the evolution of coral reef fishes. *PLoS ONE*, 9(7):1–11.

Losos, J. B. (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in Comparative Biology. *The American Naturalist*, 177(6):709–727.

Losos, J. B., Jackman, T. R., Larson, A., De Queiroz, K., and Rodríguez-Schettino, L. (1998). Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards. *Science*, 279(5359):2115–2118.

Lu, G. and Bernatchez, L. (2011). Correlated Trophic Specialization and Genetic Divergence in Sympatric Lake Whitefish Ecotypes (Coregonus clupeaformis): Support for the Ecological Speciation Hypothesis Author (s): Guoqing Lu and Louis Bernatchez CORRELATED TROPHIC SPECIALIZATION AND. *Society*, 53(5):1491–1505.

Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a Binary Character's Effect on Speciation and Extinction. *Systematic Biology*, 56(5):701–710.

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

McGhee, G. R. (2011). *Convergent evolution: limited forms most beautiful*. MIT Press.

Meyer, B., Meusemann, K., and Misof, B. (2011). MARE v0.1.2-rc.

Miller, T. L. and Cribb, T. H. (2007). Phylogenetic relationships of some common Indo-Pacific snappers (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. *Molecular Phylogenetics and Evolution*, 44(1):450–460.

Mirarab, S. and Warnow, T. (2015). ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31:44–52.

Newman, S. (2009). First record of Randall's snapper Randallichthys filamentosus (Perciformes: Lujanidae) from the eastern Indian Ocean (north-western Australia). *Journal of Fish Biology*, 75.

Pavlidis, M. A. and Mylonas, C. C. (2011). *Sparidae: Biology and Aquaculture of Gilthead Sea Bream and Other Species*. Wiley-Blackwell, 1st editio edition.

Peres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, 49(4):974–997.

Pigot, A. L., Sheard, C., Miller, E. T., Bregman, T. P., Freeman, B. G., Roll, U., Seddon, N., Trisos, C. H., Weeks, B. C., and Tobias, J. A. (2020). Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology and Evolution*, 4(2):230–239.

Præbel, K., Knudsen, R., Siwertsson, A., Karhunen, M., Kahilainen, K. K., Ovaskainen, O., Østbye, K., Peruzzi, S., Fevolden, S.-E., and Amundsen, P.-A. (2013). Ecological speciation in postglacial e uropean whitefish: rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and evolution*, 3(15):4970–4986.

Price, S. A., Friedman, S. T., Corn, K. A., Martinez, C. M., Larouche, O., and Wainwright, P. C. (2019). Building a Body Shape Morphospace of Teleostean Fishes. *Integrative and Comparative Biology*, pages 1–15.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., and Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714):392–395.

Rabosky, D. L. and Goldberg, E. E. (2017). FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution; international journal of organic evolution*.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution*, 63(12):3258–3268.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223.

Ribeiro, E. D., Davis, A. M., Rivero-Vega, R. A., Ortí, G., Betancur-R, R., Ortı, G., and Betancur-R, R. (2018). Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes. *Proceedings of the Royal Society B: Biological Sciences*, 285(1893):20182010.

Robertson, D. and Allen, G.-R. (2015). *Shorefishes of the Tropical Eastern Pacific: online information system.* Balboa, version 2. edition.

Robertson, D. R. and Van Tassell, J. (2016). *Shorefishes of the Greater Caribbean: online information system.* Balboa, version 1. edition.

Rundle, H. D., Nagel, L., Boughman, J. W., and Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science*, 287(5451):306–308.

Sallan, L., Friedman, M., Sansom, R. S., Bird, C. M., and Sansom, I. J. (2018). Vertebrate Diversification. *Science*, 464(October):460–464.

Sanciangco, M. D., Carpenter, K. E., and Betancur-R, R. (2016). Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). *Molecular Phylogenetics and Evolution*, 94:565–576.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology*, 57(5):758–71.

Taquet, M. and Diringer, A. (2013). *Fishes of the Indian Ocean and Red Sea*. Editions Quae, 1st editio edition.

Tavera, J., Acero P., A., and Wainwright, P. C. (2018). Multilocus phylogeny, divergence times, and a major role for the benthic-to-pelagic axis in the diversification of grunts (Haemulidae). *Molecular Phylogenetics and Evolution*, 121(December 2017):212–223.

Tavera, J. J., Acero P, A., Balart, E. F., and Bernardi, G. (2012). Molecular phylogeny of grunts (Teleostei, Haemulidae), with an emphasis on the ecology, evolution, and speciation history of New World species. *BMC Evolutionary Biology*, 12(1):57.

Velotta, J. P., McCormick, D., and Jones, A. W. (2018). Reduced Swimming Performance Repeatedly Evolves on Loss of Migration in Landlocked Populations of Alewife. *Physiological and Biochemical Zoology*, 91(2):814–825.

Wake, D. B., Wake, M. H., and Specht, C. D. (2011). Homoplasy: From Detecting Pattern to Determining Process and Mechanism of Evolution. *Science*, 331(6020):1032–1035.

Walker, J. A. (1997). Ecological morphology of lacustrine threespine stickleback Gasterosteus aculeatus L. (Gasterosteidae) body shape. *Biological Journal of the Linnean Society*, 61(1):3–50.

White, W. T., Last, P. R., Dharmadi, Faizah, R., Chodrijah, U., Prisantoso, B. I., Pogonoski, J. J., Puckridge, M., and Blader, S. J. M. (2013). *Market Fishes of Indonesia*.

Willacker, J. J., Von Hippel, F. A., Wilton, P. R., and Walton, K. M. (2010). Classification of threespine stickleback along the benthic-limnetic axis. *Biological Journal of the Linnean Society*, 101(3):595–608.

Chapter 2

Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes

Published in *Proceedings of the Royal Society B* (https://doi.org/10.1098/rspb.2018.2010)

Emanuell Duarte-Ribeiro, Aaron M. Davis, Rafael A. Rivero-Vega, Guillermo Ortí, Ricardo Betancur-R

2.1 Abstract

Ecological opportunity arising in the aftermath of mass extinction events is thought to be a powerful driver of evolutionary radiations. Here, we assessed how the wake of the Cretaceous-Paleogene (K-Pg) mass extinction shaped diversification dynamics in a clade of mostly marine fishes (Carangaria), which comprises a disparate array of benthic and pelagic dwellers including some of the most astonishing fish forms (e.g., flatfishes, billfishes, remoras, archerfishes). Analyses of lineage diversification show time-heterogeneous rates of lineage diversification in carangarians, with highest rates reached during the Paleocene. Likewise, a remarkable proportion of Carangaria's morphological variation originated early in the history of the group and in tandem with a marked incidence of habitat shifts. Taken together, these results suggest that all major lineages and body plans in Carangaria originated in an early burst shortly after the K-Pg mass extinction, which ultimately allowed the occupation of newly released niches along the benthic-pelagic habitat axis.

2.2 Introduction

Patterns of initial bursts of diversification in the origin and propagation of high level taxa are usually explained by Simpsonian theory on adaptive radiation—one where a rapidly proliferating lineage evolves into great ecological diversity and morphological disparity as a result of increased availability of resources and limited competition (Schluter, 2000; Simpson, 1953). These ecological opportunities may occur when previously inaccessible resources become available via acquisition of key evolutionary innovations, colonization of new areas, or removal of competitors owing to external mechanisms of habitat depauperation (Simpson, 1953).

The formation of vacant niches in the wake of mass extinction events is a major source of ecological opportunity (Erwin, 2001). Among the five major mass extinctions in the history of life on Earth, the Cretaceous-Paleogene (K-Pg) event (ca. 65 Ma) is thought to have triggered parallel rapid radiations in numerous tetrapod clades including amphibians and placental mammals (Feng et al., 2017; Pollux et al., 2014). In Actinopterygii (ray-finned fishes), the proportion of incumbent diversity that became extinct by end of the Mesozoic is high at the family level (19%; Cavin, 2002) and likely represented an important source of ecological opportunity that modulated diversification dynamics in surviving acanthomorph lineages (spiny-rayed teleost fishes; a subclade of ray-finned fishes; Cavin, 2002; Friedman, 2009, 2010). Indeed, the stratigraphic distribution of acanthomorph fossils suggests significant restructuring of marine fish communities in the aftermath of the K-Pg (Patterson, 1993), which further coincides with the expansion of the group's morphological disparity in areas of the ecospace emptied by the extinction of their (non-acanthomorph) teleost counterparts (Friedman, 2010). In the wake of the K-Pg, fishes also experienced a pronounced increase in abundance relative to sharks, which was presumably spurred by ecological release and ultimately prompted what has come to be known as the "new age of fishes" (Sibert and Norris, 2015).

In agreement with the fossil evidence, time-calibrated phylogenetic trees also reveal patterns in which the origin of several major acanthomorph subclades chronologically overlap with the K-Pg boundary (Alfaro et al., 2018; Betancur-R et al., 2013; Harrington et al., 2016; Near et al., 2013). One such clade featuring an explosive pulse of diversification near the K-Pg is the Carangaria (Alfaro et al., 2018), a diverse group with over a thousand species that includes a disparate array of benthic (e.g., flatfishes, threadfins) and pelagic (e.g., billfishes, remoras, barracudas) fish dwellers. The Carangaria also encompasses some of the most extreme morphological and ecological adaptations in vertebrates, including the asymmetric body plan of flatfishes, the endothermic body heat regulation in marlins, billfishes, and swordfishes, and the hunting behaviour of archerfishes, which generate bullets of water to feed on terrestrial prey. It thus appears that the greatest phenotypic diversity in Carangaria

is associated with the early evolution of disparate morphologies that prompted adaptation to a broad array of habitats, including clear instances of adaptive peaks lying at the extremes of the benthic-pelagic spectrum in fishes (e.g., open-water billfishes and substrate-burrowing flatfishes).

Here, we investigate the dynamics of diversification, phenotypic evolution and habitat transitions in Carangaria. Based on the above observations, we hypothesize that lineage diversification varies as a function of time, with high rates reached near the clade's origin (at the Mesozoic-Cenozoic boundary) followed by a rapid drop as the carrying capacity of species diversity is reached. Furthermore, in agreement with Simpsonian predictions on adaptive radiation, we also expect that an initial expansion of morphological disparity would be subsequently replaced by a period of morphospace packing as niches become filled. To test these ideas, we estimated a multi-locus time tree that includes all major lineages of carangarians and used a suite of recently-developed phylogenetic comparative approaches to assesses how rates of lineage diversification, multivariate phenotypic evolution, and habitat transitions vary throughout the clade's history.

2.3 Material and Methods

Taxonomic sampling and phylogenetic inference

Carangaria's diversity is represented by a sample of 125 (out of ca. 1,100) species, including representatives from 26 valid families out of 28 (only Lactariidae and Paralichthodidae were not examined) and over half of the genus-level diversity (95 out of 187) in the group. This taxonomic sampling strategy comprises a diversified scheme designed to maximize both phylogenetic and eco-morphological diversity within Carangaria, under the assumption that missing lineages are phylogenetically and eco-morphologically nested within the sampled ones.

The molecular dataset is based on a recent study that generated multi-locus sequences from 20 nuclear loci (19,461 sites; Betancur-R. et al., 2013). We expanded the molecular matrix to incorporate 10 additional outgroup species that represent major acanthomorph lineages. We used BEAST v.1.8.4 (Drummond et al., 2012) to simultaneously estimate topology and divergence times using a set of 16 fossil-based calibration points (based on Harrington et al., 2016). Lower bounds of clade ages were defined via minimum age of earliest fossil representatives; 95% soft upper bounds were estimated based on maximum ages of the oldest fossils assigned to successive outgroups for each clade. Convergence of analyses was assessed after conducting two independent runs of 300 million generations each. To account for phylogenetic uncertainty, one hundred trees were evenly sampled from the posterior distribution, providing a robust framework for downstream macroevolutionary analyses.

Detailed taxonomic sampling, phylogenetic analyses (Figure B.1), and fossil calibration information is provided in Appendix B. To further address potential phylogenetic uncertainties, we repeated some analyses using alternative trees for Carangaria estimated by other recent studies (Harrington et al., 2016; Rabosky et al., 2018, see below).

Body-shape data

The laterally-compressed body plan of most carangarians makes this group well-suited for the summarization of multivariate morphological axes using two-dimensional geometric morphometric approaches. We carefully assembled a specimen imagery database for 116 out of the 125 carangarian species in our tree. The database consists of digitized specimens from museum collections or curated images retrieved from online repositories (Appendix B). We selected 15 landmarks that are extensively used to summarize general body shape variation in percomorphs (Chakrabarty, 2005), including both type I and type II points (Figure B.2). Whereas type I landmarks are strictly homologous points, type II landmarks include points whose homology is supported by geometric evidence rather than histological data and are frequently used to describe inflexion points such as the sharpest curvature of a tooth or tips of caudal fin lobes (Bookstein, 1997). We used tpsDig2 (Rohlf, 2001) to place the landmarks and summarized the extant species' body shape diversity using Procrustes superimposition and Principal Component Analyses (PCA), as implemented in the R package geomorph (Adams and Otárola-Castillo, 2013). Next, we subjected the morphological data to a phylogenetically-corrected principal component analysis (pPCA) to account for possible distortions of the PCA arising from phylogenetic non-independence.

Tempo and mode of lineage diversification

We assessed time variation in lineage diversification rates in Carangaria using CoMET, a Bayesian statistical model implemented in the R package TESS (Höhna et al., 2015). CoMET estimates the number of lineage diversification rate shifts along with their timing and rate parameters (i.e., speciation and extinction rates). We used TESS' Bayes factor model selection to explicitly test the relative fit of the following series of alternative branching models to our comparative dataset: (1) time-homogeneous birth-death, (2) continuously-decreasing-rate birth-death, and (3) episodically-varying-rate that incorporates the diversification parameters (i.e., number and timing of episodic rate-shifts) obtained with CoMET. Model comparison analyses were applied for both the maximum clade credibility (MCC) tree and 100 trees sampled from the posterior distribution. To accommodate biases inherent to incomplete taxonomic sampling, we applied a diversified sampling strategy correction (for both CoMET and model-fitting), which is appropriate in cases where taxonomic sampling is designed to maximize phylogenetic diversity (Hohna, 2014). To further account for potential biases as-

sociated with incomplete taxonomic sampling and phylogenetic uncertainties, we conducted posterior-predictive simulation tests using our MCC tree and a set of independently estimated time trees that incorporate different taxonomic sampling schemes ranging from 5% of the clades diversity (Harrington et al., 2016) to a nearly complete taxonomic sampling (Rabosky et al., 2018), Appendix B.

Tempo and mode of morphological evolution

To assess the tempo and mode of morphological evolution in Carangaria, we initially estimated how multivariate morphological disparity accumulated through time using ancestral state reconstructions derived from rate-heterogeneous models of continuous trait evolution (Cooney et al., 2017; Revell, 2012). We estimated ancestral state values using the maximum likelihood (ML) fastAnc function implemented in the R package phytools (Revell, 2012). To account for unequal rates of evolution among the different shape axes, we estimated ancestral values for each pPC separately using rate-transformed trees in which branch lengths depict the rate of morphological change. Rate-transformed trees were estimated in BayesTraits (available from http://www.evolution.rdg.ac.uk/) using default priors. Multivariate morphological disparity was then calculated as the sum of the variances across the different pPC axes in time-slices of one million-year (Myr). We compared the observed disparity against 500 curves of disparity accumulation simulated under a constant-rate Brownian motion (BM) null model of continuous trait evolution. Because of the non-directional nature of trait change simulated using this BM model, we expect the underlying balance between morphospace expansion and packing to be effectively equal and constant over time. Thus, any period of time that shows substantial deviations in the observed patterns of morphological disparity accumulation (compared to the simulated null trajectories) would indicate that one of the processes (either morphospace expansion or packing) dominated over the other.

We also explicitly assessed the fit of alternative evolutionary models of body shape evolution in a ML framework using the R package mvMORPH (Clavel et al., 2015). We first fitted three alternative models of single-mode continuous-trait evolution: (1) a single rate BM model, (2) a single regime Orstein-Uhlenbeck (OU) model, and (3) an early burst (EB) of morphological evolution. Given that shifts in the mode of evolution may provide a more realistic explanation for the processes generating morphological disparity (Slater, 2013), we further considered three additional models in which processes generating disparity shift episodically: (4) EB to independent-rates OU shift, (5) BM to independent rates OU shift, and (6) EB to independent-rates BM shift. We used the function mvSHIFT, which can fit models of trait change within a mode of evolution after a fixed point. We modelled post-shift independent-rates OU and BM models by allowing the drift parameter to vary after a fixed point. A temporal shift window of 46 Ma was selected based on the time of

transition between stages of morphospace expansion and packing, as obtained by comparing the observed and simulated trajectories of morphological disparity (see Results).

It has been suggested that limiting macroevolutionary analyses to a narrow subset of shape dimensions (i.e., first few principal component axes) may produce erroneously strong support to more complex models, such as early burst (EB; Uyeda et al., 2015). More recently, mvMORPH has been shown to produce misleading results when the N:p ratio is sufficiently low (where N is the number of species and p the number of traits). It should be noted, however, that an adequate N:p ratio level required to confidently assess the fit of alternative models using mvMORPH is still elusive (Adams et al., 2019). To account for possible biases regarding the use of multivariate shape data, we performed analyses using two different trait subsets selected according to the proportion of body-shape variance summarized (using 5% and 1% thresholds; see Results). The 5% subset was run using both the MCC tree and a set of 100 trees drawn from the posterior distribution; due to computational limitations, the 1% subset was run using the MCC tree only.

Tempo and mode of ecological diversification

To evaluate whether the rate of habitat transitions varied as a function of time, we first assigned species in Carangaria into three major habitat categories: benthic (bottom-dwellers), pelagic, and benthopelagic (intermediate habitat states; species that swim just above the bottom; Helfman et al., 2009). The habitat occupancy dataset was compiled by aggregating information from a wide range of sources, including FishBase (Froese and Pauly, 2016), Catalog of Fishes (Fricke et al., 2016), and the primary literature. We then used SIMMAP stochastic mapping, as implemented in the R package phytools (Revell, 2012), to reconstruct the history of trait changes in our MCC tree and to estimate the rate of habitat transitions from root through present (i.e., number of transitions divided by the total edge length in 5 Myr time slices). Three hierarchical transition models—equal rates (ER), symmetrical rates (SYM) or all rates different (ARD)—were assessed by ML with results averaged across all runs; the best-fitting model for SIMMAP was identified using likelihood ratio tests. We also explicitly tested the relative fit of models of discrete character evolution to identify the best explanation concerning the distribution of events of habitat transition throughout the evolutionary history of carangarians. We used fitDiscrete as implemented in the R package Geiger (Harmon et al., 2008) to assess the fit of two contrasting models of discrete trait evolution: (1) a constant-rate model, and (2) an EB of discrete trait diversification. For this approach, we used the same set of 100 trees sampled from the Bayesian posterior distribution (see above), as well as the same set of independently estimated time trees that incorporate different taxonomic sampling schemes (Harrington et al., 2016; Rabosky et al., 2018).

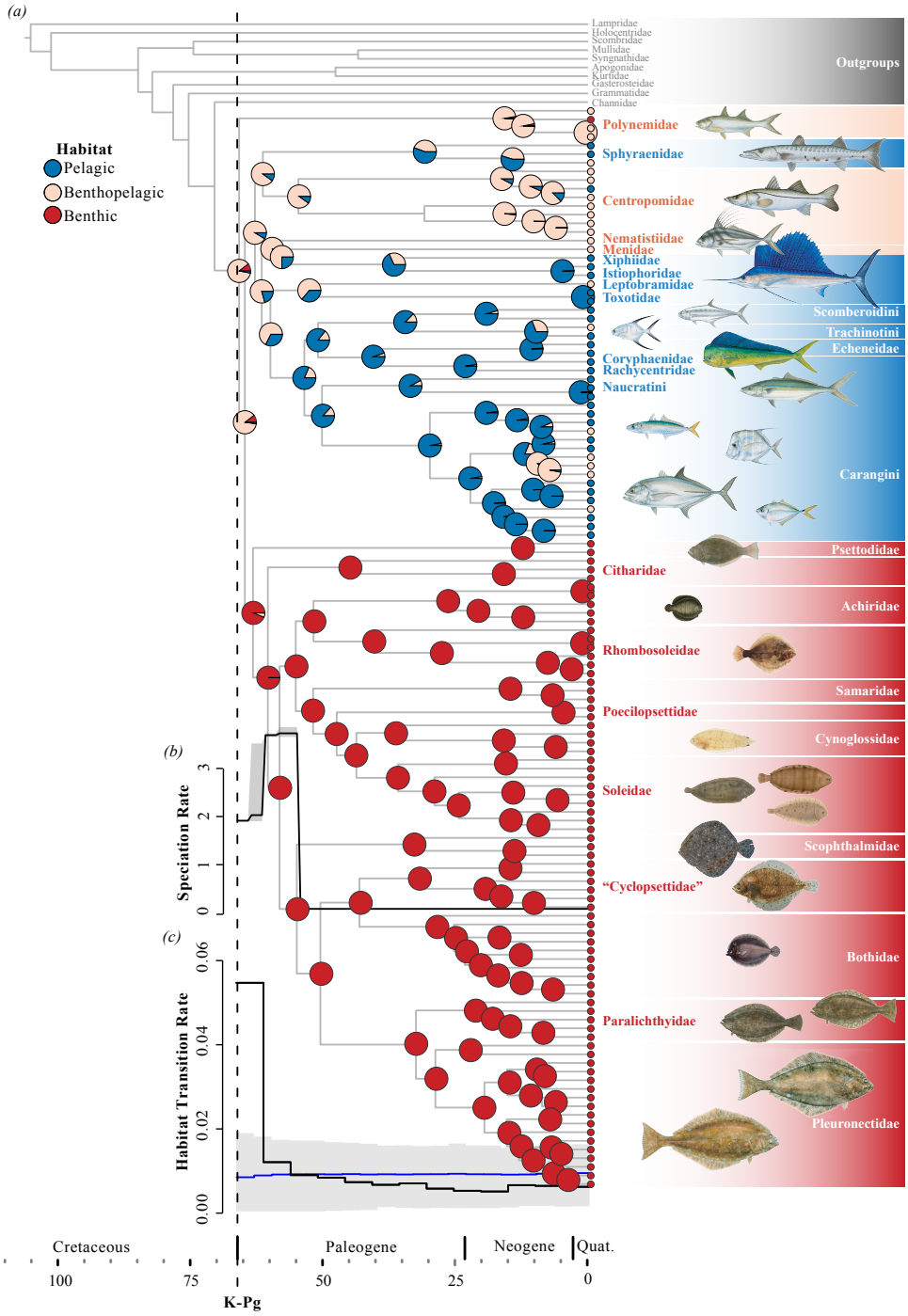

Figure 2.1: (a) Maximum clade credibility (MCC) time tree estimated for Carangaria, including pie charts for ancestral habitat reconstructions. (b) Rates of speciation through time estimated from the MCC tree using the COMET function in TESS. (c) Rates of habitat transitions through time, estimated as number of transitions divided by the total edge length in 5 Myr time slices; blue line indicates constant rate Brownian motion model. Dashed line indicates the time of the Cretaceous–Palaeogene (K-Pg) mass extinction. Most fish illustrations are reproduced with permission from Diane Rome Peebles

2.4 Results

Phylogenetic reconstruction and divergence times

Trees, datasets and R code used for comparative analyses are available from the Figshare digital repository (doi:10.6084/m9.figshare.5727096). The inferred tree (Figure 2.1, B.1) is largely congruent with previous multi-locus analyses of Carangaria (Betancur-R et al., 2013; Harrington et al., 2016), although placement for some lineages (e.g., barracudas and threadfins) along the backbone often varies due to the rapid nature of speciation events at the onset of the group's evolution. Divergence time estimates are likewise concordant with the age intervals derived from a recent study based on a phylogenomic analysis for 45 species in Carangaria (Harrington et al., 2016), as well from previous estimates using multi-locus datasets and multiple calibration points across the fish diversity (Betancur-R et al., 2013, 2017; Near et al., 2012, 2013). The evolutionary timescale inferred suggests that the origin of major carangarian lineages took place close to the Cretaceous-Paleogene boundary, with a mean clade age ranging from 71 Ma (total group; 95% HPD 78–64 Ma) to 66 Ma (crown group; 95% HPD 72–61 Ma) (Betancur-R et al., 2013; Harrington et al., 2016).

Body-shape data

The first four Principal Component (PC) axes accounted for more than 85% of the total shape variance and are presented as morphospace scatter plots in Figure 2.2. The PC1 (67% of overall shape variation) describes head morphology and the length of the dorsal and anal fin bases, features that have been identified as one of the major axes of evolution in acanthomorphs (Claverie and Wainwright, 2014; Friedman, 2010). In flatfishes in particular those characters are linked to some of the most extreme morphological adaptations experienced by vertebrates; i.e., the partial loss of bilateral symmetry arising from eye migration, and the dorsal advancement of median fins towards the cranium. Indeed, this is represented as a bimodal distribution of PC1 values that distinguish flatfishes from all non-pleuronectiform carangarians in our analyses (Figure 2.2a). The PC2 (7% of total variation) summarizes differences in body elongation, a major axis of shape variation in several fish clades (Claverie and Wainwright, 2014). The PC3 and PC4 (6% and 5% of overall variation, respectively) also encompass ecologically relevant aspects of fish morphology that are frequently represented in traditional morphometric measurements (caudal peduncle depth and snout length, respectively; Figure 2.2b). Subsequent PC axes explain lower proportions of body-shape variance. As noted in the Methods section, we selected two different trait subsets of pPC axes for downstream analyses (Adams and Collyer, 2018; Uyeda et al., 2015). The 5% and 1% subsets comprised the first 4 pPC (69% of total shape variation) and first 12 pPC (95% of total shape variation) axes, respectively.

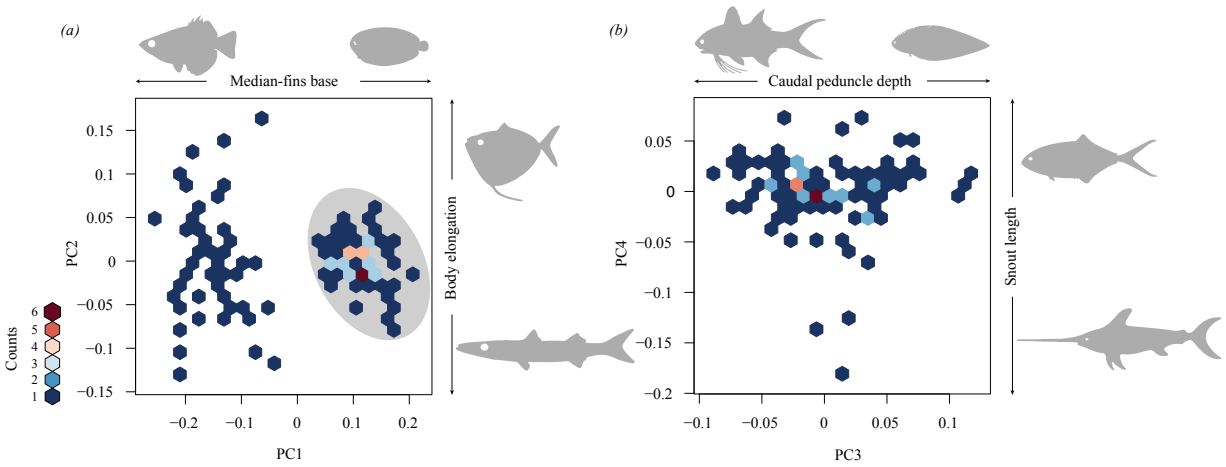

Figure 2.2: Body-shape morphospace in Carangaria. Fish silhouettes represent extreme values for each axis: (a) PC1 versus PC2 (shaded elliptical area represents the morphospace occupied by pleuronectifoms); (b) PC3 versus PC4. Colour scale indicates the number of species per hexagon.

Tempo and mode of lineage diversification

These analyses show a burst of lineage diversification rate at the onset of carangarian history, with post-Cretaceous rates decreasing abruptly after the Paleocene–Eocene boundary (56 Ma). CoMET results were rather inconsistent about the clade's evolutionary dynamics and highly sensitive to the choice of hyper-priors. However, one recurrent scenario—high initial speciation followed by a decline in speciation rates around 55 Ma (Figure 2.1b)—demonstrates the existence of a strong signal supporting a change in the diversification regime, a result that seems robust to analytical artefacts (Figure B.3). TESS' marginal likelihood model comparison showed a preference (BF >100) for variable-rates models (continuously-decreasing-rate birth-death and episodically-varying-rate) over a time-homogeneous birth-death mode for most pruned resampled trees, confirming our expectation that that time-homogeneous processes cannot explain lineage diversification dynamics in the group (Figure 2.3a-c). Moreover, comparisons between the two variable-rates models reveal that 99% of the trees provide decisive support (BF >100) for the episodically-varying rate model that incorporates one diversification rate-shift at 55 Ma. We obtained similar results for model-fit comparisons using the original set of resampled trees (Figure B.4). Finally, our results on lineage diversification appear to be robust to the utilization of alternative phylogenetic trees incorporating a broad array of taxonomic sampling schemes as well as to the implementation of posterior-predictive simulation tests (Figures B.5) or other simpler statistics (Figure B.6; see Appendix B for details).

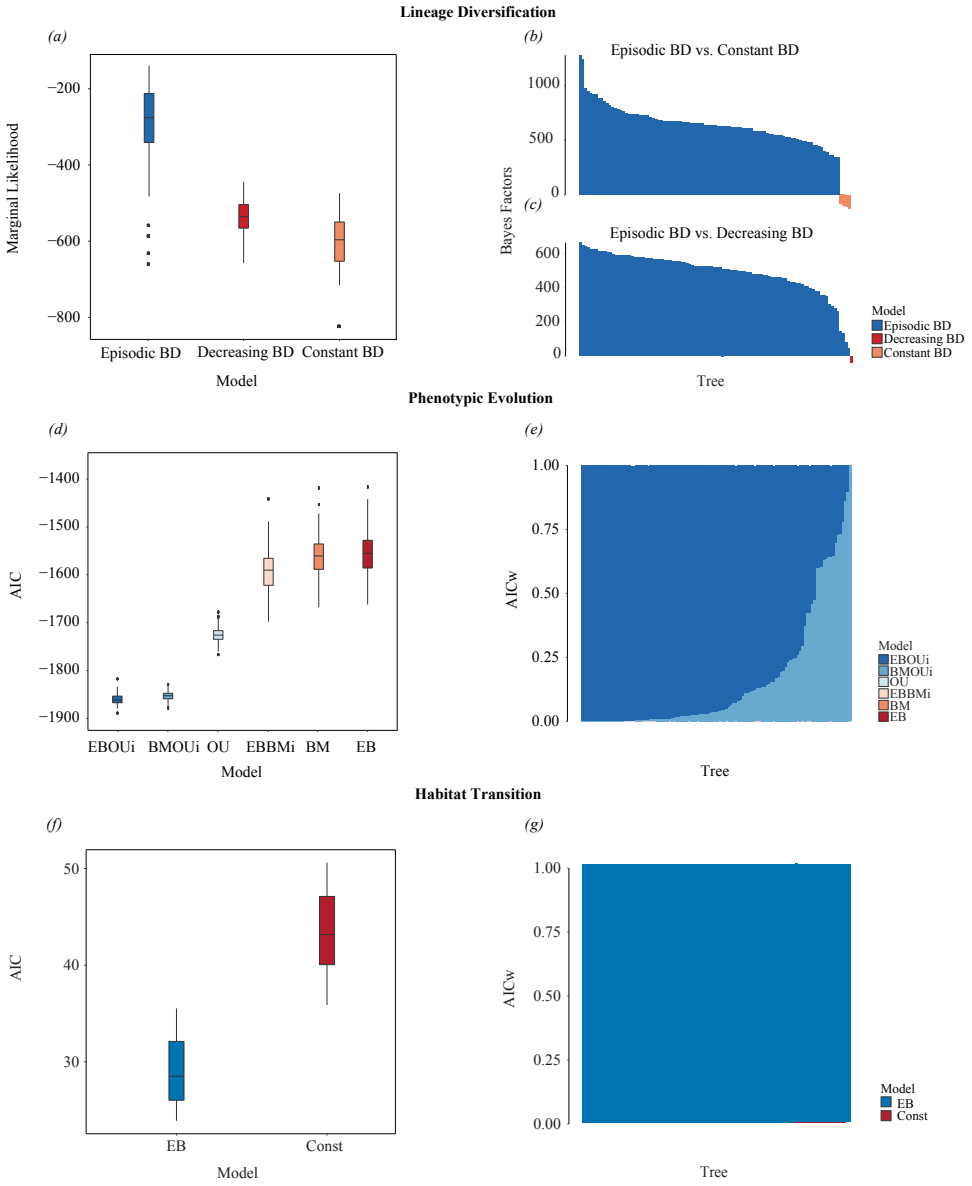

Figure 2.3: Model-fit comparisons based on a set of 100 trees evenly sampled from the posterior distribution. (a–c) Comparisons for alternative models of lineage diversification (using pruned versions of our 100 empirical that excludes recent cladogenetic events): (a) distribution of the marginal likelihood for the three alterna- tive branching models; (b) Bayes factors comparing episodic birth-death (BD) and constant BD models for the 100 resampled trees; and (c) Bayes factors comparing episodic BD and decreasing BD for the 100 resampled trees. (d,e) Comparisons of alternative models of morphological evolution using the 5% threshold trait subset: (d) distribution of the Akaike information criterion (AIC) values for the six alternative models of continuous trait evolution (EBOUi, shift from a single rate EB into a multiple independent rates OU; BMOUi, shift from a single rate BM into a multiple independent rates OU; OU, Ornstein-Uhlenbeck; EBBMi, shift from a single rate EB into a multiple independent rates BM; BM, Brownian motion; EB, early burst); (e) AIC weights of each alternative model based on each resampled tree. (f,g) Comparisons of alternative models of ecological evolution: (f) distribution of the AIC values for the two models of discrete trait evolution; (e) AIC weights of each alternative model based on each resampled tree.

Tempo and mode of morphological evolution

The multivariate disparity-through-time analyses reveal a remarkable proportion of the Carangaria morphological variation originating early in the clade's history. Both trait subsets analysed (4 and 12 pPCs) revealed similar patterns of morphological disparity accumulation, with 60% of the total variance arising before the Paleocene-Eocene boundary within a time interval of just 10 Myr (Figure 2.4a). This proportion is particularly relevant when compared to the total variance expected to be accumulated at 55 Ma under a BM null model of trait evolution (only 5%). Comparisons against a BM also indicate a dominance of morphospace expansion early in carangarian history (Figure 2.4c). This initial stage of accelerated morphological evolution lasted for about 20 Myr (until around 46 Ma) and was subsequently replaced by a period of morphospace packing, presumably reflecting a Simpsonian process of vacant niche filling. We also obtained congruent results for model fitting based on both pPC trait subsets (4 and 12), indicating that the results are robust to the number of shape axes included. A simple EB model of morphological evolution—rates slowing down exponentially through time—presented the worst fit among all the competing models (Akaike Weights or w_A <0.01). However, we found strong support (w_A >0.50 for 90% of the trees) for a time-heterogeneous model in which body-shape evolution switches from an initial EB into a random-walk with multiple and independent stationary peaks (Figure 2.3d,e). These independent peaks appear to represent different adaptive zones corresponding to major body-plans in Carangaria, with the random-walks showing the exploration of niche space after the transition from morphospace expansion to morphospace packing ca. 46 Ma.

Tempo and mode of ecological diversification

In agreement with results obtained for lineage diversification and morphological evolution, plots of habitat transition rates through time show that the distribution of ecological shifts in the group is notably uneven, with initial high rates that drop slightly before the Paleocene-Eocene boundary (Figures 2.1c). Moreover, comparisons of models of discrete character evolution indicate a decisive support for an early burst model in all resampled trees (w_A >0.99), suggesting that the skewed distribution of events of habitat transition would be plausibly explained by a model in which the rate of evolution decreases exponentially through time (Figure 2.3f-g). Similar results are obtained using taxonomically-denser trees for Carangaria.

2.5 Discussion

By implementing complex models of lineage, morphological, and ecological evolution, our study supports post-Cretaceous bursts of diversification as a probable explanation concerning

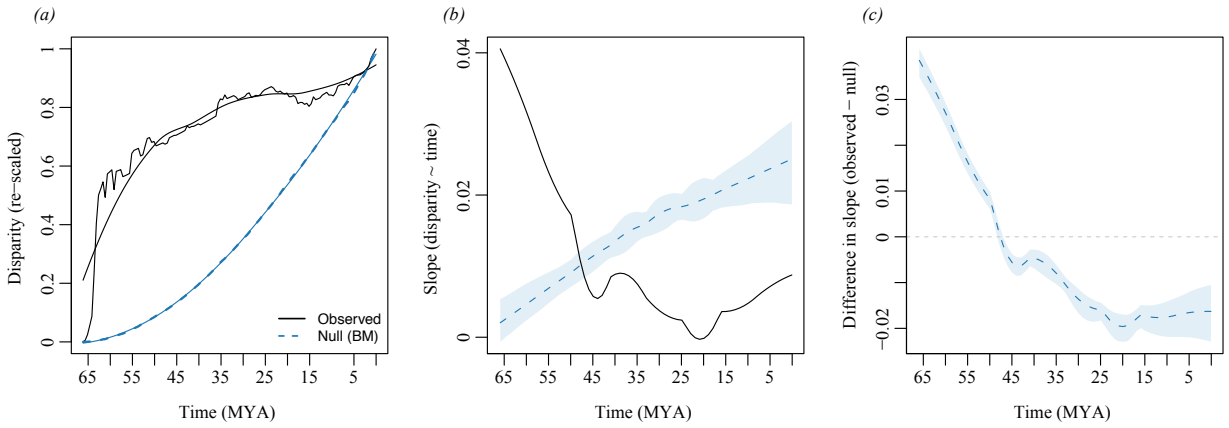

Figure 2.4: Disparity-through time plots showing the evolution of morphospace filling in Carangaria using the % threshold trait dataset (the highest 12 pPC axes). (a) Accumulation of multivariate disparity through time in 1 Myr time slices (thick black line, observed data; thin black line, after locally estimated scatterplot smoothing (LOESS) smoothing; blue line, constant rate Brownian motion (BM) null model). (b) Comparison of slopes for the two competing models; shaded areas represent 95% confidence intervals. (c) Differences in slope for the observed data and the BM null model; values above and below zero indicate the dominance of morphospace expansion versus morphospace packing, respectively.

the evolutionary trajectory of carangarians, aligning with observations from mammals, frogs and other tetrapod groups. It also highlights the apparent role of ecological release stemming from the extinction/absence of competitors in triggering ancient radiations along the benthic-pelagic axis—a well-characterized mode of diversification in recent groups of temperate fishes (e.g., sticklebacks, whitefishes; Hollingsworth et al., 2013) whose significance is otherwise poorly understood from a macroevolutionary perspective.

Initial assessments of time variation in the rates of lineage accumulation indicated a stage of high diversification during the Paleocene followed by a period of relative stasis towards the present. Recent efforts to assess time variation across major acanthomorph groups (as well as specific subclades) have failed to detect signatures of the K-Pg mass extinction in diversification rates (Arcila and Tyler, 2017; Near, 2013, but see Price et al., 2014). By contrast, our results reveal strong support for an uneven origination of species richness in Carangaria, with high rates of lineage diversification reached in the aftermath of the end-Cretaceous mass extinction.

Early cladogenetic events giving rise to all major lineages in Carangaria (e.g., flatfishes, billfishes, robalos, moonfishes, threadfins, remoras, jacks) are entirely restricted to the Paleocene, supporting the hypothesis that ecological opportunity arising in the wake of the K-Pg mass extinction prompted rapid radiation. Similar patterns are found in other specious acanthomorph fish clades. For instance, pelagiarians—a group comprising open-ocean fishes such as tunas, mackerels, and cutlassfishes—has likewise radiated in the aftermath of the K-Pg mass extinction, with most of its major lineages arising in the early-Paleogene (Miya et al.,

2013). Additionally, there are signs of rapid radiations in many species-rich reef-fish families, their early divergence in most cases also dates back to this time (e.g., wrasses, grunts, surgeonfishes, and blennies; Alfaro et al., 2018). The episodic decline in Carangaria's lineage diversification appears to coincide with global climatic changes during the Paleocene–Eocene Thermal Maximum (PETM; 55.8±0.2 Ma), a brief interval of extreme perturbation in the global carbon cycle that resulted in record-high levels of global warming (5° to 10°C; Röhl et al., 2000). Recent work hints that such severe environmental conditions affected reef-fish diversification dynamics (Price et al., 2014) and are potentially linked to clade-wide extinctions in the acanthomorph order Tetraodontiformes (ocean sunfishes, pufferfishes, and allies; Arcila and Tyler, 2017). A mode of classic niche filling, however, cannot be rejected as a plausible explanation for the decline in Carangaria's lineage diversification rates during the Paleocene-Eocene boundary.

Plots of multivariate disparity accumulation through time revealed a remarkable proportion of Carangaria's morphological variation originating early in the clade's history, with 60% of the total clade disparity being reached before the Paleocene-Eocene boundary. The notable dominance of morphospace expansion followed by a period of morphospace packing fits Simpsonian predictions on morphospace filling and reinforces the evolutionary role of the K-Pg mass extinction. This result is in line with the stratigraphic distribution of acanthomorph fossils, which reveals an early-Cenozoic expansion of the acanthomorph body-shape disparity in areas of the morphospace emptied after the mass extinction event (Friedman, 2010). Some of Carangaria's modern body-plans were already represented during the early Cenezoic. For instance, the late-Paleocene *†Mene purdyi* resembles the contemporary morphology of its congener, the moonfish (Friedman and Johnson, 2005). Other modern taxonomic groups, such as jacks, robalos, and stem flatfishes have also been documented from the Eocene (49 Ma) deposits of Monte Bolca in northern Italy (Bellwood, 1996).

The paleontological record is rich in evidence supporting the macroevolutionary trend of animal clades reaching high morphological disparity early in their evolutionary history (Hughes et al., 2013). However, phylogenetic comparative studies have challenged the relevance of early bursts in explaining the morphological evolution in well-established examples (though mostly younger) adaptive radiations (Harmon et al., 2010). While a simple EB model also proved to be a poor explanation for the dynamics of morphological evolution in Carangaria, the apparent incompatibility between the patterns of morphospace filling and the results of model-fit comparisons was reconciled by accounting for more realistic models that incorporate variation in the processes generating morphological evolution. We found unequivocal support for a model that incorporates a shift in the mode of evolution from EB into a multiple selective peak random walk at the recovered time of transition between morphospace expansion and packing (ca. 46 Ma). Although bursts of morphological evolution may be a common macroevolutionary feature, it has been demonstrated that our ability to

detect them would be affected by factors such as the ecological relevance of analysed traits (Cooney et al., 2017), the phylogenetic scale (Hopkins and Smith, 2015), and the use of overly simplistic evolutionary models (Slater, 2013). This latter factor is likely the source of conflict in our analyses.

An important prediction of the adaptive radiation theory is that both speciation and morphological adaptations must be significantly associated to the occupation of divergent environments (Schluter, 2000). Habitat transitions along the benthic-pelagic axis have had important outcomes in the diversification dynamics of relatively recent freshwater fish groups such as sticklebacks, whitefishes, cichlids, minnows, and perches (Burress et al., 2017; Hollingsworth et al., 2013), as well as grunts of the marine family Haemulidae (Tavera et al., 2018). However, the effects of the adoption of divergent ecological regimes remain largely unexplored at deeper macroevolutionary scales.

Ancestral state reconstructions revealed that carangarians experienced higher rates of habitat transitions along the benthic-pelagic axis during the Paleocene, notably overlapping with the early lineage diversification and morphological evolution in major clades. A notable example is the loss of the bilateral symmetry experienced by flatfishes, with laterally compressed bodies featuring both eyes on the same side of the head. Although dorsally-flattened (depressed) body plans are recurrent among benthic dwellers (e.g., rays, skates, suckermouth-armoured catfishes, and batfishes), flatfishes' laterally compressed plan is unusual among benthic-living species. Flatfishes further evolved other key adaptations to facilitate their burrowing into the substrate (e.g., the recessus orbitalis, a muscular sac that enables eye protrusion). Carangaria also comprises several clades that have invaded in parallel the pelagic realm, such as istiophoriforms (swordfish, marlins and billfishes), sphyraenids (barracudas), and many carangiforms (e.g., dolphinfishes, amberjacks, and the rainbow runner). Those open-water, fast-moving predators have convergently developed streamlined bodies, forked tail fins, and slender tail bases (caudal peduncle). In agreement with our observations, reconstructions of the trajectory of morphological evolution in the fossil record of acanthomorphs have shown that a major component of the early-Cenozoic morphospace expansion reflected a process of ecological replacement of the pelagic non-acanthomorph fauna that became extinct by the end of the Mesozoic (Friedman, 2009, 2010).

Our results also show that the uneven distribution of habitat transitions events is temporally associated with the asymmetric accumulation of species richness, with an initial stage of high rates of ecological transitions that subsequently slows down, aligning with expectations under a BM model during the Paleocene-Eocene boundary. In agreement with this pattern, comparisons of the fit of alternative models of discrete trait evolution strongly support an exponential decrease in the rates of habitat transitions. Taken together, our results highlight the importance of the adoption of divergent ecological regimes in the origin and recovery of marine fish clades in the wake of mass extinction.

While we find bursts of evolution in carangarians to be chronologically associated with post-Cretaceous ecological release, we recognize that other possible sources of ecological opportunity have likely played an important role in their radiation during the early Cenozoic. For instance, the loss of bilateral symmetry in flatfishes and the elongation of the premaxilla bone in billfishes rank among the most extreme phenotypic adaptations in vertebrates and are candidates for key innovations that may have created additional sources of ecological opportunity in these subclades. Colonization of novel habitat regimes is another source of adaptive radiation that we here show operated in synchrony with newly-released niches in the wake of the K-Pg, allowing the diversification of carangarians along the benthic-pelagic axis. Although decoupling the relative importance of these different sources of ecological opportunity may be difficult, the chronological order of events—extinction of Mesozoic marine fish fauna in the K-Pg followed by high rates of habitat transitions and the origin of singular morphologies (Figure 2.1)—suggests that post-Cretaceous niche vacancy was the main driving force behind Carangaria's evolutionary success. Comparisons between rates of lineage diversification and morphological evolution provide important insights into the mechanisms of origination and diversification of higher-level taxa (Foote, 1993). The results presented herein for Carangaria reveal variable dynamics during the clade's history, with high levels of lineage, morphological, and ecological diversity being reached within a relatively short period in the aftermath of the K-Pg mass extinction. By and large, temporal associations of the initially accelerated rates for the three metrics investigated herein fit Simpsonian predictions on adaptive radiation. They also ultimately underscore the importance of increased ecological opportunity arising in the wake of mass extinctions by providing vacant space that prompted niche divergence along the benthic-pelagic axis and the rapid evolution of major clades.

References

Adams, D., Collyer, M., and Kaliontzopoulou, A. (2019). Geomorph: Software for geometric morphometric analyses.

Adams, D. C. and Collyer, M. L. (2018). Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations. *Systematic Biology*, 67(1):14–31.

Adams, D. C. and Otárola-Castillo, E. (2013). Geomorph: An r package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4):393–399.

Alfaro, M. E., Faircloth, B. C., Harrington, R. C., Sorenson, L., Friedman, M., Thacker, C. E., Oliveros, C. H., Černý, D., and Near, T. J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary.

Arcila, D. and Tyler, J. C. (2017). Mass extinction in tetraodontiform fishes linked to the Palaeocene – Eocene thermal maximum. *Proc. R. Soc. B*, 284.

Bellwood, D. R. (1996). The eocene fishes of Monte Bolca: The earliest coral reef fish assemblage. *Coral Reefs*, 15(1):11–19.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton, J. C., Zhang, F., Buser, T., Campbell, M. a., 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., Betancur-R., R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. a., Ballesteros, J. a., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, Apr 18(APR 2013):1–45.

Betancur-R., R., Henhong, C. L. I., Unroe, T. H. A. M., Allesteros, J. E. A. B., Rtí, G. U. O., Betancur, R., Li, C., Munroe, T. a., Ballesteros, J. a., and Ortí, G. (2013). Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology*, 62(5):763–785.

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., and Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1):162.

Bookstein, F. L. (1997). *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.

Burress, E. D., Holcomb, J. M., Tan, M., and Armbruster, J. W. (2017). Ecological diversification associated with the benthic-to-pelagic transition by North American minnows. *Journal of Evolutionary Biology*, 30(3):549–560.

Cavin, L. (2002). Effects of the Cretaceous-Tertiary Boundary Event on Bony Fishes. pages 141–158. Springer, Berlin, Heidelberg.

Chakrabarty, P. (2005). Testing Conjectures about Morphological Diversity in Cichlids of Lakes Malawi and Tanganyika. *Copeia*, 2005(2):359–373.

Clavel, J., Escarguel, G., and Merceron, G. (2015). mvmorph : an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution*, 6(11):1311–1319.

Claverie, T. and Wainwright, P. C. (2014). A Morphospace for Reef Fishes: Elongation Is the Dominant Axis of Body Shape Evolution. *PLoS ONE*, 9(11):e112732.

Cooney, C. R., Bright, J. A., Capp, E. J. R., Chira, A. M., Hughes, E. C., Moody, C. J. A., Nouri, L. O., Varley, Z. K., and Thomas, G. H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, 542(7641):344–347.

Drummond, A. J., Xie, W., and Heled, J. (2012). Manual * BEAST. pages 1–18.

Erwin, D. H. (2001). Lessons from the past: Biotic recoveries from mass extinctions. *Proceedings of the National Academy of Sciences*, 98(10):5399–5403.

Feng, Y.-J. J., Blackburn, D. C., Liang, D., Hillis, D. M., Wake, D. B., Cannatella, D. C., and Zhang, P. (2017). Phylogenomics reveals rapid, simultaneous diversification of three major clades of Gondwanan frogs at the Cretaceous-Paleogene boundary. *Proceedings of the National Academy of Sciences of the United States of America*, 114(29):E5864–E5870.

Foote, M. (1993). Discordance and concordance between morphological and taxonomic diversity. *Paleobiology*, 19(02):185–204.

Fricke, R., Eschmeyer, W. N., van der Laan, R., Fricke, R., and van der Laan, R. (2016). Catalog of fishes: genera, species, references.

Friedman, M. (2009). Ecomorphological selectivity among marine teleost fishes during the end-Cretaceous extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13):5218–23.

Friedman, M. (2010). Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proceedings of the Royal Society B: Biological Sciences*, 277(1688):1675–1683.

Friedman, M. and Johnson, G. D. (2005). A New Species of Mene (Perciformes: Menidae) from the Paleocene of South America, with Notes on Paleoenvironment and a Brief Review of Menid Fishes. *Journal of Vertebrate Paleontology*, 25(4):770–783.

Froese, R. and Pauly, D. (2016). FishBase: World Wide Web electronic publication.

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., Schulte, J. A., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T., and Mooers, A. T. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution*, 64(8):2385–2396.

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., and Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics*, 24(1):129–131.

Harrington, R. C., Faircloth, B. C., Eytan, R. I., Smith, W. L., Near, T. J., Alfaro, M. E., and Friedman, M. A. (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evolutionary Biology*, pages 1–14.

Helfman, G. S., Collette, B. B., Facey, D. E., Bowen, B. W., Collette, B. B., Facey, D. E., and Bowen, B. W. (2009). *THE DIVERSITY OF FISHES: Biology, Evolution and Ecology*, volume Second.

Hohna, S. (2014). Likelihood Inference of Non-Constant Diversification Rates with Incomplete Taxon Sampling. 9(1):17–20.

Höhna, S., May, M. R., and Moore, B. R. (2015). TESS: An R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. *Bioinformatics*, 32(5):789–791.

Hollingsworth, P. R., Simons, A. M., Fordyce, J. A., Hulsey, C. D., Hollingsworth Jr, P. R., Simons, A. M., Fordyce, J. A., Hulsey, C. D., Hollingsworth, P. R., Simons, A. M., Fordyce, J. A., and Hulsey, C. D. (2013). Explosive diversification following a benthic to pelagic shift in freshwater fishes. *BMC evolutionary biology*, 13:272.

Hopkins, M. J. and Smith, A. B. (2015). Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 112(12):3758–63.

Hughes, M., Gerber, S., and Wills, M. A. (2013). Clades reach highest morphological disparity early in their evolution. *Proceedings of the National Academy of Sciences*, 2013(39):36–38.

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., and 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(9).

Near, T. J. (2013). Supporting Information ©. (1):1–20.

Near, T. J., Eytan, R. I., Dornburg, a., Kuhn, K. L., Moore, J. a., Davis, M. P., Wainwright, P. C., Friedman, M., and Smith, W. L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences*, 109(34):13698–13703.

Near, T. J. T. J., Dornburg, A., Eytan, R. I. R. I., Keck, B. P. B. P., Smith, W. L. L., Kuhn, K. L. K. L., Moore, J. A. J. a., Price, S. A. S. A., Burbrink, F. T. F. T., Friedman, M., and Wainwright, P. C. P. C. (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proceedings of the National Academy of Sciences of the United States of America*, 110(31):12738–12743.

Patterson, C. (1993). An overview of the early fossil record of acanthomorphs. *Bulletin of Marine Science*, 52(1):29–59.

Pollux, B. J. a., Meredith, R. W., Springer, M. S., and Reznick, D. N. (2014). The evolution of the placenta drives a shift in sexual selection in livebearing fish. *Nature*, 513(7517):233–236.

Price, S. A., Schmitz, L., Oufiero, C. E., Eytan, R. I., Dornburg, A., Smith, W. L., Friedman, M., Near, T. J., and Wainwright, P. C. (2014). Two waves of colonization straddling the K-Pg boundary formed the modern reef fish fauna. *Proceedings. Biological sciences*, 281(1783):20140321.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., and Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714):392–395.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223.

Röhl, U., Bralower, T., Norris, R., and Wefer, G. (2000). New chronology for the late Paleocene thermal maximum and its environmental implications. *Geology*, 28(10):927.

Rohlf, F. (2001). TPS Dig 2.0.

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Sibert, E. C. and Norris, R. D. (2015). New Age of Fishes initiated by the Cretaceous-Paleogene mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 2015(22):1–6.

Simpson, G. G. (1953). *Major features of evolution*. Columbia University Press: New York.

Slater, G. J. (2013). Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution*, 4(8):734–744.

Tavera, J., Acero P., A., and Wainwright, P. C. (2018). Multilocus phylogeny, divergence times, and a major role for the benthic-to-pelagic axis in the diversification of grunts (Haemulidae). *Molecular Phylogenetics and Evolution*, 121(December 2017):212–223.

Uyeda, J. C., Caetano, D. S., and Pennell, M. W. (2015). Comparative Analysis of Principal Components Can be Misleading. *Systematic Biology*, 64(4):677–689.

Chapter 3

Phylogenomic and comparative genomic analyses support a single evolutionary origin of flatfish asymmetry

Formatted for submission as "matters arising" manuscript in response to Lü et al. (2021) *Nature Genetics*, 53(5):742–751

Emanuell Duarte-Ribeiro, Ulises Rosas-Puchuri, Matt Friedman, Gavin C. Woodruff, Lily C. Hughes, Kent E. Carpenter, William T. White, John J. Pogonoski, Mark Westneat, Juan Martin Diaz de Astarloa, Jeffrey T. Williams, Mudjekeewis D. Santos, Omar Domínguez-Domínguez, Guillermo Ortí, Dahiana Arcila, and Ricardo Betancur-R

The eye migration that characterizes flatfish cranial development provides a unique opportunity to study the molecular mechanisms underlying this anatomical asymmetry. In a recent paper in *Nature Genetics*, (Lü et al., 2021, hereafter LEA) conducted the first comparative genomic assessment of flatfish asymmetry, and claimed that the two major lineages (Pleuronectoidei and Psettodoidei) are polyphyletic, each evolving asymmetric bodies convergently from different symmetric ancestors (flatfish polyphyly or FP; Fig. 1b). Here, we revisit this finding by analyzing three independent genome-scale datasets, including LEA's, showing that support for FP results from a failure to accommodate lineage-specific variation in base composition. We also reanalyzed LEA's genomic dataset to identify positively selected genes (PSGs) but using instead an inferred tree that groups flatfishes as monophyletic (FM), implying a single evolutionary origin of the asymmetric body plan (Fig. 1a). These results reveal a key evolutionary role of thyroid hormones (THs) and bone morphogenetic proteins (BMP), bridging evidence from the fossil record (Friedman, 2008) and single-species

molecular assays (Campinho et al., 2018).

Evolutionary non-homogeneous processes leading to base compositional non-stationarity (BCNS) are known to be major source of phylogenetic error(Betancur-R et al., 2013; Jarvis et al., 2014; Romiguier et al., 2010). Unaccounted BCNS among carangarian lineages, a major clade that comprises flatfishes and their closest symmetrical relatives, has been shown to mislead phylogenetic inference, challenging the FM hypothesis(Betancur-R et al., 2013). We assessed the effects of BCNS in flatfishes by analyzing three genome-scale datasets: the LEA dataset (18 species and 1693 exon markers), a non-coding ultraconserved elements (UCE) dataset(45 species and 596 markers; Harrington et al., 2016), and a newly generated exonic dataset (up to 389 species and 990 markers). We used two models of nucleotide substitution: a standard homogeneous model (HM) and a non-homogeneous model (NHM; Crotty et al., 2020) that permits rate variation among lineages (see SM). Contrary to LEA, our results provide overwhelming support for the single-origin FM hypothesis (Figure 3.1c). While both HM and NHM analyses using non-coding data consistently resolve the FM tree (see also Harrington et al., 2016), many HM analyses based on protein-coding sequences favor the FP topology. These results align with recent findings (Jarvis et al., 2014), suggesting that the effects of BCNS are more severe when using protein-coding markers than with introns or UCEs. Together, our analyses indicate that appropriate modeling of lineage-specific variations in protein-coding data is sufficient to favor the FM topology, reconciling the results obtained with coding and non-coding datasets.

LEA discussed additional morphological and molecular evidence to support their assertion that the flatfish cranial asymmetry has two independent origins. We argue that this is subject to interpretation and is inconsistent with other evidence from paleontology and comparative anatomy (Figure 3.1b). The earliest definitive flatfish fossils are from the early Eocene of Bolca, Italy (ca. 48.5 Ma) and include the crown pleuronectoid †Eobothus minimus plus taxa interpreted as stem pleuronectiforms: *†Amphistium paradoxum*, *†Heteronectes chaneti*, and, less certainly, *†Anorevus lorenzonii* (Bannikov and Zorzin, 2020; Friedman, 2008). In particular, *†Amphistium* and *†Heteronectes* show strong but incomplete orbital migration and are thus considered the link between flatfishes and their symmetrical relatives. Slightly older fossils (ca. 56 Ma) showing incomplete cranial asymmetry are reported in the literature but remain undescribed (Bannikov and Zorzin, 2020). Morphologically, multiple anatomical synapomorphies support the FM hypothesis (e.g., pseudomesial bar, orbital migration, recessus orbitalis, and asymmetrical pigmentation; Harrington et al., 2016), including traits that are unrelated to asymmetry (e.g., axial and fin skeletons, and otoliths; Harrington et al., 2016) and therefore are less likely to have evolved via convergent evolution (see also Chanet et al., 2020). LEA also identified chromosome rearrangements shared between two species of Pleuronectoidei that are absent in Psettodoidei. They interpreted this as additional support for the polyphyletic origin of flatfishes, but we argue that an equally likely explanation is

that pleuronectoid-specific rearrangements evolved after the split between Pleuronectoidei and Psettodoidei and, therefore, should not be used as evidence against FM. Finally, LEA reported analyses based on relative evolutionary rates and lineage-specific substitutions to support FP, but several pleuronectoid species have overlapping rates with Psettodes (see details in Appendix C Note 4).

To investigate the presence of lineage-specific adaptations under the FM topology (Figure 3.1a), we tested for positive selection among the 1693 orthologs identified by LEA using the aBSREL model of codon evolution(Smith et al., 2015). We used two different foreground schemes (see Appendix C Note 3): (1) stem flatfishes, which aimed to detect genes responsible for the initial break of symmetry in the single branch leading to all extant flatfish species; and (2) crown flatfishes, which aimed to detect genes responsible for further adaptations experienced later in the flatfish radiation. We identified 67 PSGs in the stem flatfish lineage, and 588 PSGs shared between Psettodoidei and at least one pleuronectoid lineage (Figure 3.2a). Remarkably, 15 (31%) and 162 (53%) of the lineage-specific PSGs that LEA reported using their FP topology for Pleuronectoidei and Psettodoidei, respectively, are under positive selection in the crown flatfish lineage when analyzed using the FM tree (Figure 3.2a). Although a scenario involving parallelism cannot be disregarded, such substantial overlap in PSGs provides additional corroboration for the single-origin hypothesis.

Gene Ontology (GO) characterization of PSGs revealed genes involved in the determination of left/right symmetry during embryonic development (bmp4, mkks, flr) and components of several essential signaling pathways responsible for tissue development and cell proliferation in stem flatfishes: BMP (bmp4), WNT (cxxc4, dcdc2b, carf, mkks, tinagl1), RA (asxl1), FGF (mil), NOTCH (dlk1, dlk2), and HOX (shox). Several of these are regulated by thyroid hormones (THs), a key signaling pathway that orchestrates vertebrate metamorphosis (Mourouzis et al., 2020). In flatfishes, the disruption of the thyroid axis with methimazole (MMI) inhibits asymmetric skull development by (i) reducing dermal cell proliferation in the sub-ocular region, and (ii) ablating a TH-responsive asymmetric centre, localized just ventral to the migrating eye and determined by deiodinase 2 (dio2) expression (see Figure 3.2b-f; Campinho et al., 2018). While the mechanisms by which THs regulate dio2 asymmetric expression in teleosts remain elusive, in humans it is well known that dio2 expression is stimulated by the Thyroid Transcription Factor-1 (TTF-1; Gereben et al., 2001). TTF-1 is required for the preoptic region development in zebrafish (Manoli and Driever, 2014) and is differentially expressed in the metamorphosing head of the Atlantic halibut (Alves et al., 2016). Notably, TFF-1 was also identified as a PSG in stem flatfishes when using the FM topology, a result not reported by LEA using their FP tree.

Another important developmental aspect involving the TH-responsive asymmetric centre is that its ossification correlates with the origin of the pseudomesial bar. This bone structure plays a central role in driving the complete eye migration, which is unique to crown flat-

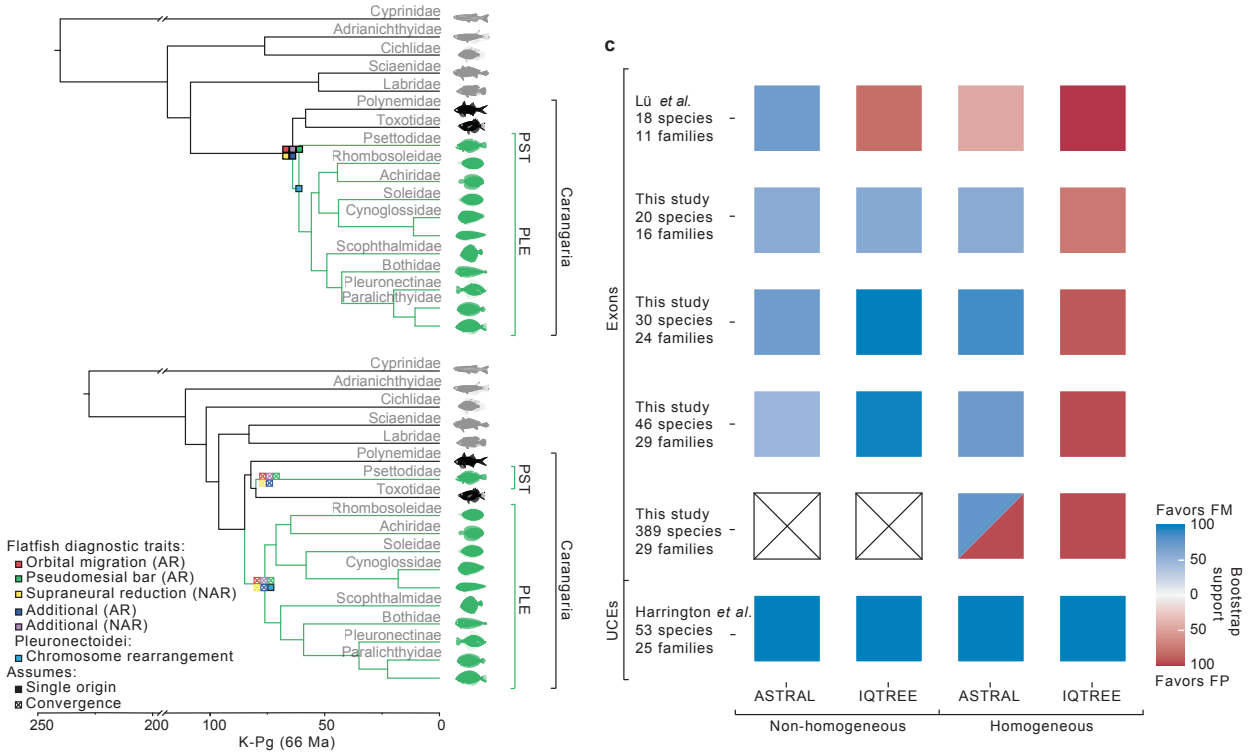

Figure 3.1: Phylogenetic relationships, divergence times and support values for the two competing hypotheses. (a) FM tree estimated using LEA's dataset with ASTRAL under a non-homogeneous model (GHOST) of nucleotide substitution (see SM Note 2 for details on time calibration). (b) FP tree illustrates the phylogenetic hypothesis and divergence times proposed by LEA. (a, b) Branch and fish silhouette colors denote major flatfish (Pleuronectoidei [PLE], Psettodoidei [PST]; green) and non-flatfish carangarian (black) clades. Color-coded squares at branches denote diagnostic characters defining flatfishes. Some diagnostic characters represent morphological features already identifiable in the earliest flatfish fossils, including asymmetry-related (AR; orbital migration) and non asymmetry-related (NAR; supraneural reduction) traits. Further diagnostic characters shared between extant Pleuronectoidei and Psettodoidei also include ARs (recessus orbitalis, pseudomesial bar, and asymmetrical pigmentation) and NARs (e.g., absence of supraneurals, dorsal-fin insertion above skull, depression on inner face of saccular otolith). Full squares represent traits evolving under the single-origin assumption, and crossed squares under the assumption of convergent evolution. (c) Meta-table shows bootstrap support for the FM and FP hypotheses using different genomic data types, datasets, phylogenetic reconstruction methods, and nucleotide substitution models (HM and NHM). Each row represents a different dataset or subset; each column represents a different phylogenetic method and substitution model. Blue squares denote support for FM, and shades show bootstrap values (0 to 100%); red squares indicate support for FP; white squares denote lacking analyses (statistically intractable; see Appendix C). Coalescent analyses were run in ASTRAL using maximum likelihood (ML) gene trees estimated with IQTREE (both HM and NHM) based on either the complete gene tree topology or a less resolved topology after collapsing clades with low bootstrap support (¡20%). Note that ASTRAL analyses based on our full exonic dataset produced contrasting results (analyses using collapsed gene trees support FM). Concatenation ML analyses were conducted in IQTREE (both HM and NHM).

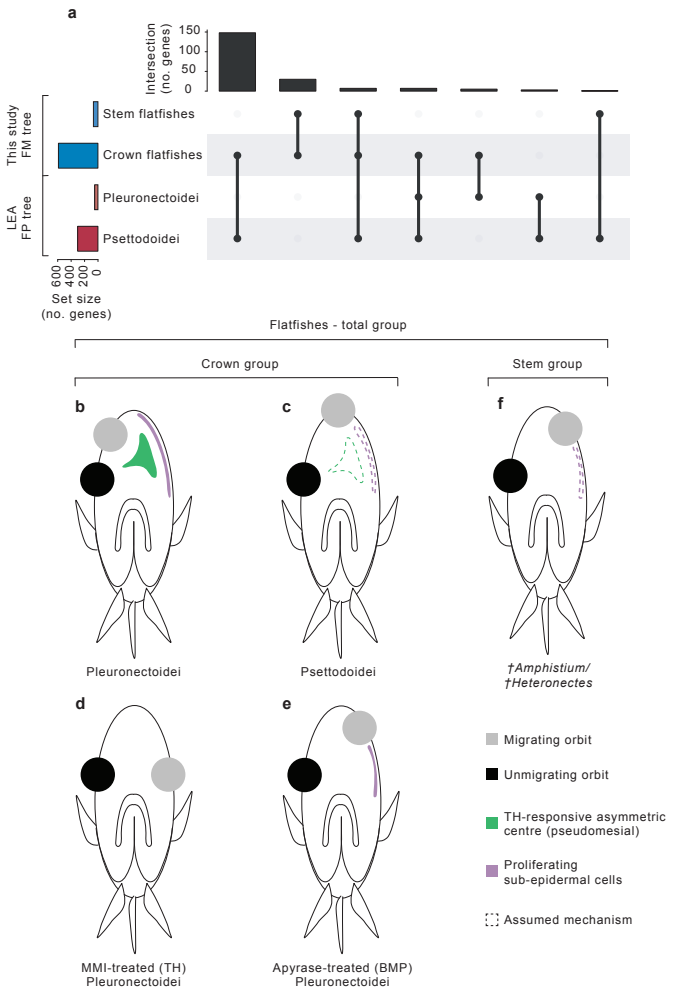

Figure 3.2: (a) Upset plot comparing positively selected genes (PSGs) identified with aBSREL for the stem and crown (foreground) flatfishes using the FM tree (horizontal blue bars), and the PSGs reported by LEA for Pleuronectoidei and Psettodoidei using the FP tree (horizontal red bars). Black dots connected by black lines represent all existent dataset intersection combinations; vertical bars indicate the number of PSGs within a particular intersection. The substantial overlap (intersection size) of PSGs identified in the two major flatfish lineages and across different analyses suggests that Pleuronectoidei and Psettodoidei share the same molecular mechanisms of asymmetric development, ultimately corroborating the single-origin hypothesis. (b-f) The developmental model summarizes the hypothesized evolutionary mechanisms underlying flatfish asymmetric development (Harrington et al., 2016; Campinho et al., 2018). (b) Simplified representation of a Pleuronectoidei post-metamorphic juvenile highlighting the two major developmental events responsible for the eye migration—sub-ocular dermal cell proliferation (purple) and the TH-responsive asymmetric centre (green) that later develops into the pseudomesial bar (flatfish neomorph that frames the migrated orbit). (c) In post-metamorphic Psettodoidei, the position of the sub-ocular dermal cell proliferation (dashed purple line) and the TH-responsive asymmetric centre (dashed green line) are inferred based on the presence of the pseudomesial bar and the complete eye migration in the adult, where migrating orbit eclipses the body mid-line. (d) Eye migration is entirely inhibited in the MMI-treated larvae due to the disruption of the thyroid axis. Apyrase-treated larvae (e) have incomplete orbit migration caused by inhibition of the TH-responsive asymmetric centre ossification3, closely resembling the primitive condition found in the earliest flatfish fossils that lack the pseudomesial bar (f)2. Note that TFF-1 (a transcription factor that regulates the TH signaling pathway), and bmp4 (involved in the heterotypic ossification of the TH-responsive asymmetric centre and the origin of the pseudomesial bar), were identified as PSGs in our analyses using the FM topology, but not in LEA's study using the FP topology. Fish illustrations are adapted from (Harrington et al., 2016; Campinho et al., 2018)

fishes (Friedman, 2008). Apyrase treatment hinders eye migration in the metamorphosing Senegalese sole by inhibiting the heterotypic dermal ossification regulated by BMP (Campinho et al., 2018). Remarkably, apyrase-treated larvae develop other indexes of metamorphic progression, such as sub-ocular dermal cell proliferation, closely resembling the primitive condition found in the earliest flatfish fossils that lack the pseudomesial bar (see Figure 3.2d; Campinho et al., 2018; Friedman, 2008). ¬¬The accumulation of amino acid substitutions at the bmp4 prodomain affects its regulation and has been identified as the major driver of craniofacial adaptations in African cichlids and Galapagos finches (Parsons and Albertson, 2009). Contrary to the LEA findings, bmp4 is positively selected in stem flatfishes, suggesting that its regulation plays a major role in the origin of the flatfish asymmetry.

Our results demonstrate that, when unproperly modeled, genome-scale datasets are prone to phylogenetic biases artificially favoring the dual-origin asymmetry hypothesis. Using the FM topology as the evolutionary framework, we find that the molecular basis of the complex phenotypic modifications in flatfishes involves an intricate interaction between THs and many important developmental pathways. These results ultimately highlight the importance of accounting for topological uncertainty and model violations in phylogenetically-informed comparative genomic analyses.

References

Alves, R. N., Gomes, A. S., Stueber, K., Tine, M., Thorne, M. A., Smáradóttir, H., Reinhard, R., Clark, M. S., Rønnestad, I., and Power, D. M. (2016). The transcriptome of metamorphosing flatfish. *BMC Genomics*, 17(1):1–24.

Bannikov, A. F. and Zorzin, R. (2020). A new genus and species of percomorph fish ("stem pleuronectiform") from the Eocene of Bolca in northern Italy. *Miscellanea Paleontologica*, 17(Studi e ricerche sui giacimenti terziari di Bolca, XX):5–14.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton, J. C., Zhang, F., Buser, T., Campbell, M. a., 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., Betancur-R., R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. a., Ballesteros, J. a., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, Apr 18(APR 2013):1–45.

Campinho, M. A., Silva, N., Martins, G. G., Anjos, L., Florindo, C., Roman-Padilla, J., Garcia-Cegarra, A., Louro, B., Manchado, M., and Power, D. M. (2018). A thyroid hormone regulated asymmetric responsive centre is correlated with eye migration during flatfish metamorphosis. *Scientific Reports*, 8(1):1–12.

Chanet, B., Mondéjar-Fernández, J., and Lecointre, G. (2020). Flatfishes interrelationships revisited based on anatomical characters. *Cybium*, 44(1):9–18.

Crotty, S. M., Minh, B. Q., Bean, N. G., Holland, B. R., Tuke, J., Jermiin, L. S., and Haeseler, A. V. (2020). GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. *Systematic Biology*, 69(2):249–264.

Friedman, M. (2008). The evolutionary origin of flatfish asymmetry. *Nature*, 454(July):209–212.

Gereben, B., Salvatore, D., Harney, J. W., Tu, H. M., and Larsen, P. R. (2001). The human, but not rat, dio2 gene is stimulated by thyroid transcription factor-1 (TTF-1). *Molecular Endocrinology*, 15(1):112–124.

Harrington, R. C., Faircloth, B. C., Eytan, R. I., Smith, W. L., Near, T. J., Alfaro, M. E., and Friedman, M. A. (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evolutionary Biology*, pages 1–14.

Jarvis, E. D., Ye, C., Liang, S., Yan, Z., Zepeda, M. L., Campos, P. F., Missael, A., Velazquez, V., Samaniego, J. A., Avila-arcos, M., Martin, M. D., Barnett, R., Ribeiro, A. M., Mello, C. V., Lovell, P. V., Almeida, D., Maldonado, E., Pereira, J., Sunagar, K., Philip, S., Dominguez-bello, M. G., Bunce, M., Lambert, D., Brumfield, R. T., Sheldon, F. H., Holmes, E. C., Gardner, P. P., Steeves, T. E., Stadler, P. F., Burge, S. W., Li, C., Ho, S. Y. W., Faircloth, B. C., and Nabholz, B. (2014). A Phylogeny of Modern Birds. *Science*, 346(6215):1126–1138.

Lü, Z., Gong, L., Ren, Y., Chen, Y., Wang, Z., Liu, L., Li, H., Chen, X., Li, Z., Luo, H., Jiang, H. H., Zeng, Y., Wang, Y., Wang, K., Zhang, C., Jiang, H. H., Wan, W., Qin, Y., Zhang, J., Zhu, L., Shi, W., He, S., Mao, B., Wang, W., Kong, X., and Li, Y. (2021). Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan. *Nature Genetics*, 53(5):742–751.

Manoli, M. and Driever, W. (2014). Nkx2.1 and Nkx2.4 Genes Function Partially Redundant During Development of the Zebrafish Hypothalamus, Preoptic Region, and Pallidum. *Frontiers in Neuroanatomy*, 8(DEC):1–16.

Mourouzis, I., Lavecchia, A. M., and Xinaris, C. (2020). Thyroid Hormone Signalling: From the Dawn of Life to the Bedside. *Journal of Molecular Evolution*, 88(1):88–103.

Parsons, K. J. and Albertson, R. C. (2009). Roles for Bmp4 and CaM1 in shaping the jaw: Evo-devo and beyond. *Annual Review of Genetics*, 43:369–388.

Romiguier, J., Ranwez, V., Douzery, E. J., and Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Research*, 20(8):1001–1009.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5):1342–1353.

Appendix A

Appendix: Evolutionary determinism and convergence associated with water-column transitions in marine fishes

Trees, tables, datasets and scripts used for comparative analyses are available from the Figshare digital repository (https://doi.org/10.6084/m9.figshare.20102951.v1)

Material and Methods

DNA extractions, exon capture and sequencing

DNA was extracted in a 96-well plate format on a GenePrep and following manufacturer's instructions at the Laboratory of Analytical Biology at the Smithsonian Institution National Museum of Natural History in Washington, DC. The quality of DNA extractions was checked by visually inspecting whether high molecular weight DNA stained with GelRed (Biotium) was visible on a 1% agarose gel. Arbor Biosciences performed library preparation using the dual round ('touchdown') capture protocol of Li et al. (2013), using eight samples multiplexed per capture. Target capture probes were designed based on alignments of 1,105 single-copy exons for all ray-finned fishes (Hughes et al., 2018, 2020), though one marker was excluded due to alignment complexity arising from high levels of sequence divergence. One sequence for each of four lineages that span the diversity of eupercarian fishes was used for probe design. These lineages included Perciformes, Gerreiformes, Tetraodontiformes, and Lutjaniformes (taxonomy following 2). Probes of 120 bp were designed to be staggered across the reference sequences every 20 bp and were filtered for potential self-hybridization and repeats using the

RepeatMasker.org database, with probes having more than 25% repeats eliminated. Several exons that were not included in Hughes et al. (2018), but that have been in wide use in fish phylogenetics were also added to the probe set: TBR1, MYH6, KIAA1239, PLAGL2, PTCHD1, RIPK4, SH3PX3, SIDKEY, SREB2, ZIC1, SVEP1, GPR61, SLC10A3, UBE3A, and UBE3A-like. Probes were synthesized with a MYBaits1 custom probe kit at Arbor Biosciences (Ann Arbor, Michigan), which is available upon request. Four mitochondrial (mtDNA) markers (COI, CYTB, 12S and 16S) were also captured with probes, but were highly diluted compared to the nuclear probes in order to improve library normalization of mtDNA and nuclear sequences (Hughes et al., 2020). Samples were sequenced at the University of Chicago Genomics facility on one lane of a HiSeq 4000 with paired-end 100 bp reads.

Data assembly and alignment

Fastq files were trimmed for adapter contamination and low-quality base calls with Trimmomatic v0.36 (Bolger et al., 2014). Reads were mapped against reference sequences used in probe design with BWA-MEM (Li et al., 2009) and potential PCR duplicates were removed with Samtools v1.9 (Li et al., 2008). Mapped reads were extracted for each locus, and an initial contig for each exon was assembled with Velvet v1.2.10 (Zerbino and Birney, 2008). The longest contig assembled by Velvet for each locus was then used as a reference for aTRAM 2.0 (Allen et al., 2017) to obtain longer contigs. aTRAM was run for a maximum of five iterations, using Velvet as the underlying assembler. Redundant contigs were removed with CD-Hit-EST with a threshold of 99% similarity (Fu et al., 2012). The coding2genome algorithm in Exonerate (Slater and Birney, 2005) was used to find reading frames by aligning it to a percomorph reference sequence that was previously verified by visual inspection (Hughes et al., 2018) to the assembled contig. If more than one contig had a reading frame for each locus, the longest contig was retained. Exons were aligned using TranslatorX (Abascal et al., 2010), with Mafft v7.421 (Katoh and Standley, 2013) as the underlying aligner. Sequences in each alignment that had more than 0.5 average pairwise distance from all other sequences were flagged with a custom python script (AlignmentChecker.py; https://github.com/lilychughes/FishLifeExonCapture), to check for possible misaligned or outlier sequences. Flagged sequences were checked visually and edited or removed on a case-by-case basis. Sequences that spanned less than 50% of the alignment were also removed for each exon (Hughes et al., 2020).

Phylogenomic analyses of exon markers

We combined genomic data from 85 newly sequenced species with sequences for 26 additional species acquired from Genbank. Including outgroups, and before eliminating dupli-

cate species tips, we concatenated individual exon alignments into a supermatrix consisting of 1,115 genes and 132 taxa (474,132 bp). To find the set of genes and taxa with minimal proportions of missing data for phylogenomic reconstructions (reduced matrix), we applied MARE (matrix reduction) v0.1.2-rc (Meyer et al., 2011), an algorithm for reducing genome-scale datasets to a subset of taxa and genes with minimal proportions of missing data. The MARE approach resulted in the retention of 1,047 orthogroups and 103 taxa (448,410 bp). For both expanded and reduced datasets, duplicate species tips were eliminated (leaving only one terminal taxon per species), and a final expanded matrix of 1,115 genes and 111 lutjanid species was used for downstream analyses.

From the expanded matrix, 13 random subsets were assembled by dividing them into seven subsets of 89 loci and six subsets of 90 loci, all of which overlap in only four genes (ATP6, COI, CYTB and RAG1). The best-fitting partitioning scheme was determined for complete datasets and subsets using PartitionFinder2. In each case, maximum-likelihood (ML) trees were estimated in RAxML v8.2.4 (Stamatakis, 2006; Stamatakis et al., 2008) using the best-fit partition selected via the Bayesian Information Criterion (BIC) and the GTRGAMMA model. For each dataset or subset, we conducted 30 independent ML searches and assessed support using non-parametric bootstrapping. The number of bootstrap replicates was determined automatically via the autoMRE function in RAxML, with bootstrap bipartitions subsequently drawn onto the best ML tree. We also estimated individual gene trees in RAxML using by-codon partitions based on sequence alignments from all individual loci. Finally, gene trees were used as input for coalescent-based analyses in ASTRAL-II v4.7.12 (Mirarab and Warnow, 2015).

Phylogenetic dating

The complete matrices and subsets were run in MCMCTree using the approximate likelihood method under the HKY85 model (Reis and Yang, 2011). Prior parameters for the MCMCTree runs were as follow: independent rate relaxed-clock model, BDparas: 1, 1, 0.80; kappa_gamma: 6, 2; alpha_gamma: 1, 1; rgene_gamma: 2, 200, 1; sigma2_gamma: 2, 5, 1. Two independent runs of the complete matrices (1,115 genes) were run for 14 million generations; subsets were run for 4 million generations. To check for convergence, we visually examined traces and effective sampling size values (ESS >200) for each parameter, after a 10% burn-in using Tracer v1.6.

Fossil calibrations

Based on recommendations by Parham et al. (2012), we used the youngest age interpretation of the fossils. All MCMCTree calibrations used uniform distributions.

(1) Root (Lutjaniformes). MRCA: *Lutjanus lutjanus*, *Pomadasys empherus*. Hard lower

bound: *†Ottaviania mariae* (Sorbini, 1983), *†Ottaviania leptacanthus* (Agassiz, 1833), *†Veranichthys ventralis* (Agassiz, 1833), *†Goujetia crassispina* (Agassiz, 1833), *†Lessinia horrenda* (Bannikov, 2014), and *†Lessinia sp.* (Bannikov, 2006; Carnevale et al., 2014). Diagnosis and phylogenetic placement: the placement of these six fossils (total group Lutjanidae) has not yet been supported by a comparative morphological phylogenetic study, and some of these may lack synapomorphies of extant lutjanids as identified by Johnson (Johnson, 1980). Therefore, the calibration is placed as stem Lutjanidae (one node below). Stratigraphic horizon and locality: early Eocene, upper Ypresian, Monte Bolca, Italy (Bannikov, 2014). Absolute age estimate: 48.5 Ma (Friedman and Carnevale, 2018). Soft upper bound: 66 Ma (see below). Prior setting MCMCTree: B(0.485,0.66,1e-300,0.05). Comments: this calibration is a combination of a primary calibration, given by minimum age of the six fossils, and a secondary calibration, where the maximum age corresponds to previous estimates of the timing of diversification in the Fish Tree of Life using multiple fossil calibrations (e.g., 2, 19–24). While this is typically treated as a stem calibration (with MRCA *Lutjanus lutjanus*, *Pristipomoides typus*), here it is instead applied as crown calibration one node below due to limitations in the MCMCTree implementation.

(2) Crown Lutjanidae. MRCA: *Etelis oculatus*, *Lutjanus lutjanus*. Hard lower bound: *†Hypsocephalus atlanticus* (Swift and Ellwood, 1972). Diagnosis and phylogenetic placement: this fossil was first described in Hoplopagrini (e.g., along the *Hoplopagrus* stem); however, the only characters suggesting a close relationship with this fossil and the extant *Hoplopagrus* are the conical canines on dentaries and premaxillaries. These are characters related to trophic behavior, which are often subject to strong selection and convergence. Furthermore, it seems that this fossil did not have a particularly large nasal capsule, as observed in *Hoplopagrus* (Pfeiffer, 1964). The fossil description clearly matches characters, however, observed in other crown lutjanids, such as the overall morphology in ethmoid regions and the generalized snapper dentition. Also, the ethmoid region, maxillaries, and premaxillaries in *†Hypsocephalus* and other extant lutjanids show the ability to expand the oral cavity both ventrally and laterally. We therefore apply a more conservative placement for this fossil in crown Lutjanidae. We note that Frédérich and Santini (2017) used the fossil to calibrate a more nested clade within crown lutjanids (i.e., the "lutjanines" + "caesionines" clade); however, no morphological evidence was provided for this decision. Stratigraphic horizon and locality: late Eocene, Operculinoides-Asterocyclina Zone in the Crystal River formation in north Florida area (Puri and Vernon, 1959). Absolute age estimate: 33.9 Ma (Swift and Ellwood, 1972). Soft upper bound: 48.5 Ma. Prior setting MCMCTree: B(0.339,0.485,1e-300,0.05). Comment: Soft upper bounds are estimated using the hard lower bound of the root calibration.

Geologic calibrations based on trans-isthmian geminate taxa

Several geminate species pairs in Lutjanidae, including terminal clades occurring on both sides of the Isthmus of Panama (Jordan and Evermann, 1898), were used to apply geologic calibrations in our tree. The timing of the final closure of the Isthmus of Panama, which separated the Eastern Pacific 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., 20), recent studies have challenged the timing of the final closure of the Panama Isthmus (Coates and Obando, 1996). More specifically, Montes et al. (2015) proposed the Middle Miocene as the final closure of the Central American Seaway, 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 during Pleistocene. Given these ongoing controversies, we set a lower hard bound of 2.8 Ma (with density 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).

(3) Geminate *Lutjanus peru-Lutjanus campechanus*. MRCA: *Lutjanus peru*, *Lutjanus campechanus*.

(4) Geminate *Lutjanus inermis-Ocyurus chrysurus*. MRCA: *Lutjanus inermis*, *Ocyurus chrysurus*.

(5) Geminate *Lutjanus argentiventris-Lutjanus alexandrei*. MRCA: *Lutjanus argentiventris*, *Lutjanus alexandrei*.

(6) Geminate *Lutjanus synagris-Lutjanus guttatus*. MRCA: *Lutjanus synagris*, *Lutjanus guttatus*.

(7) Geminate *Lutjanus cyanopterus-Lutjanus novemfasciatus*. MRCA: *Lutjanus cyanopterus*, *Lutjanus novemfasciatus*.

Ancestral range reconstructions

We used the R package BioGeoBEARS (Matzke, 2013), which compares competing models of range evolution in a phylogenetic framework. We implemented a maximum likelihood framework to build 12 different biogeographical models, including DEC (Dispersal-extinction-cladogenesis), DIVA (dispersal-vicariance analyses), and BayArea (Bayesian Inference of Historical Biogeography for Discrete Areas), each of them combined with and without the founder-speciation event (j) and the dispersal matrix power exponential (w) parameters. The j parameter allows the founding of a new area by a daughter lineage while the splitting-sister linage stays at the ancestral area (Matzke, 2014). The w parameter is used to infer the optimal dispersal multiplier matrix, which acts as an exponent on that matrix using maximum likelihood (Dupin et al., 2017). We set this 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 (65-12 Ma, 12-2.8 Ma, and 2.8-0 Ma), to account for connectivity changes between regions over geological time. The Tethys Sea region was added to first time slice only (65-12 Ma) to reflect the existence of this ancient basin. Both the dispersal-multiplier and areas-allowed matrices account for the dynamics of biogeographical 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 separated or disconnected areas. From 65 to 12 Ma, we allowed high dispersal probability (1.0) between WIO and EA through the Tethys seaway. The final closure of the Tethys seaway occurred 12 Ma (Steininger and Rögl, 1979). Thus, from 12 Ma onwards, we only allowed low dispersal probability value (0.05) between WIO and EA to reflect this closure but also to allow dispersal through the South African coast (Rocha et al., 2005). To account for the final closure of the Panama Isthmus, 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 (0.0001) between WA and TEP. Finally, for all time-slices, we set a low dispersal probability (0.05) between CP and TEP to reflect dispersal limitations associated with the crossing of the Eastern Pacific Barrier (Bellwood and Wainwright, 2002; Lessios and Robertson, 2006). We assessed the AIC scores of the twelve different biogeographical models and the best-fitting model was selected. We also summarized the six biogeographic areas initially defined into three major ocean realms by merging EA and WA into the Atlantic, WIO, CIP, and CP into the Indo-Pacific, and leaving the TEP as originally coded (Figure 1.1). All BioGeoBEARS analyses (with three and six areas) used the 'master tree' inferred with RAxML as input.

Geometric morphometrics on body shape

We generated three alternative datasets based on digitized landmarks: (i) A full-body shape dataset that comprises a set of 18 functionally homologous landmarks (Figure A.1), as well as a set of semi-landmarks that are allowed to slide along curves that outline the dorsal, anal, and caudal fins according to a minimized bending energy algorithm; (ii) a body-only dataset which is limited to the set of 18 homologous landmarks; and (iii) a fins-only dataset that includes the set of sliding semi-landmarks designed to capture fin shape variation. To account for intraspecific variation, we analyzed a maximum of four individuals per species. After performing Procrustes superimposition for each dataset, we calculated species-average coordinates, and performed principal component analyses (PCA) using the R package geomorph (Adams and Collyer, 2019). To account for possible distortions of the PCA arising from phylogenetic non-independence, we subjected the morphological data to a phylogenetically corrected principal component analysis (pPCA) (Revell, 2009). Finally, we determined the number of meaningful PC axes using the broken-stick model (Jackson, 1993; Peres-Neto et al., 2005), which minimizes loss of signal while avoiding noise from less relevant axes.

Convergence analyses

We tested the relative fit of four alternative evolutionary models using mvMORPH, a method that compares a range of evolutionary models under maximum likelihood (Clavel et al., 2015). We then fitted four alternative models of continuous-trait evolution: (i) a single-rate Brownian Motion (BM) model, (ii) a single-regime Orstein-Uhlenbeck (OU) model, (iii) a multiple-selective-regime BM (BMM) model with distinct adaptive optima for specific modes of habitat occupation (as determined based on the ancestral habitat reconstructions), and (iv) a multiple-selective-regime OU (OUM) model. Although mvMORPH is not strictly designed to test for convergent evolution, we expect to find support to the OUM model for midwater dweller lineages evolving towards the same adaptive peak (Davis and Betancur-R., 2017).

We also tested for an association between habitat occupancy and the four most relevant PC axes using the threshold model, which assesses the correlation between a discrete trait and a continuous character that co-vary according to an underlying (unobserved) trait called liability (Felsenstein et al., 2012). We used a Bayesian MCMC function (threshBayes) as implemented in the R package phytools (Revell, 2012). We ran analyses for 100 million generations, discarding the first 25% as burn-in. We then used the posterior distribution to determine whether correlation coefficients differed significantly from zero.

We explicitly tested for convergent evolution using convevol, an approach that uses distance-based metrics (C1-C4) to quantify the amount of phenotypic distance between two lineages that becomes reduced by subsequent evolution (Stayton, 2005). While C1 measures the magnitude of phenotypic distance in multidimensional space closed by evolution (ranging from 0 to 1; where 1 indicates complete convergence), it can be scaled to permit comparisons within and between different taxa and datasets (C2-C4). To test the significance of our measures of C1–C4, we compared the observed measures against null expectations generated by 1000 BM simulations. Due to computational limitations we limited the convevol analyses to the 'master tree.'

To further evaluate the strength of morphological convergence for taxa assigned to the same habitat category, we also used the Wheatsheaf index as implemented in the R package Windex (Arbuckle et al., 2014). This index, before investigating similarity, generates phenotypic distances from any number of traits across species, penalizing by phylogenetic distance. Finally, we used the multivariate data-driven approach implemented in the R package ℓ1ou (Khabbazian et al., 2016) to estimate the optimal number of selective regimes under an Ornstein-Uhlenbeck process applied to the least absolute shrinkage and selection operator (LASSO). We applied two methods to select the number of model shifts in ℓ1ou: the widely used Akaike information criterion (AICc), and the more conservative Bayesian information criterion (pBIC) (Davis and Betancur-R., 2017). To complement the ℓ1ou analyses, we also used the SURFACE method (Ingram and Mahler, 2013)for data-driven identification of

clades featuring convergent evolution.

State-dependent diversification

We first used BiSSE, as implemented in the R package diversitree (Fitzjohn, 2012), to estimate habitat-dependent rates of diversification in a Bayesian framework. For the HiSSE (Hidden State Speciation and Extinction) analyses, we tested the relative fit of a set of alternative branching models to our comparative dataset that includes null models (i.e., no state dependence), and a combination of state-dependent diversification models that incorporate unobserved hidden state within the focal habitat states. Because model-based tests of SSE methods are sensitive to model inadequacy (e.g., when the set of tested models depart substantially from the true evolutionary history of the group), we also applied the nonparametric FiSSE approach, which has shown to be robust to phylogenetic pseudoreplication and model misspecification (Rabosky and Goldberg, 2017). FiSSE compares the distributions of branch lengths for lineages with and without the focal habitat state and has been proposed as a complement to model-based SSE methods.

Results

Phylogenomic reconstruction and divergence times

The reduced matrix assembled using the MARE approach (Meyer et al., 2011) comprises 1,047 exons and a total 448,410 DNA sites for 84 species (16% of data missingness). The complete concatenated dataset contains 1,115 exons with an expanded data matrix consisting of 474,132 DNA sites for 111 species (37% of data missingness). In agreement with results from previous studies (Alfaro et al., 2018; Betancur-R et al., 2013, 2017; Frédérich and Santini, 2017; Hughes et al., 2018; Near et al., 2011; Rabosky et al., 2018), the family Lutjanidae (to the exclusion of Caesionidae) was deemed non-monophyletic based on both concatenated and coalescent-based analyses (Figure A.2-A.5). Relationships among major clades of snappers and fusiliers were resolved with strong support on the basis of analyses conducted using the reduced and the expanded matrices, largely revealing strong concordance to previous studies (Frédérich and Santini, 2017; Gold et al., 2011), with some notable exceptions explained below. All analyses invariably resolved seven major lutjanid clades (Figure A.2-A.5): the first-branching clade is composed of two reciprocally monophyletic subfamilies: (i) Apsilinae (*Apsilus*, *Lipocheilus*, and *Paracaesio*) and (ii) Etelinae (*Aprion*, *Aphareus*, *Etelis*, *Pristipomoides*, and *Randallichthys*). (iii) The next clade includes a monophyletic subfamily Paradicichthyinae with two monotypic genera, *Symphorus* and *Symphorichthys*. Next, Clade A (iv) and Clade B (v), as defined by Frédérich and Santini (2017), are sister groups, differing from the placement in their study where Clade B is clustered within Clade C. Clade

A includes *Lutjanus adetii* and *L. sebae* sister to *Pinjalo lewisi*, *P. pinjalo*, and several additional species of *Lutjanus* (*L. sanguineus*, *L. malabaricus*, *L. dodecacanthoides*, and *L. timoriensis*); Clade B is composed of *Lutjanus bohar*, *Lutjanus gibbus*, *Macolor macularis* and *M. niger*, and the fusiliers (formerly Caesionidae: *Pterocaesio*, *Caesio*, *Gymnocaesio*, and *Dipterygonotus*). We identified a substantially different placement for *Lutjanus bohar* (Clade C), estimated with GenBank sequences from two different specimens from Australia and Asia, and our trees (Clade B), based on a single specimen from Australia that was target-captured for the complete gene set. Analyses of individual gene trees suggest that Frédérich and Santini's phylogenetic placement for *Lutjanus bohar* was compromised due to miss-identification of the Asian specimen. (vi) The next lineage includes *Hoplopagrus guentherii* (sometimes placed in a separate subfamily, Hoplopagrinae), which constitutes the sister species of Clade C (vii), a large subclade that includes several lineages that span most of the diversity of *Lutjanus* as well as two monotypic genera, *Ocyurus* and *Rhomboplites*, which are nested within Lutjanus. These three genera together with *Hoplopagrus*, *Macolor*, and *Pinjalo* form the subfamily Lutjaninae (Nelson et al., 2016). It should be noted that both Lutjaninae and *Lutjanus* are taxonomic waste baskets that are grossly polyphyletic in all trees, including species in 4 of the 7 delineated clades. Many of the *Lutjanus* subclades resolved, however, tend to be clustered within major biogeographic basins (see below). Other genera that were not resolved as monophyletic include *Paracaesio*, *Pristipomoides*, and *Pterocaesio*. Taken together, these and other previous results (Frédérich and Santini, 2017; Gold et al., 2011) call for a revised taxonomy of genera and subfamilies in Lutjanidae.

The relationships estimated with the expanded matrix, in which 118 species are placed on the basis of just 1115 genes, were highly consistent with those in the reduced matrix, which features minimal proportions of missing cells (16%) providing a robust phylogenomic framework. Additionally, the placement of the GenBank species for which we lacked genomic data, where included, were resolved in the expected placement according to previous studies (Frédérich and Santini, 2017; Gold et al., 2011).

In addition to the major expanded and reduced datasets, we analyzed independent subsets derived from the expanded matrix to incorporate uncertainty in divergence times and relationships for downstream comparative analyses. Preliminary tests including a higher number of subsets, each with fewer genes (25 subsets), resulted in high levels of topological discrepancy, in particular for trees estimated with ASTRAL-II. Subsequently, we reduced the number of subsets to 13 (seven with 89 genes, and six with 90 genes), all of which produced trees with lower levels of topological discordance compared to those obtained using fewer genes. Some relationships among major lutjanid clades were not obtained in a large proportion of subset trees, despite being resolved in trees estimated with full gene sets (expanded and reduced matrices). For example, the monophyly of Clade A + Clade B, which was resolved in all analyses based on expanded and reduced matrices, was only obtained in 12 of

the 26 subset trees. To further assess topological disparity, we estimated tree space plots for the 28 trees using a multidimensional scaling (MDS) visualization implemented in phytools. The MDS plots place the RAxML and ASTRAL-II trees in opposite areas of the tree space. The ASTRAL-II trees also show greater topological disparity compared to the RAxML trees (including the 'master tree' reference; Figure A.6). We hypothesize that non-overlapping tree spaces for RAxML and ASTRAL-II trees is the result of gene tree error affecting species tree inferences—a possibility that remains to be tested using simulations. Regardless of the of the source of incongruence between RAxML and ASTRAL-II trees, however, we emphasize that most comparative methods performed here account for topological uncertainty.

Dates inferred from the 13 subsets with age estimates for MCMCTree analyses are provided in Figure A.7-A.8. Divergence-time estimates are reasonably in good agreement compared to the age of the lutjanid stem, as estimated by multi-locus analyses. Studies that did not include internal calibrations for lutjanids placed the origin of the crown group in the early Eocene (Betancur-R et al., 2017; Frédérich and Santini, 2017; Rabosky et al., 2018). In contrast, we date the age of crown lutjanids to the middle Eocene (46 Ma, 95% HPD: 40-49 Ma). The stem age of the lutjanids dated close to the Cretaceous–Paleogene (K-Pg) boundary, around 64 Ma. The Apsilinae + Etelinae clade dates from the Middle Eocene (40 Ma, 95% HPD 34-44 Ma). Estimates of subfamily-level clade ages were as follows: the subfamilies Apsilinae, Etelinae, and Paradicichthyinae, are Miocene in age, 21 Ma (95% HPD 15.34-27.25 Ma), 23.92 Ma (95% HPD 19.2-28.83 Ma), and 11 Ma (95% HPD 7.6-15 Ma), respectively. Clade A and Clade B divergences took place in the Oligocene with a clade age of 28 Ma (95% HPD 24-32 Ma). Caesionines split from other members of the Clade B around 20 Ma (95% HPD 16.83-23.86 Ma). The species-rich Clade C diverged from *Hoplopagrus guentherii* around 27Ma (95% HPD 26-31 Ma).

Biogeographic analyses

The best-supported biogeographic model for lutjanids based on six areas was the BayAREA + j + w (AICw= 0.61; Figure A.9). For three areas the DEC + j + w model had a better fit (AICw= 0.43; Figure A.10), although support for BayAREA + j + w was also substantial (AICw= 0.2). For simplicity, we thus report all results (for six and three areas) based on the BayAREA + j + w model (Figure A.11). Our ancestral area reconstruction analyses suggest that the family Lutjanidae originated in the Indo-Pacific Ocean (WIO + IO + CP) with subsequent independent colonization events of the New World (WA and TEP) via multiple routes. The main diversification of lujtanid lineages occurred within the Indo-Pacific. Lujtanines, apsilines, caesionines, and hoplopagrines originated from a widespread ancestor (WIO + CIP + CP) at 29.3 Ma (95% HPD 25.3-33.4), 21.3 Ma (95% HPD 15.3-27.2), 15.9 Ma (95% HPD 13.2-19.3 HPD), and 3.5 Ma (95% HPD 2.15-4.9),

respectively. In contrast, the subfamilies Etelinae and Paradicichthyinae originated from a WIO + CIP ancestor at 31.4 Ma (95% HPD 26.5-36.2) and 11.1 Ma (95% HPD 7.6-15), respectively. Most of the genera also appear to have an Indo-Pacific (WIO, CIP or CP) origin, except for *Ocyurus* and *Rhomboplites*, which likely originated in the WA. Caesionines and Paradicichthyines are the only subfamily-level clades of snappers that did not disperse outside their center of origin in the Indo-Pacific.

By and large, our biogeographic reconstructions suggest that the TEP was colonized 5-7 different times with likely dispersal paths taking place eastwards across the East Pacific Barrier. Four of these are currently present in the TEP and/or the Indo-Pacific but do not occur in the WA. Others (e.g., genera in the subfamilies Etelinae, Apsilinae and Lutjaninae) are present in the WA but not in the TEP, despite the fact that their most likely dispersal route likely involved the TEP. Lutjanines colonized the WA through the EA at least twice. The first event occurred westwards from the Indo-Pacific before the closure of the Tethys seaway (12 Ma), suggesting that the colonization event happened through tropical waters across the Tethys seaway (rather than through a subtropical path via Cape of Good Hope in South Africa). The second event took place after the closure for the Tethys seaway, likely requiring lineages to colonize the Atlantic via South Africa. Finally, five geminate species pairs in lutjanines had a WA origin with subsequent colonization of one species of each pair into the TEP before the final closure of the Isthmus of Panama (see comments under divergence-time calibrations).

Ancestral habitat reconstructions

The initial assessment of the fit of the two alternative models of discrete character evolution (ER and ARD) failed to reject a null model that assumes symmetric habitat transition (ER). However, ancestral state reconstructions using both the symmetric (ER) and asymmetric (ARD) models recovered similar results, identifying an average of 14.01 ($\pm$ 0.5) transitions between the two habitat categories. The reconstructions performed under the three alternative models converged, suggesting that midwater lutjanid lineages originated at 9.01 ($\pm$0.1) times on average throughout their evolutionary history, which is roughly twice as frequent as benthic-to-midwater transitions (mean 4.5$\pm$0.4). In line with these results, HiSSE analyses support a model that accounts for habitat dependent diversification (HiSSE benthic) and asymmetric transition rates. According to our HiSSE results, the rate of transitions from benthic to midwater habitats is approximately two times faster than the reverse (mean q 0.016 vs. q 0.008, respectively). These contradictory results obtained with fitDiscrete (i.e., model-fitting for discrete characters) and SSE methods are not surprising. A previous study demonstrated that the function fitDiscrete may lack statistical power, consistently failing to reject the null hypothesis of equal transition rates under highly asymmetrical empirical and

simulated conditions (Betancur-R et al., 2015).

By merging results of ancestral habitat (SIMMAP) and ancestral range (BioGeoBEARS), we find that the invasion of the water column took place independently at least once within each of the three major oceanic basins (Figure 1.1). While some areas feature more transitions than others (e.g., Indo-Pacific vs. Eastern Pacific; Figure 1.1), the ubiquitous nature of habitat transitions in lutjanids is a remarkable result of this study that highlights the deterministic character of these changes.

Geometric morphometric analyses

The number of meaningful PC axes varied among the three morphometric datasets. Both the full-body shape and the fins-only datasets are optimally represented by the first four PC axes (responsible for 78% and 85% of the total variance respectively); the body-only dataset was best represented by the first two PC axes, which accounted for more than 72% of the total variance. For the full-body dataset, the main trends in shape variation described by the first four PC axes are presented as morphospace scatter plots (Figure 1.2). For the full body-shape dataset, PC1 ($< 50\%$ of total variance) summarizes morphological differences in body elongation and caudal fin shape, features that have been repeatedly found to comprise two of the major components of fish evolution along the benthic-pelagic axis. Indeed, PC1 remarkably discriminates between benthic and midwater dwellers. The PC1 traitgram shows that different lutjanid midwater lineages independently evolved slender-bodies and furcate caudal fins, suggesting strong ecologically-driven evolutionary convergence in this clade. This pattern is further confirmed by the threshold model (Felsenstein, 2005), where for the full-body shape dataset we found a moderate correlation between PC1 and the two different habitat states (r2=0.70). The remaining three PC axes (PC2-4) summarize further relevant aspects in fin's shape variation and ornamentation. The same pattern was detected for the body-only (r2=0.66) and fins-only (r2=0.68) datasets, where only PC1 exhibits a moderate correlation. We found an extensive overlap between benthic and midwater species at lower PC axes. Together with the threshold model results—low correlation between the PC2-4 and habitat occupancy data (r2=0.04 -0.20)—our results suggests that ecomorphological convergence is less clearly associated with these axes than it is to the main PC1 axis (SI Appendix, Figure A.12).

Convergence analyses

Snappers and fusiliers display considerable morphological diversity in body shape concerning body depth and fin shape (Figure 1.2). We conducted several proposed methods to assess the scale and nature of convergence for each of the three separate datasets (full-body, using

PC1-PC4 axes, body only, using PC1-PC2 axes, and fins only, using PC1-PC4 axes).

Adaptive peak-based assessments of convergence

Results for the multivariate model fitting using the full-body shape dataset show split support for the two multi-selective-regime models (OUM and BMM) model. In both models, the distinct selective regimes correspond to different habitat categories in Lutjanidae. The remaining two alternative morphometric datasets (body-only, fins-only) show decisive support for each of the two multiple selective regime models. The body-only dataset unequivocally supports the OUM model, supporting the idea that independent lineages with similar habitat occupancy along the benthic-pelagic axis are strongly constrained towards the same adaptive landscape optimum. In contrast, the fins-only dataset supports the BMM model, a random walk multi-selective-regime that does not incorporate optimum trait values (θ).

Strength of convergence

We used the convevol distance-based measures (C1-C4) to assess the strength of convergence associated with incursions into the water column. The C1-C4 metrics are all statistically significant for the three alternative morphometric datasets. The C1 index measures how similar lineages have evolved to be more similar to one another when compared to their respective ancestors. Our results indicate that midwater lineages have, on average, closed slightly less than half of their phenotypic distance by subsequent convergent evolution (C1 = 42-48%). We also used C5, a frequency-based index that measures the number of lineages evolving into the focal region in the morphospace. Our results show that 3-5 lineages independently evolved into the area of morphospace delimited by midwater species; however, all C5 tests were non-significant ($p<0.57$).

We also used the Wheatsheaf index (w), a method that compares the degree of phenotypic similarity between the species in the a priori defined convergent clades and the disparity of these species from the non-convergent species. Wheatsheaf results (w= 1.2–1.35; Figure A.14) suggest that convergence in midwater species is significantly stronger (p 0.01; with a narrow confidence interval or CI) than would be expected from a random distribution of trait values simulated under a Brownian Motion model (BM) across the tree. All w values are similar, and CI overlaps among the three alternative morphometric datasets, which suggest that both body shape and fins morphology present a similar strength in convergent evolution. To further validate these results, we calculated w using benthic species as focal clades. In this case, w was significantly smaller than values simulated under BM in all three morphometric datasets (w= 0.82–1.03; $p > 0.99$). These results support the idea that morphological diversity is high among benthic species, and strong convergent evolution is largely restricted to midwater lutjanids.

ℓ1ou and SURFACE analyses

Finally, we assessed the extent of convergence evolution without a priori habitat designations using ℓ1ou and SURFACE. Similar to the simulations used by Khabbazian et al. (2016), we conducted two different tests using the 'master tree' and the alternative ASTRAL-II tree based on the full dataset. For ℓ1ou, we first assessed shifts on the first PC and pPC axis. We then explored the performance of ℓ1ou and SURFACE when applied to multiple PC axes (first four axes for full body shape and fins-only datasets and first two axes for body-only dataset; Figure A.12). For our first test, ℓ1ou (based on both AICc and pBIC) detected on average more shifts when we use PCA instead of pPCA (FBS: AICc 14.75–14.5, pBIC 5.2-5; BO: AICc 10.5-11.5, pBIC 2-2.5; FO: AICc 14.75-12.5, pBIC: 4.25-4.75 respectively). To properly account for phylogenetic co-variation we report all downstream analyses based on pPCA (Revell, 2009; Uyeda et al., 2015). When analyzing the first PC axis only on our three morphometric datasets (body only, fins only and body plus fins combined) using ℓ1ou + AICc, fewer shifts were detected relative to similar analyses based on multiple PC axes (see below), showing the gain in detection power when combining multiple axes in our second test. The more conservative ℓ1ou + pBIC test detected a single adaptive shift at the base of the fusilier clade, falling in line with traditional taxonomic delimitations of snappers and fusiliers as separate families (Carpenter, 2001, 1987, 1988, 1990, 1993; Fricke et al., 2016; Froese and Pauly, 2019). The single adaptive shift largely reflects elongation of fusilier's body plan, mouth reduction, and forking of caudal fin—key adaptations for life in the water column.

Results for the multivariate ℓ1ou and SURFACE analyses for the twenty-eight trees produced largely congruent results (Figure A.19); interpretations reported here are thus based on the 'master tree' (RAxML) and the alternative ASTRAL-II tree. The ℓ1ou model using AICc for shift detection on the full-body shape dataset identified, to their four pPC traits, 16 distinct adaptive shifts from mean trait values, which converged in eleven regimes, composed of four shifts to convergent peaks and five unique non-convergent peak shifts. Several species converged to some extent to benthic and midwater states (e.g., subfamilies Apsilinae, Etelinae, *C. cuning*, *O. chrysurus*, and *L. inermis*). Distantly related benthic species (e.g., *Lutjanus novemfasciatus* and *L. argentimaculatus*, *L. madras*, and *Symphorichthys spilurus*; Figure A.15) also show body-shape convergence. Considering body shape only, ℓ1ou identified 13 distinct adaptive shifts to their two pPC traits, from mean trait values, which converged into seven different regimes, collapsed down to four convergent peaks with similar morphologies (deep or slender body shapes). Subfamilies Apsilinae, Etelinae, *R. aurobens*, and *Pinjalo* are midwater dwellers that converge on slender body shape. By contrast, *L. cyanopterus*, *L. dentatus*, *L. madras*, *L. carnolabrum*, and *L. biguttatus* tend to converge on deep-bodied phenotypes (Figure A.16). Finally, for the fins-only analysis, ℓ1ou identified 13 distinct adaptive shifts from mean trait values (AICc= -2172.456), which collapsed to

ten distinct regimes that mostly correspond to subfamily-level clades, except for the first-branching clade (subfamilies Apsilinae and Etelinae) which did not reveal a shift (Figure A.17). As expected, the ℓ1ou results using pBIC for shift detection were more conservative for the three datasets on average. For instance, on the full-body dataset, nine adaptive shifts converged to eight regimes, while on the body only dataset, five distinct adaptive shifts collapsed to four different regimes, and in both cases, we only found a single convergent regime; for the fins-only dataset, the results yielded nine distinct adaptive shifts, which collapsed down to nine different regimes. Results using SURFACE were similar for the three datasets (Figure A.18). On the full-body dataset, we identified on average eighteen distinct adaptive peaks (k=18), with twelve distinct regimes (k=12), composed of four convergent events and eight unique non-convergent peak shifts, whereas on the body-only and fins-only datasets we detected sixteen different adaptive variations (k=16). On the body-only dataset, eight total adaptive peaks were identified as being reached multiple times by independent lineages (k=8). On the fins-only dataset we identified eleven (k=11) distinct regimes, composed of four convergent events and seven unique non-convergent peak shifts. Neither ℓ1ou nor SURFACE (Figure A.19) analyses show a complete convergence of phenotypic regimes (e.g., not all midwater species collapsed down to one regime).

To determine the extent to which identified convergences in the adaptive landscape could have occurred by chance under non-convergent processes, character histories were simulated using a null distribution of 99 random phenotypic datasets under simple BM and OU models of evolution (following 60). Convergence summary statistics were determined from each of the 99 simulations for each model, and the significance of the observed results were estimated as the frequency of combined simulated and observed values being greater than or equal to that of the best-supported model under our data.

Simulation comparisons for our three datasets under both a single-peak OU model and a BM model revealed that there was a strong evidence of significantly greater numbers of convergent shifts than would be expected by chance (Figure A.20). For the full body shape simulations, the variables with the highest frequencies were shifts (10 OU/10 BM), number of regimes (4 OU/6 BM), and number of convergent regimes (3 OU/3 BM), in contrast to the empirical data which included 15 shifts, 6 regimes and 5 convergent regimes. For the body only simulations, the variables with the highest frequencies were shifts (11 OU/10 BM), number of regimes (5 OU/6 BM), and number of convergent regimes (3 OU/3 BM), in contrast to the empirical data which yielded 10 shifts, 5 regimes and 4 convergent regimes. For the fins only simulations the variables with the highest frequencies were shifts (11 OU/10 BM), number of regimes (5 OU/6 BM), and number of convergent regimes (3 OU/2-3 BM), in contrast to the empirical data, which included 17 shifts, 7 regimes and 6 convergent regimes. These results suggest convergence of many lineages to multiple, shared adaptive peaks in body shape ecomorphology, characterizing the trait changes in Lutjanidae.

State-dependent diversification

We assessed whether the preference for different habitat states would affect rates of lineage diversification. For 20 out of the 28 trees, model fitting comparisons supported a state-dependent diversification model that incorporates a hidden state associated with benthic lineages (HiSSE benthic; Figure A.21a). While this model was not decisively favored by our data and the support was shared with two alternative null models (Figure A.22b), under the 'HiSSE benthic' model net diversification rates (speciation minus extinction) were roughly 2x faster in benthic lineages compared to their midwater counterparts. Finally, the results obtained with HiSSE were consistent with those obtained using non-parametric FiSSE and Bayesian-based BiSSE estimations of diversification rates (Figure A.23), identifying support for habitat-dependent diversification. In agreement with our hypotheses, benthic dwellers tend to show faster rates of net diversification than midwater species, including both faster speciation and slower extinction.

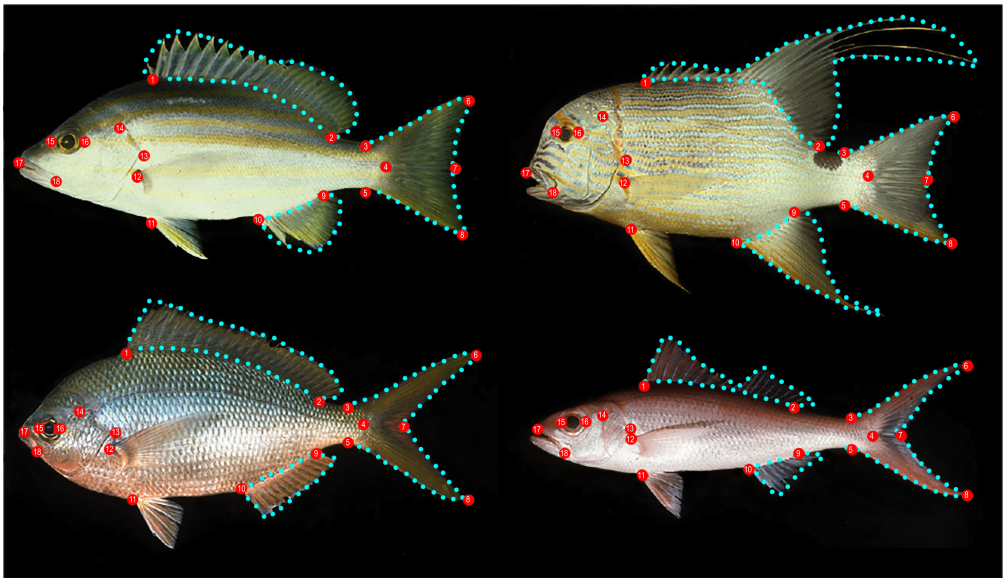

Figure A.1: Geometric morphometrics digitization scheme including 18 landmarks (red circles) selected to summarize body-shape variation in Lutjanidae: (1) anterior insertion of dorsal fin, (2) posterior insertion of dorsal fin, (3) dorsal insertion of caudal fin, (4) posterior end of lateral line, (5) ventral insertion of caudal fin, (6) end of upper lobe of caudal fin, (7) midpoint of caudal fin, (8) end of bottom lobe of caudal fin, (9) posterior insertion of anal fin, (10) anterior insertion of anal fin, (11) anterior insertion of pelvic fin, (12) upper insertion of pectoral fin, (13) caudal end of opercule, (14) dorsal end of opercule, (15) anterior margin of eye, (16) posterior margin of eye, (17) rostral tip of premaxilla, (18) caudal end of maxilla. Turquoise points outline the dorsal, anal and caudal fins indicate sliding semi-landmark curves (fins only dataset).

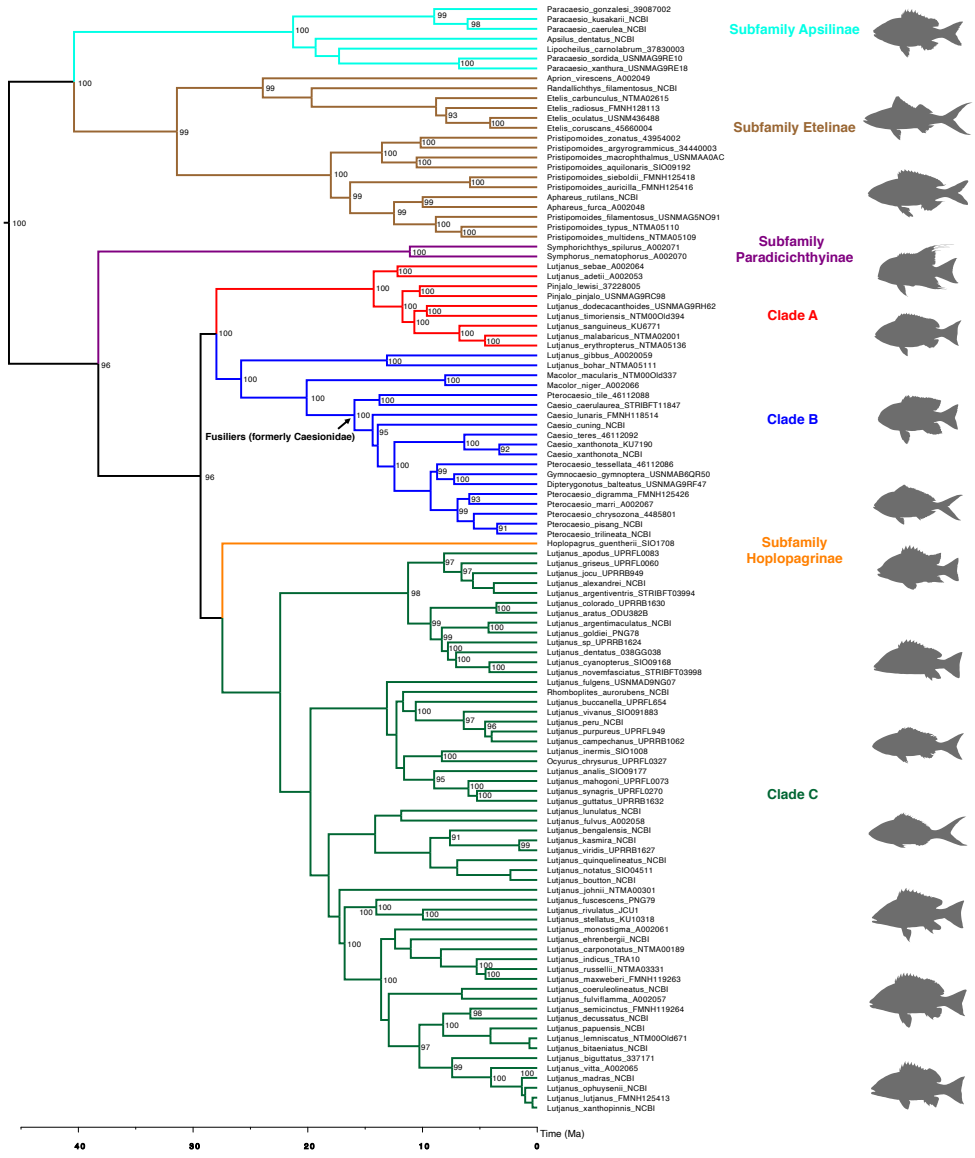

Figure A.2: Phylogenetic tree inferred with RAxML for the expanded dataset ('master tree') and time-calibrated using MCMCTree. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support.

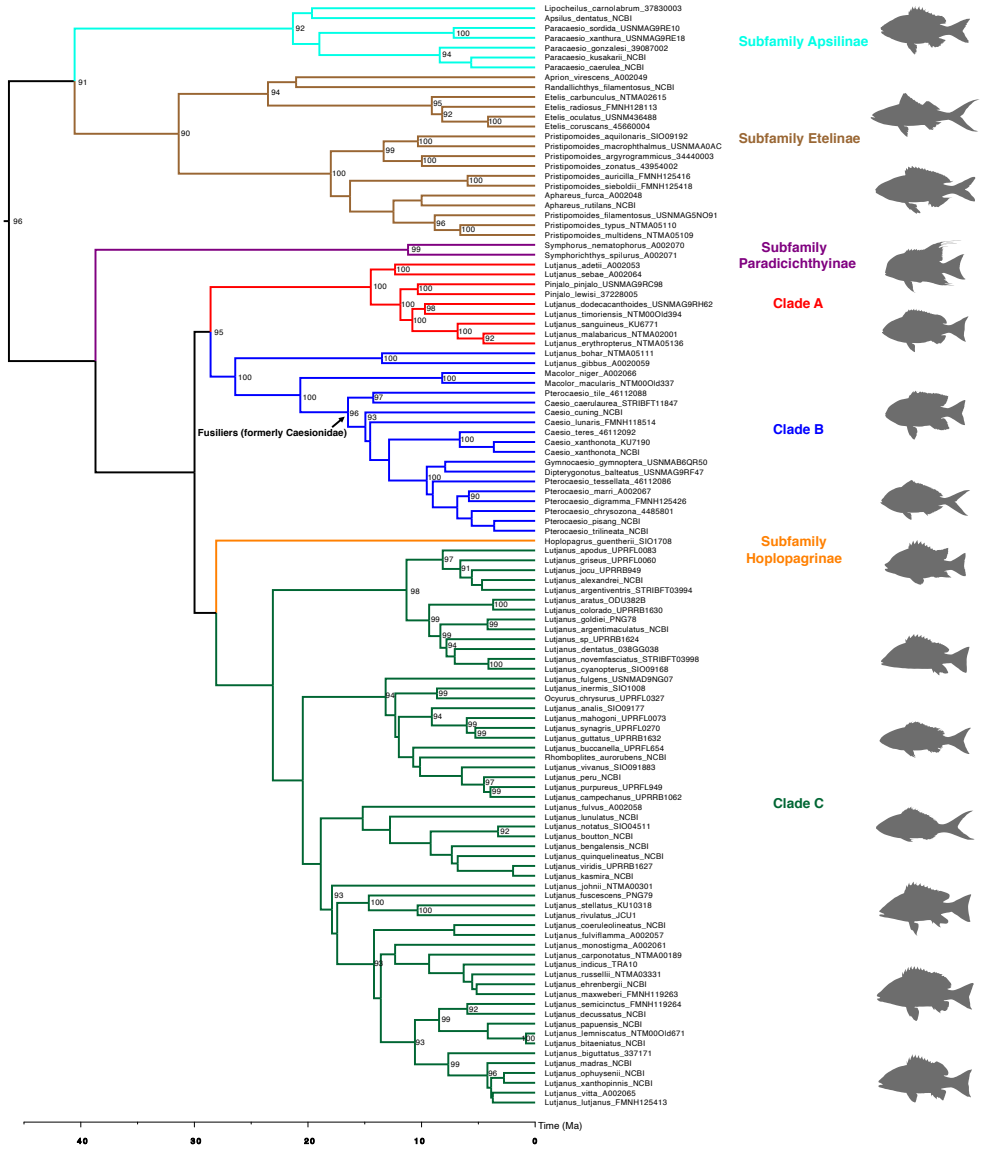

Figure A.3: Phylogenetic tree inferred with ASTRAL-II for the expanded dataset and time-calibrated using MCMCTree. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support.

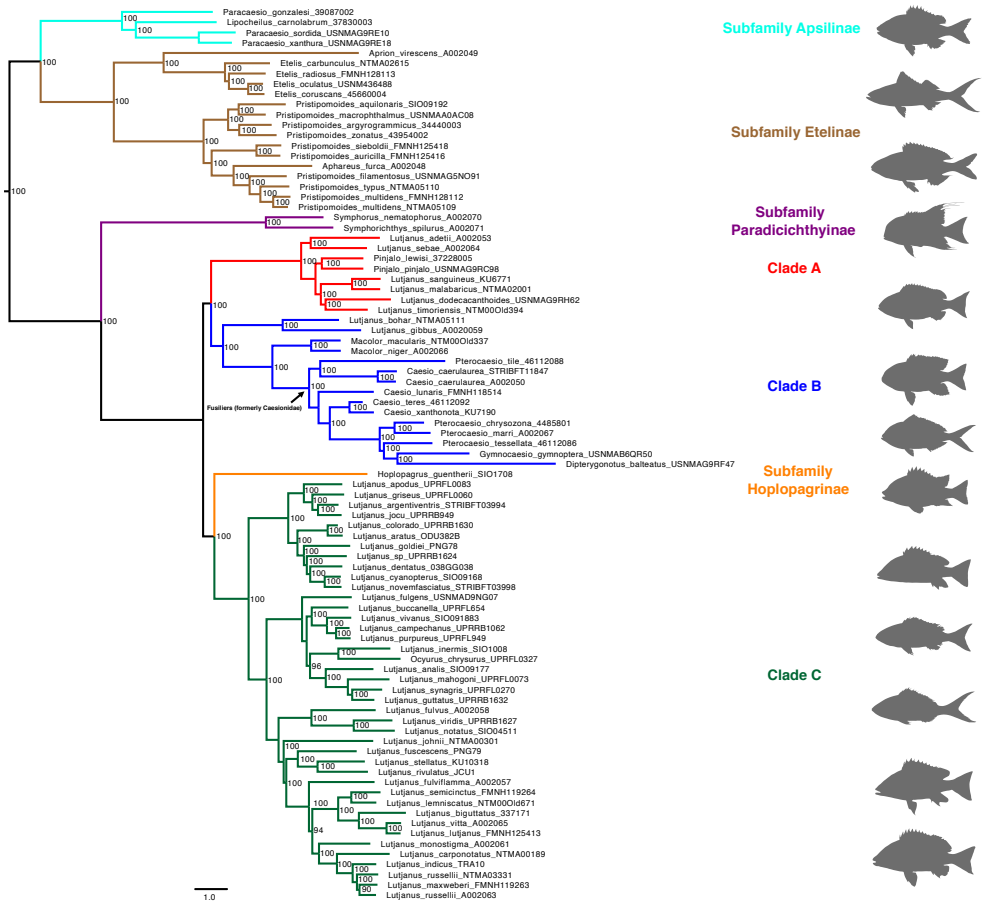

Figure A.4: Phylogenetic tree inferred with RAxML for the reduced dataset. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support. The purpose of this inference was to assess sensitivity of phylogenetic results to missing data (16%). Thus, this tree it was not time calibrated.

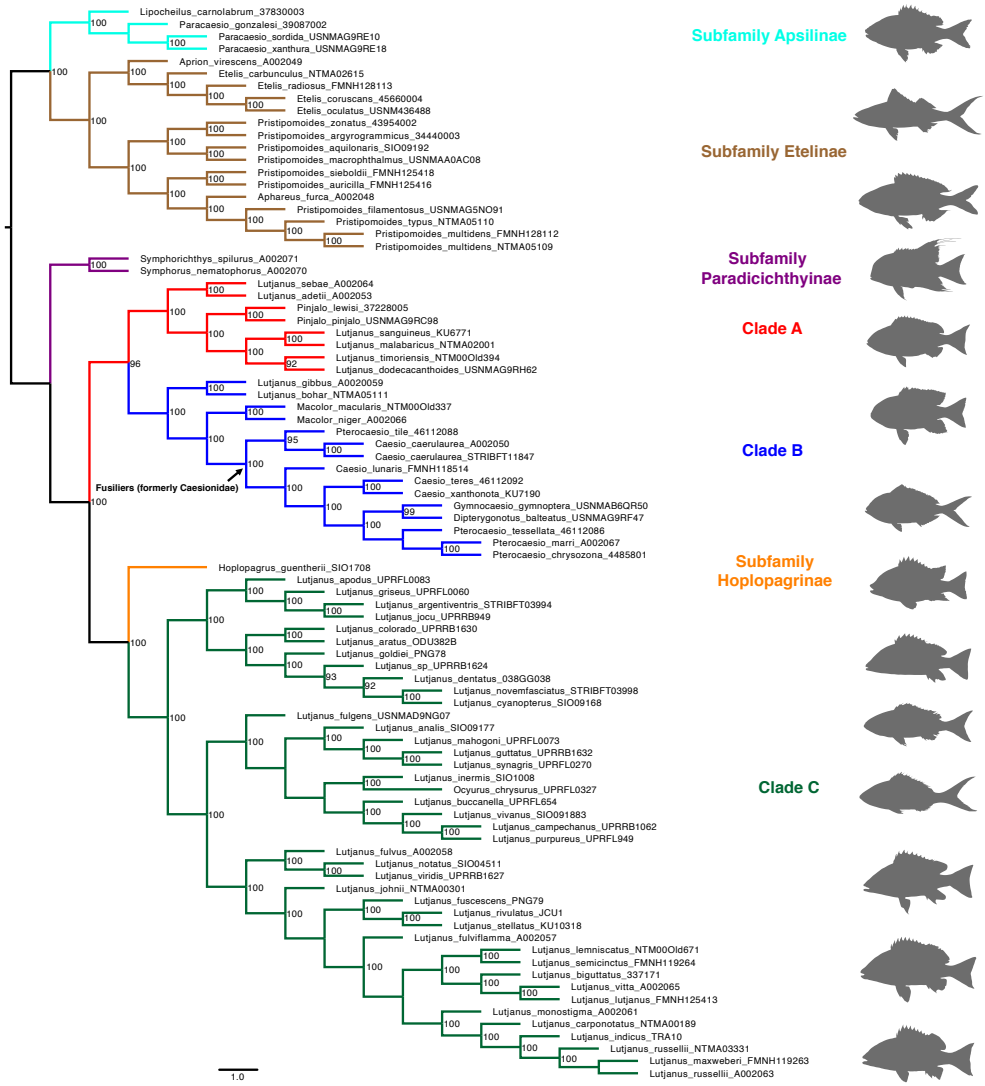

Figure A.5: Phylogenetic tree inferred with ASTRAL-II for the reduced dataset. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support. The purpose of this inference was to assess sensitivity of phylogenetic results to missing data (16%). Thus, this tree it was not time calibrated.

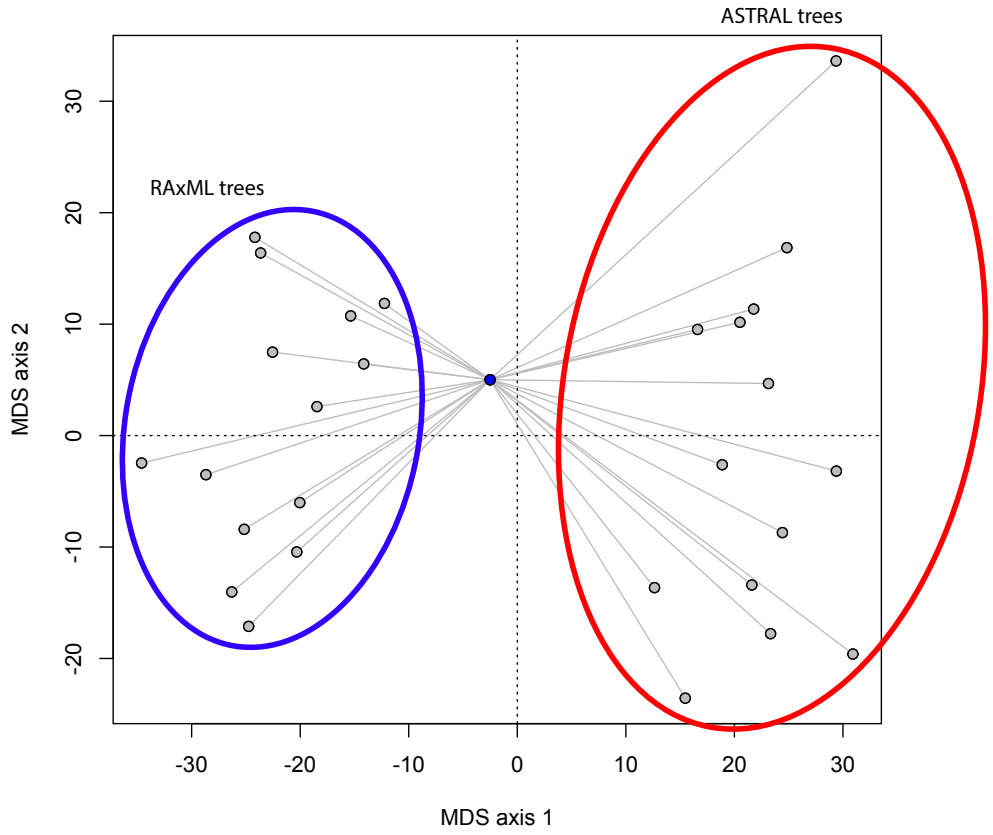

Figure A.6: Tree spaces for the twenty-eight trees estimated in this study. MT: 'master tree', AT: alternative ASTRAL-II tree based on the full dataset. The blue dot represents the average (centroid) tree in tree space.

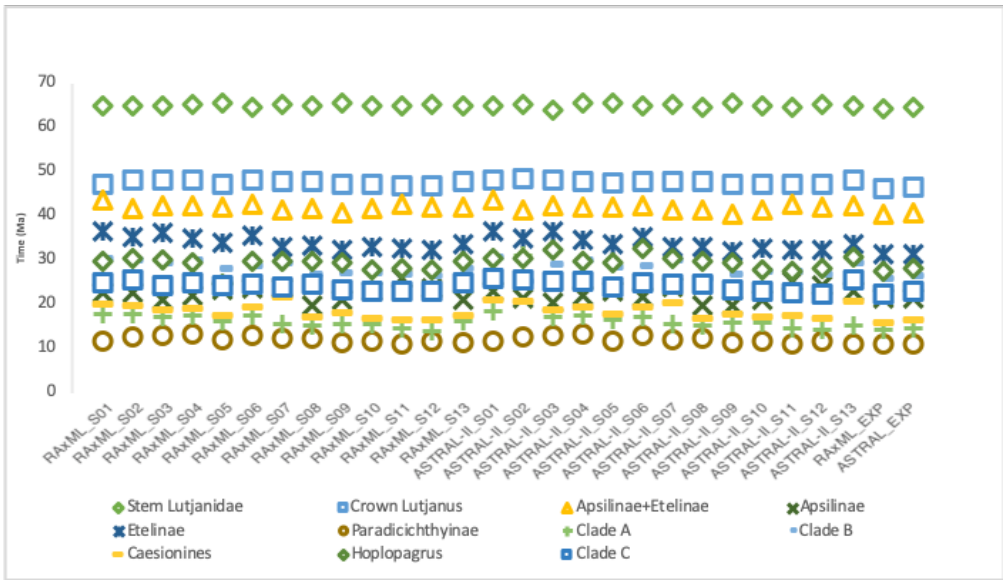

Figure A.7: Divergence date uncertainty for major lutjanid clades based on the 28 trees dated with MCMCTree.

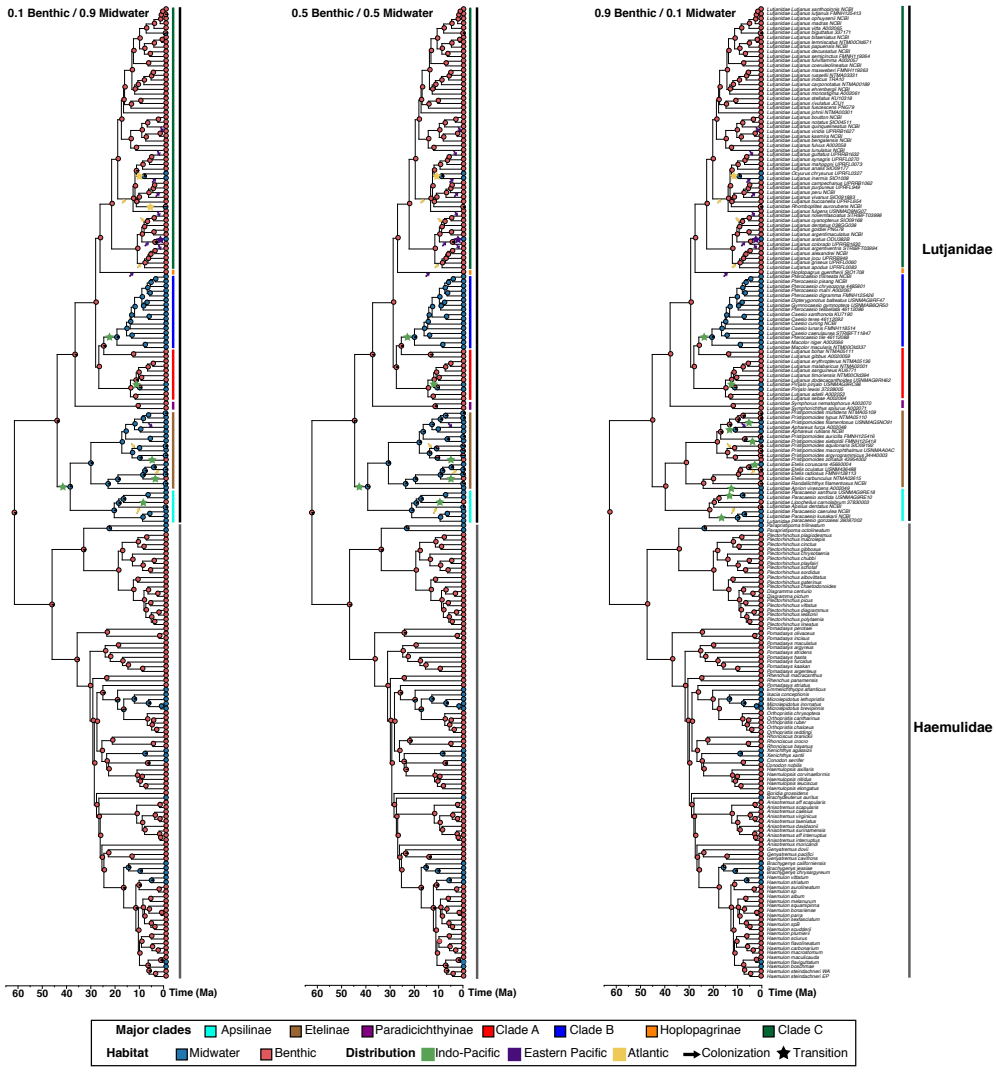

Figure A.8: SIMMAP reconstructions based on the 'master tree' (RAxML) and expanded haemulid outgroups following three alternative probability schemes for coding species with ambiguous habitat affiliations: 0.1 benthic/0.9 midwater, 0.50 benthic/0.50 midwater, 0.9 benthic/0.1midwater. This figure shows how different coding schemes affect habitat reconstructions (see also Figure A.9-A.11 for similar results based on the 28 trees). Major lutjanid clades are highlighted. Biogeographic colonization events of the Atlantic and the tropical eastern Pacific inferred with BioGeoBEARS (see Figure A.12-A.14) are indicated with arrows. Stars indicate the most likely ancestral area where SIMMAP analyses identify a habitat transition.

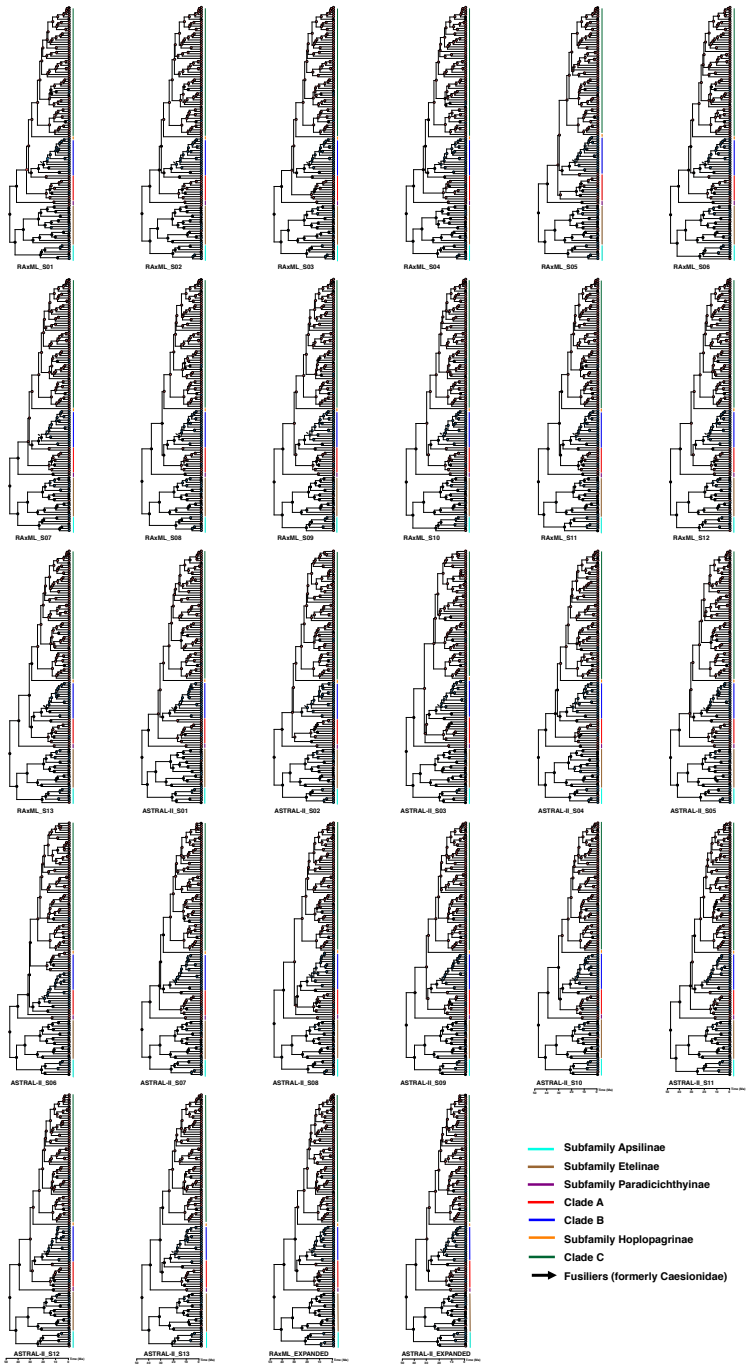

Figure A.9: SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.5 benthic/0.5 midwater probability scheme. The reconstructions included all haemulid outgroups (see Figure A.8), but these were pruned out here to facilitate visualization.

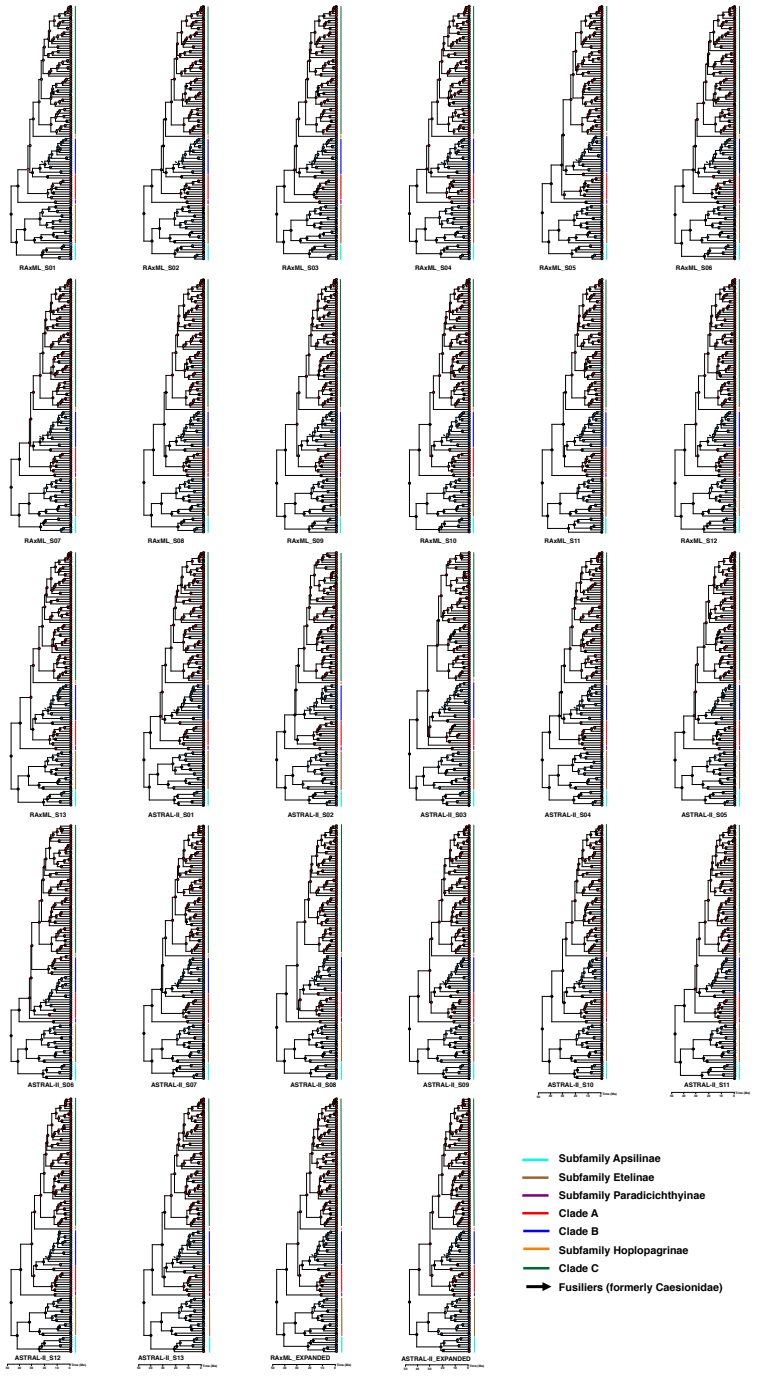

Figure A.10: SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.1 benthic/0.9 midwater probability scheme. The reconstructions included all haemulid outgroups (see Figure A.8), but these were pruned out here to facilitate visualization.

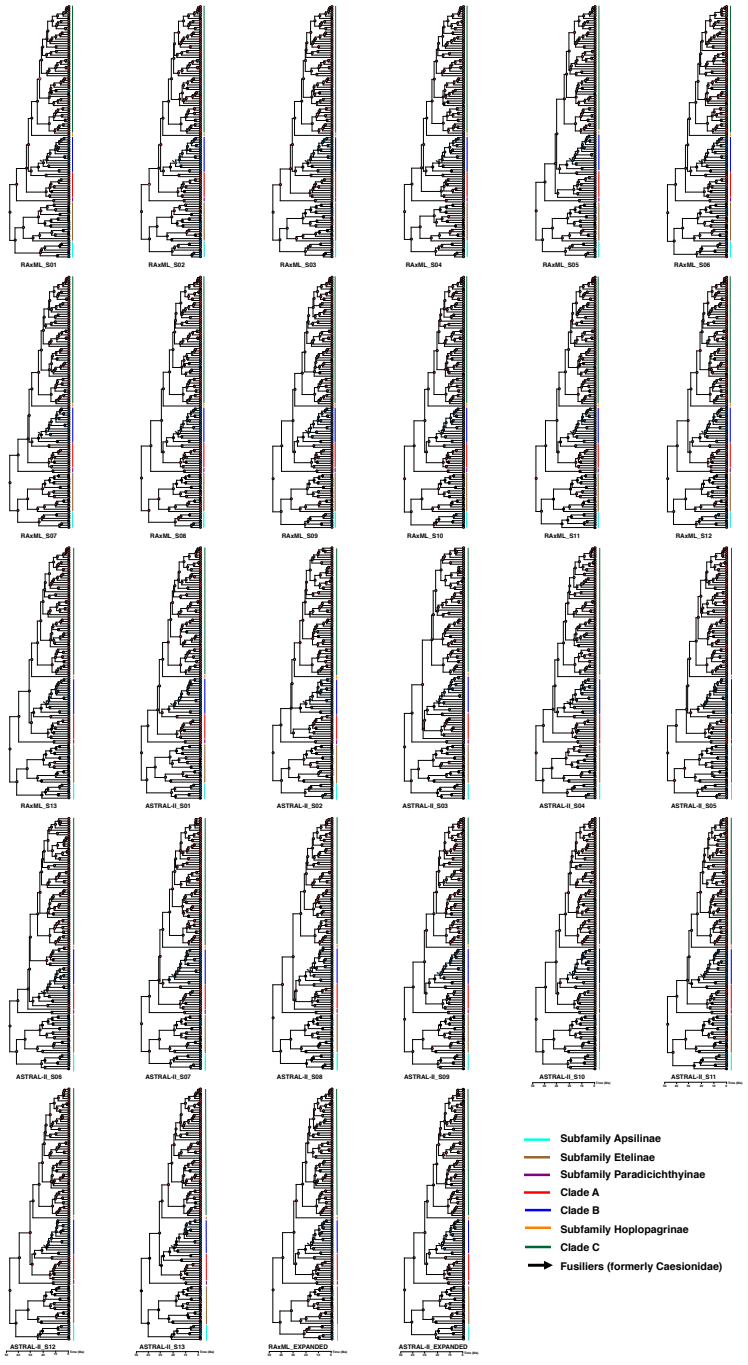

Figure A.11: SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.9 benthic/0.1 midwater probability scheme. The reconstructions included all haemulid outgroups (see Figure A.8), but these were pruned out here to facilitate visualization.

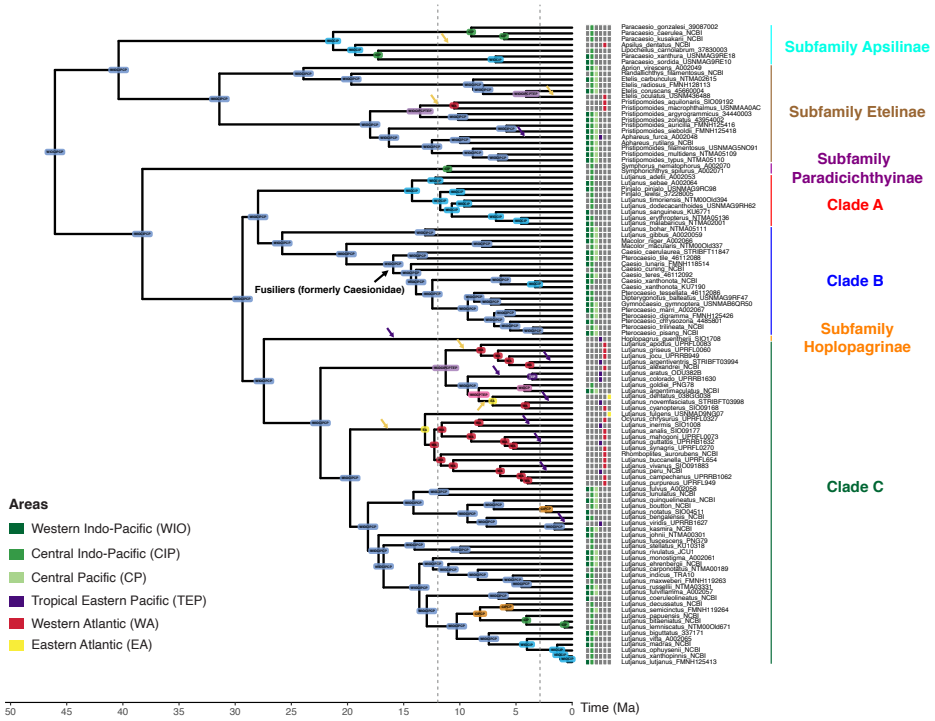

Figure A.12: Ancestral range estimations (BioGeoBEARS) for Lutjanidae using the best-supported biogeographical model for six areas (BAYAREALIKE+j+w) applied to the 'master tree.' Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond to two major biogeographic events, the Tethys Seaway closure (12 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Nine purple and six yellow arrows indicate colonization events to the TEP and the Atlantic, respectively.

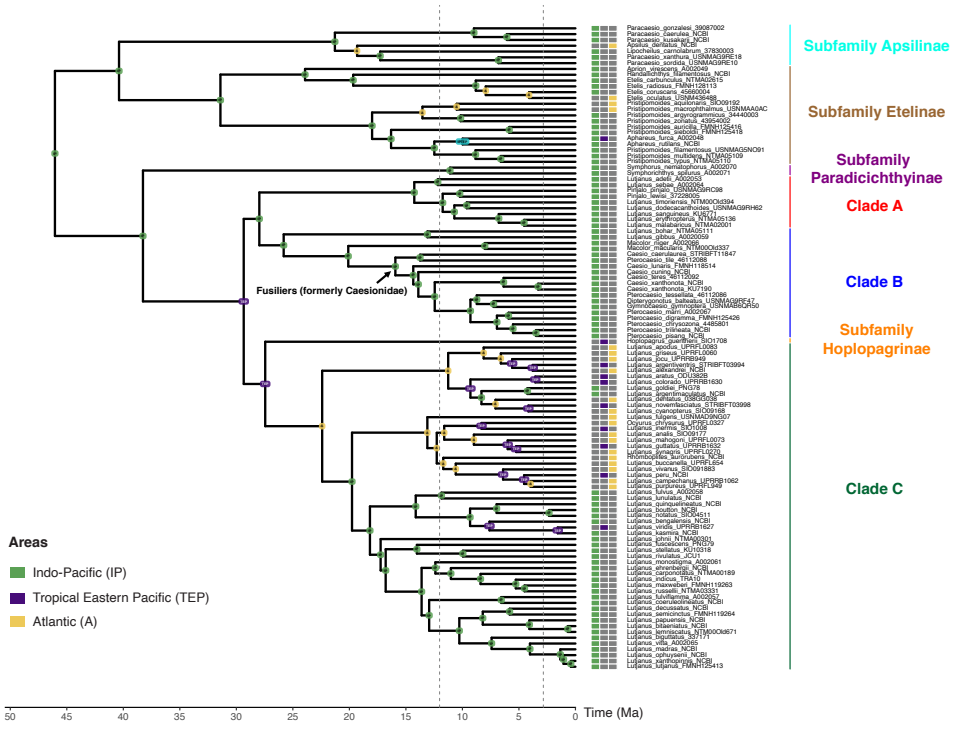

Figure A.13: Ancestral range estimations (BioGeoBEARS) for Lutjanidae using the best-supported biogeographical model for three areas (DEC +j+w) applied to the 'master tree.' Boxes represent the geographic distribution of extant species. Dotted lines represent time constraints that correspond to two major biogeographic events, the Tethys Seaway closure (12 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations).

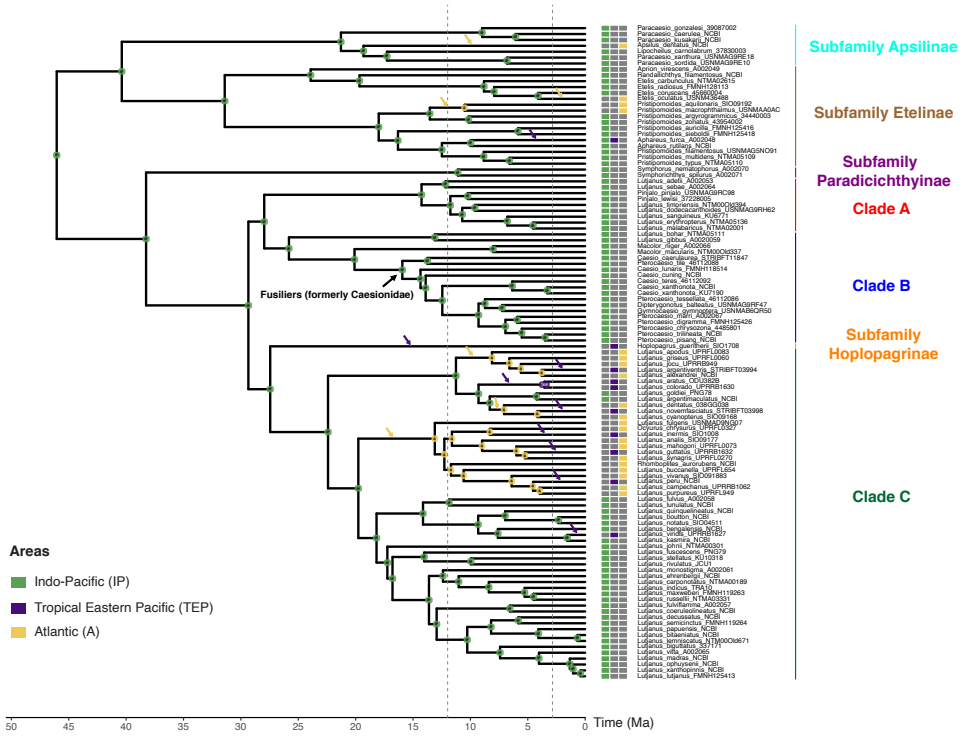

Figure A.14: Ancestral range estimations (BioGeoBEARS) for Lutjanidae based on the three-areas scheme applied to the 'master tree' but using the best-fit model for six areas (BAYAREALIKE+j+w; see Figure A.8). Boxes represent the geographic distribution of extant species. Dotted lines represent time constraints that correspond to two major biogeographic events, the Tethys Seaway closure (12 Ma) and the undisputed minimum age for the closure of the Isthmus of Panama (2.8 Ma; see comments under divergence-time calibrations). Nine purple and six yellow arrows indicate colonization events to the TEP and the Atlantic, respectively.

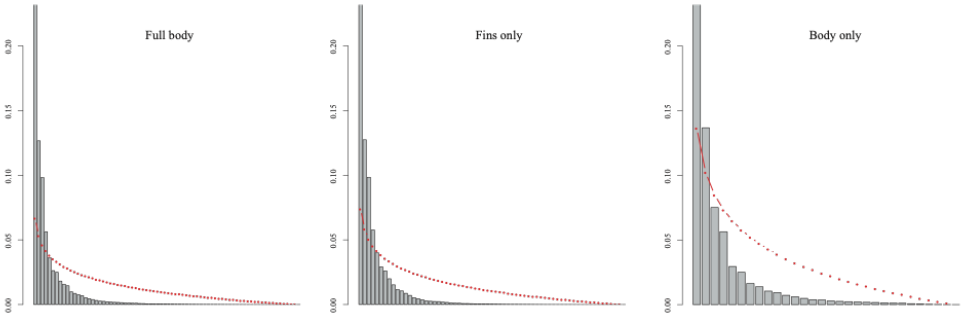

Figure A.15: Plots of the broken-stick method showing PC axis variation (x axis) for full body-shape, fins-only, and body-only datasets. Plots represent (red dashed lines) the broken stick distributions and (grey bars) the relative proportions of the variation that are summarized by all the PCs for each alternative dataset. The first value where the estimated broken stick value is larger than the observed variation summarized by that PC determines the optimal number of PCs axes.

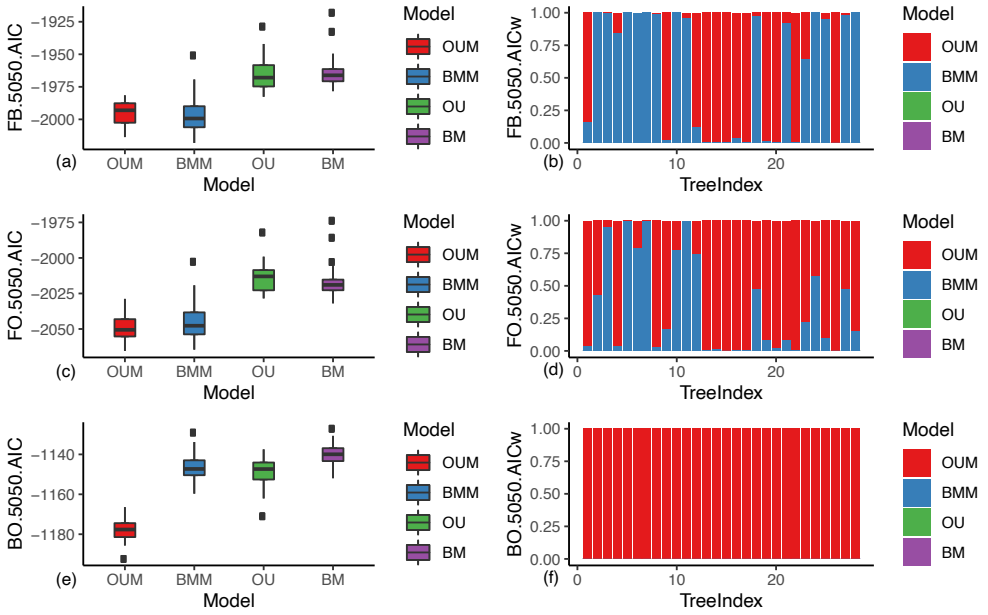

Figure A.16: Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.5 benthic/0.5 midwater probability scheme. Distribution of the Akaike Information Criterion (AIC) and Akaike weight (AICw) values for the three alternative models of continuous trait evolution (BM, OU, BMM, AND OUM) applied to the (a,b) body-only, (c,d) full body-shape, and (e,f) fins-only datasets.

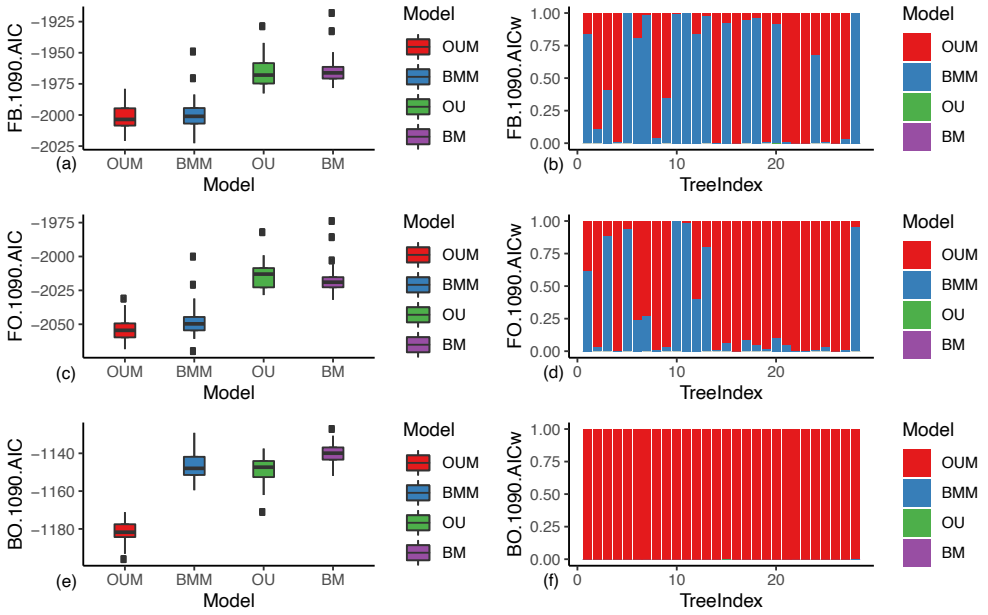

Figure A.17: Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.1 benthic/0.9 midwater probability scheme. Distribution of the Akaike Information Criterion (AIC) and Akaike weight (AICw) values for the three alternative models of continuous trait evolution (BM, OU, BMM, AND OUM) applied to the (a,b) body-only, (c,d) full body-shape, and (e,f) fins-only datasets.

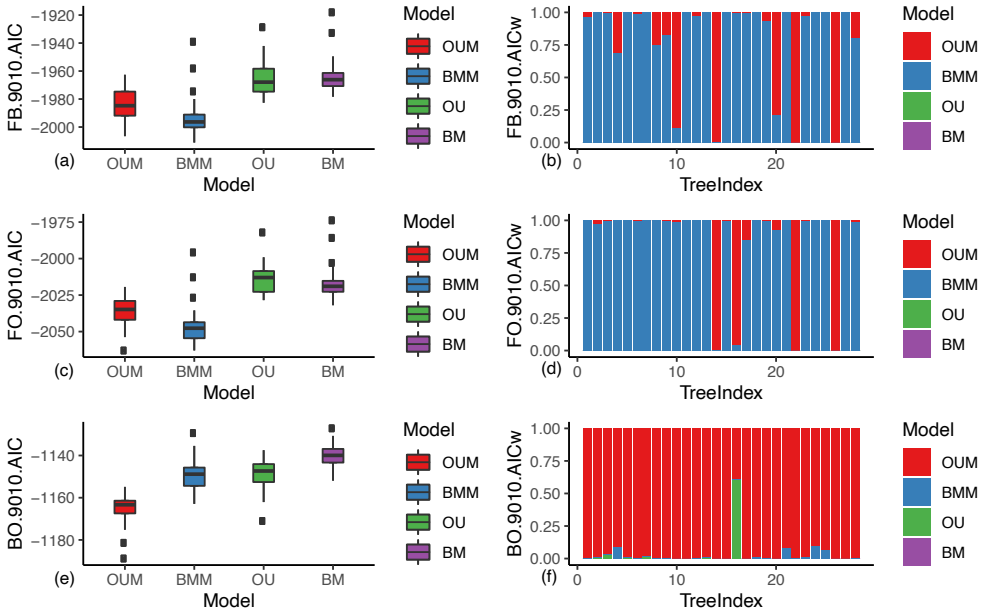

Figure A.18: Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.9 benthic/0.1 midwater probability scheme. Distribution of the Akaike Information Criterion (AIC) and Akaike weight (AICw) values for the three alternative models of continuous trait evolution (BM, OU, BMM, AND OUM) applied to the (a,b) body-only, (c,d) full body-shape, and (e,f) fins-only datasets.

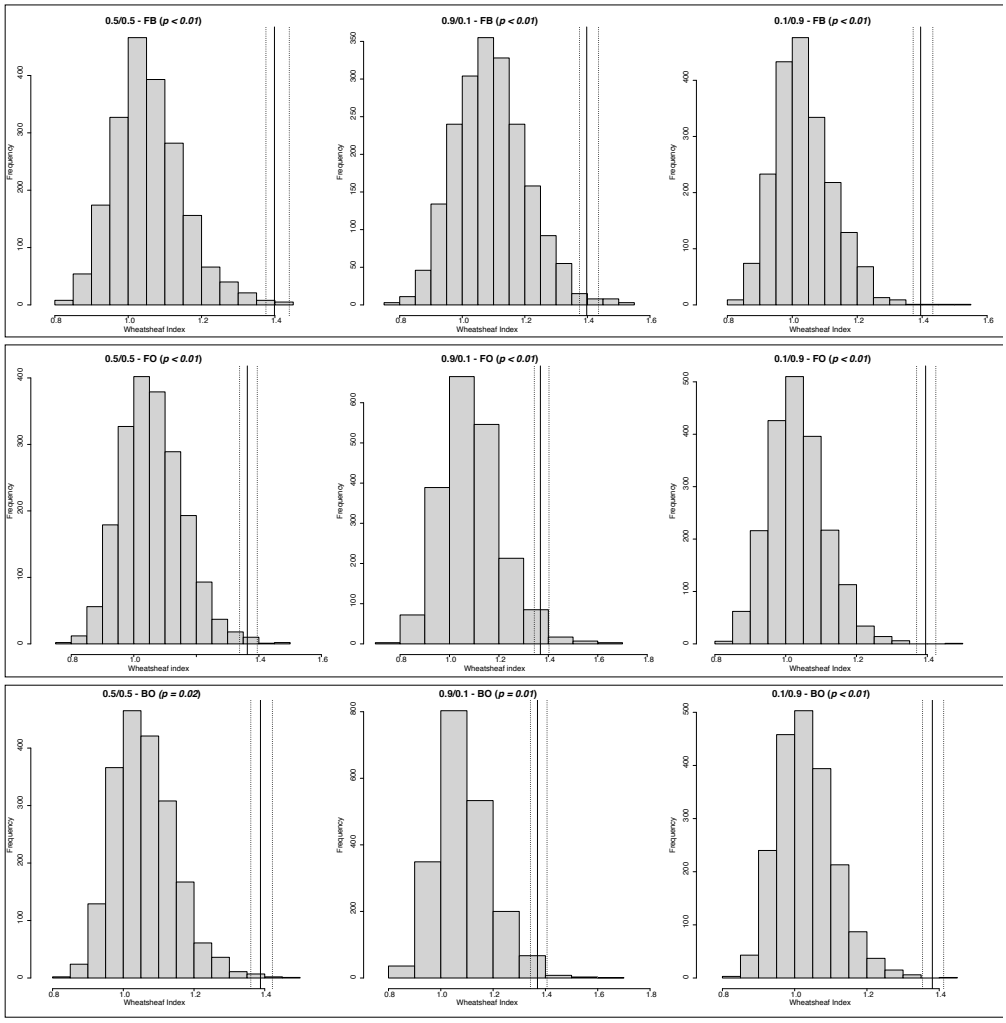

Figure A.19: Histograms representing the distribution of bootstrapped Wheatsheaf index values for all morphometric datasets based on PCA and the three different habitat coding schemes. Black thick lines represent the calculated Wheatsheaf index. Dashed lines show 95% confidence interval.

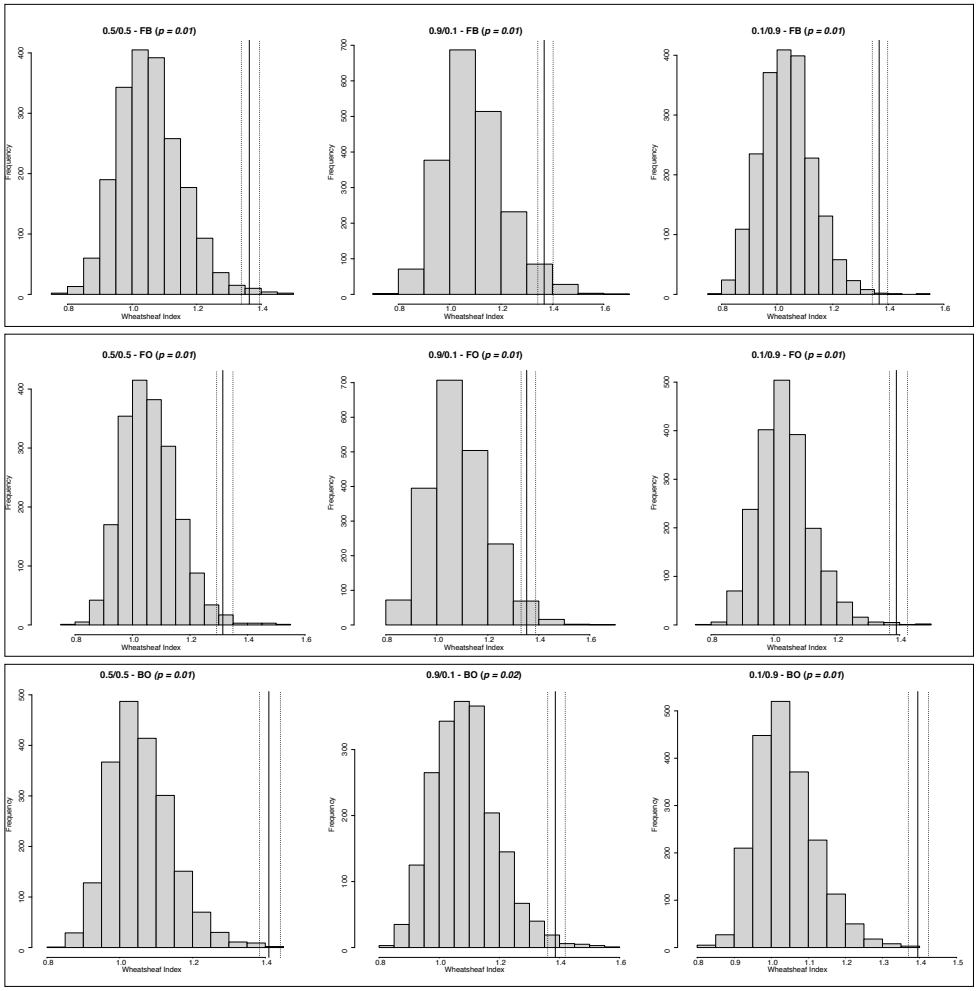

Figure A.20: Histograms representing the distribution of bootstrapped Wheatsheaf index values for all morphometric datasets based on pPCA and the three different habitat coding schemes. Black thick lines represent the calculated Wheatsheaf index. Dashed lines show 95% confidence interval.

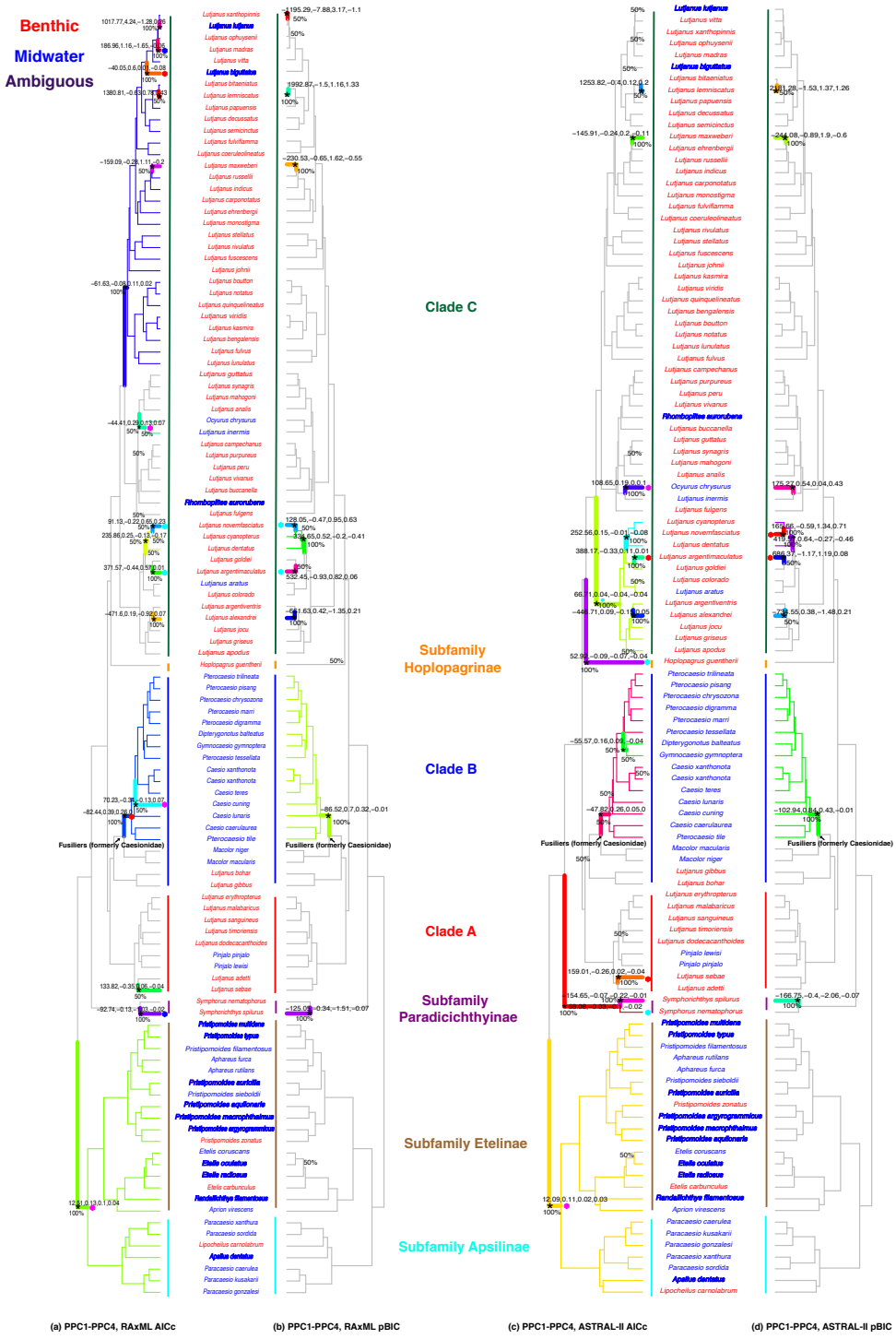

Figure A.21: Adaptive and convergent shifts in Lutjanidae for the full-body shape dataset (first four PC axes) with ℓ1ou using the AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

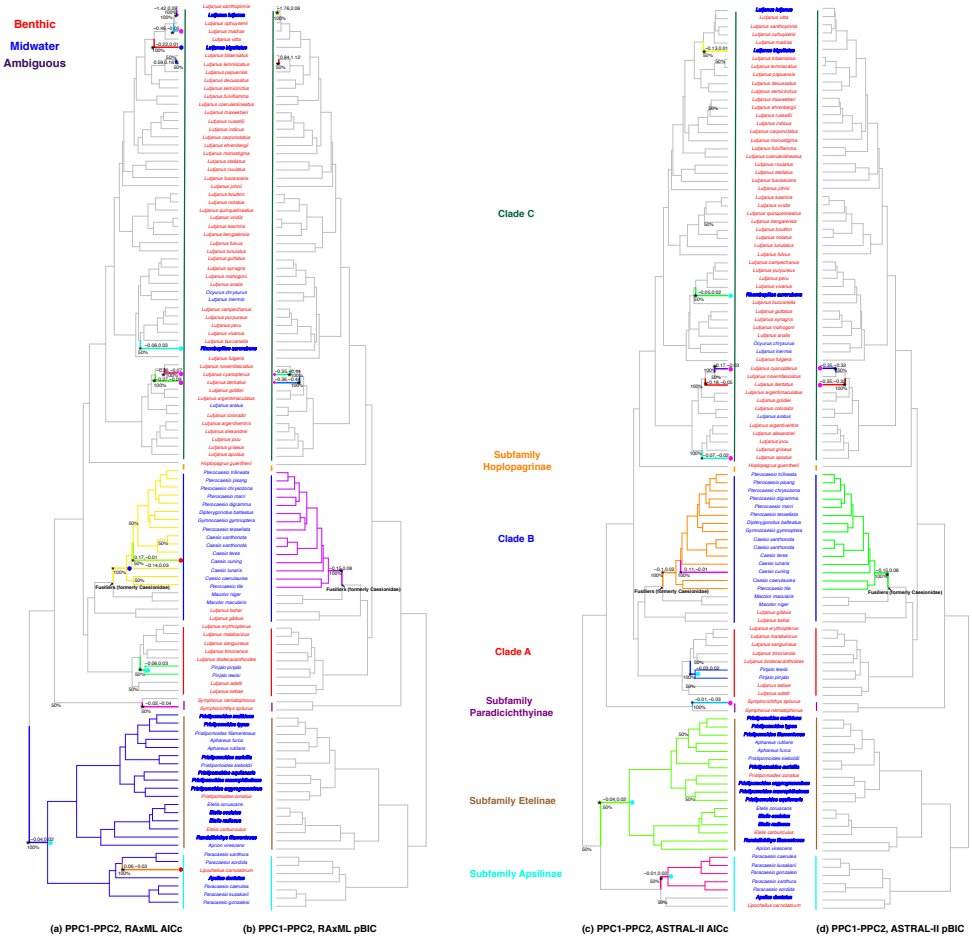

Figure A.22: Adaptive and convergent regimes in Lutjanidae for the body-only dataset (first two PC axes), with ℓ1ou using both AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

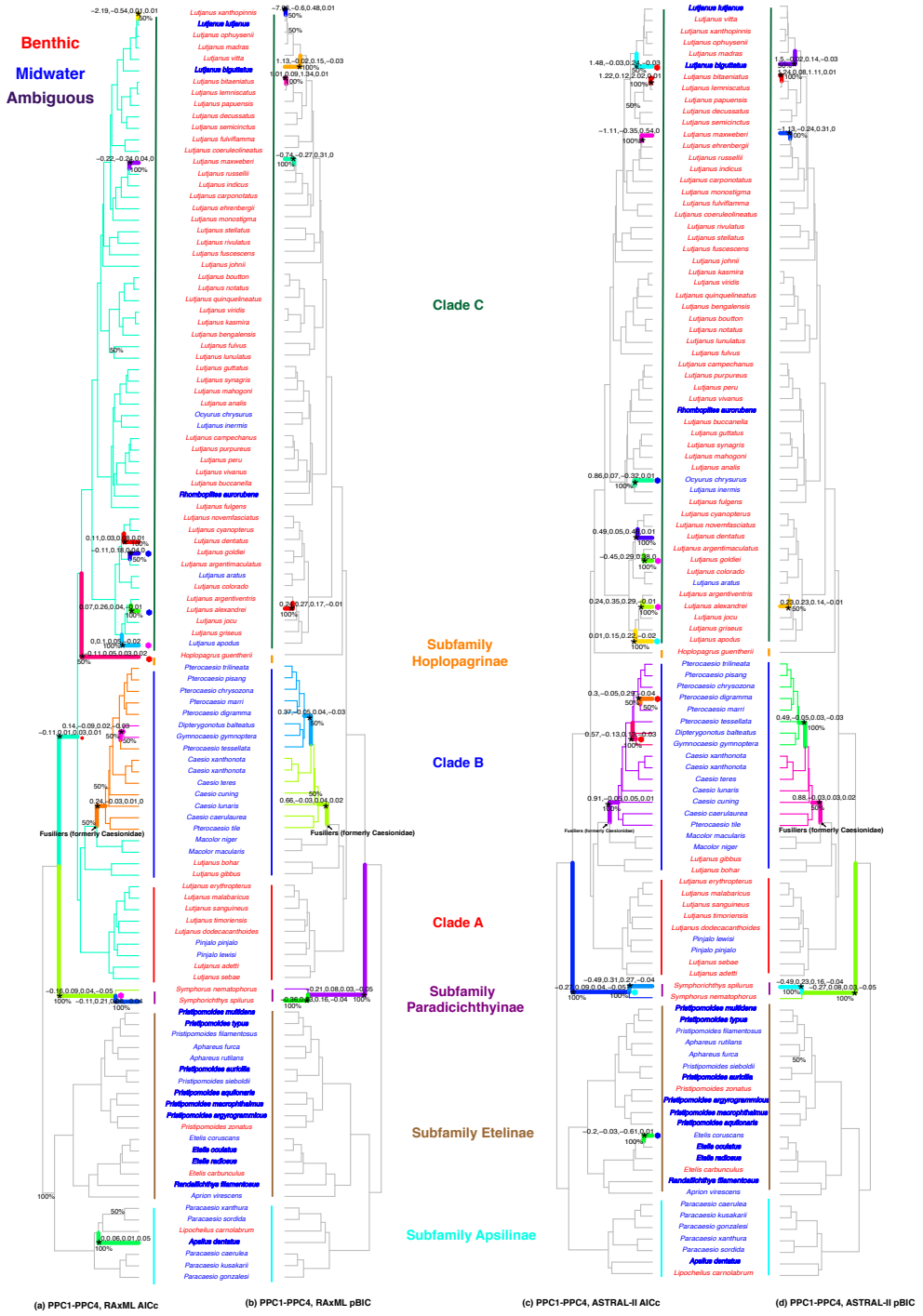

Figure A.23: Adaptive and convergent regimes in Lutjanidae for the fins-only dataset (first four PC axes), with ℓ1ou using both AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

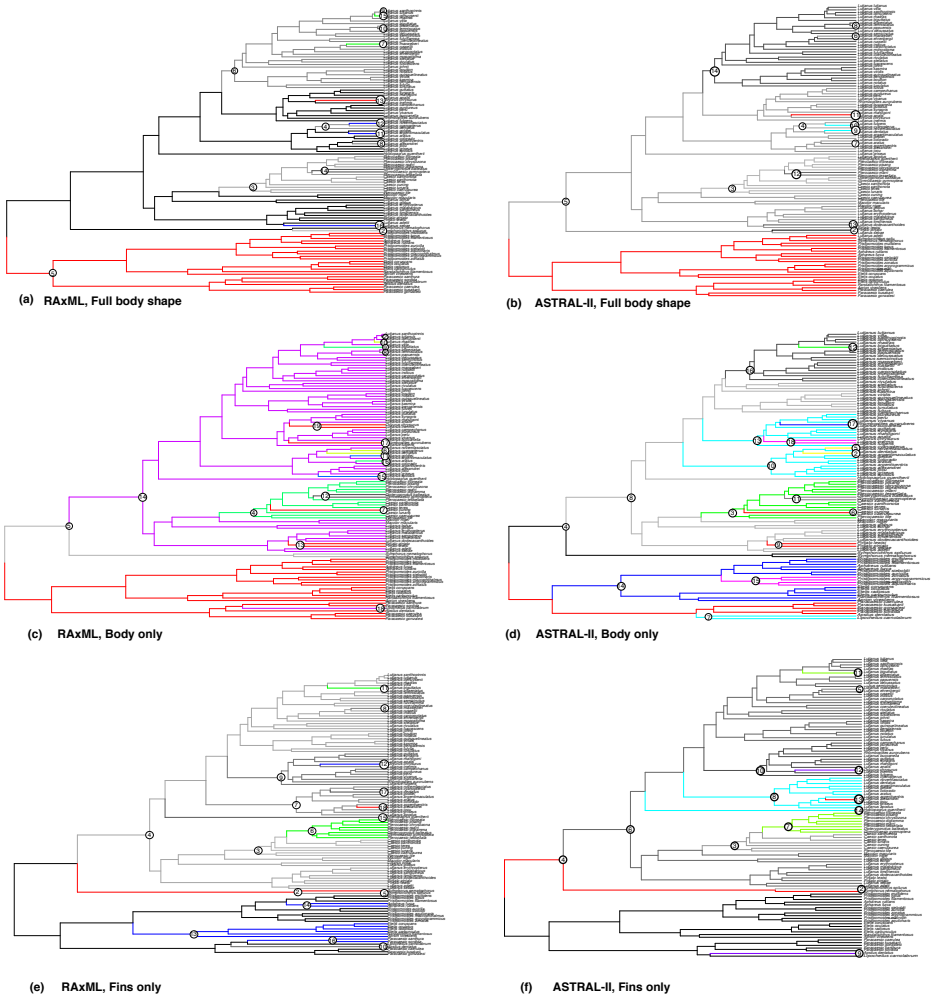

Figure A.24: Adaptive and convergent shifts in Lutjanidae for the (a) full-body shape (b) body-only, and (c) fins-only datasets with SURFACE and AICc. Numbers indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum value.

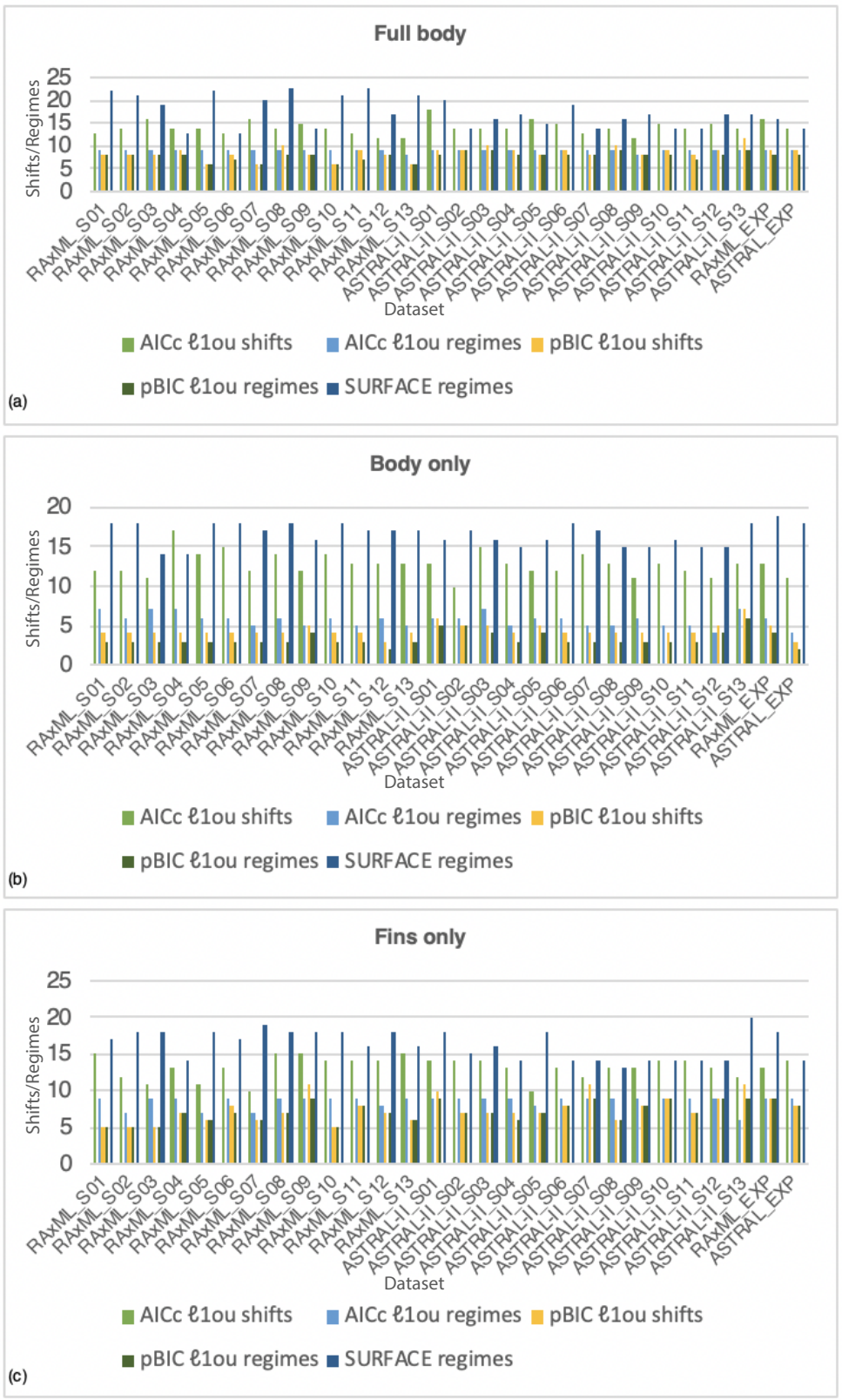

Figure A.25: ℓ1ou and SURFACE results for each alternative morphometric dataset: a) full body shape, b) body only, and c) fins only. Each panel shows the number of shifts and regimes for AICc and pBIC models from ℓ1ou analyses, and the number of regimes for SURFACE analyses.

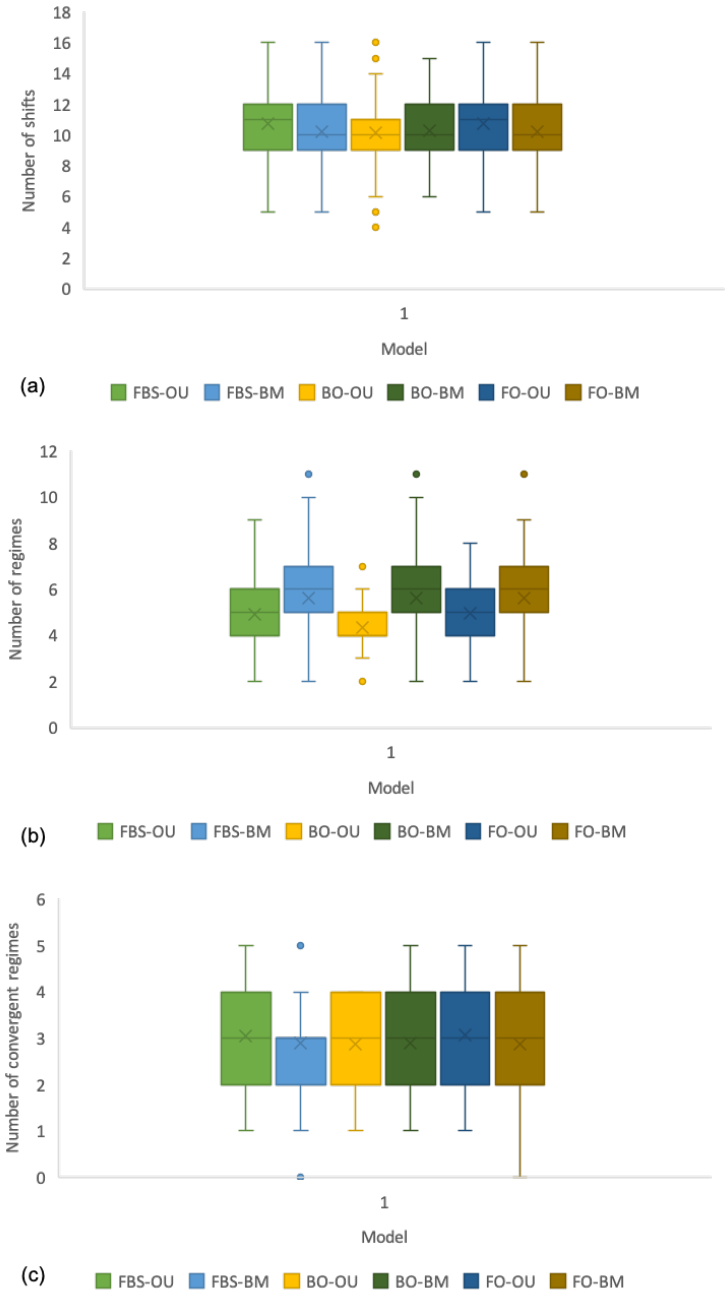

Figure A.26: Number of shifts, regimes, and convergent regimes identified with the empirical dataset against 99 simulated BM and OU model distributions for single PC axes. FBS: Full body- shape; BO: Body only; FO: Fins only. a. Number of shifts; b. number of regimes; c. number of convergent regimes.

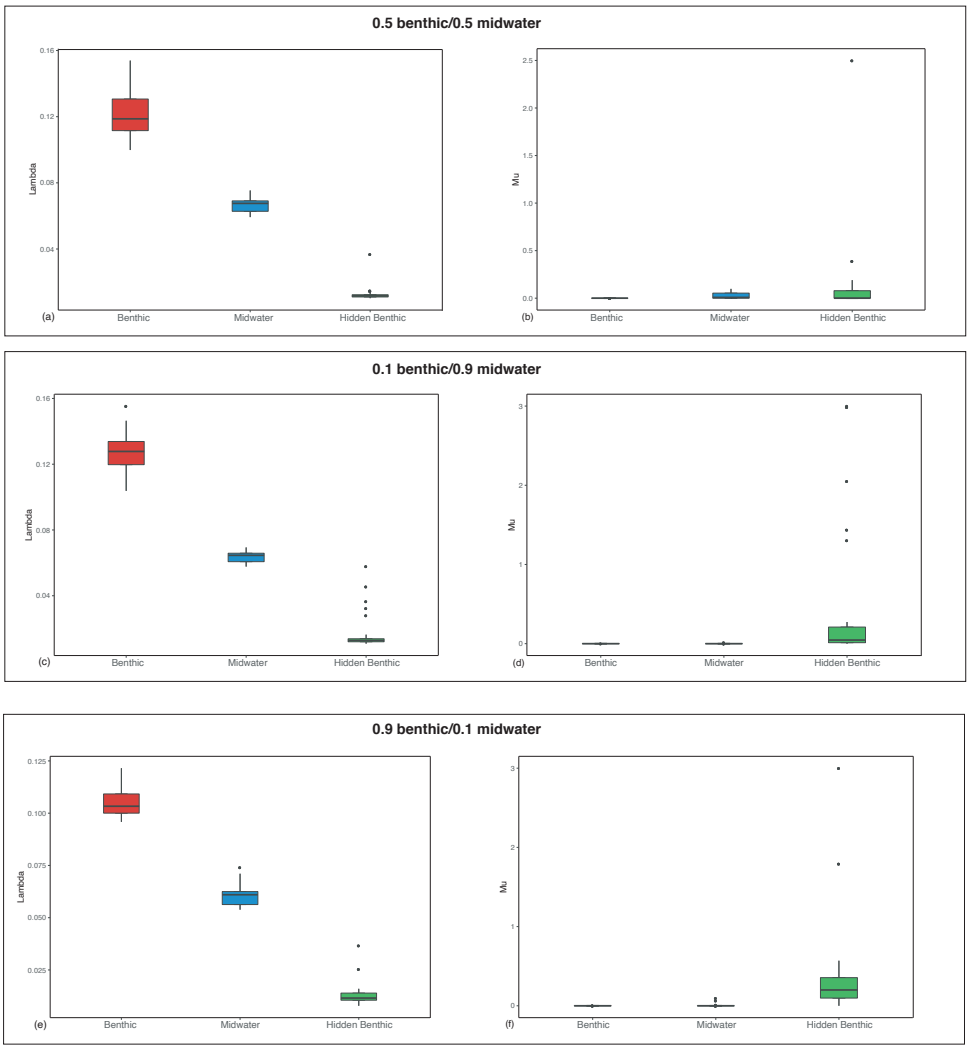

Figure A.27: Distribution of diversification rates estimated under the 'HiSSE benthic' model for the 28 trees based on three alternative habitat coding schemes. (a) Net-diversification (speciation minus extinction), (b) speciation, and (c) extinction parameters obtained for midwater (blue) benthic (red) habitats, including a hidden state (pink) associated with benthic lineages.

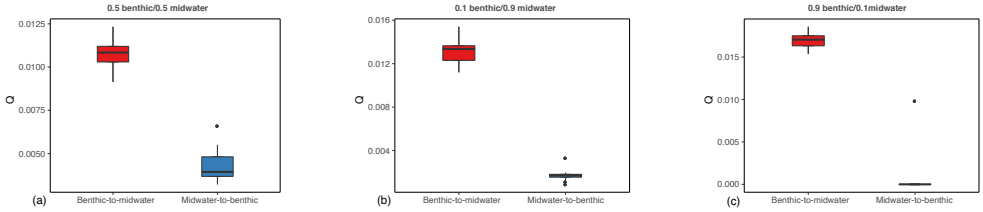

Figure A.28: Distribution of transition rates between the different habitat states estimated under the 'HiSSE benthic' model for the 28 trees using the three alternative habitat coding schemes.

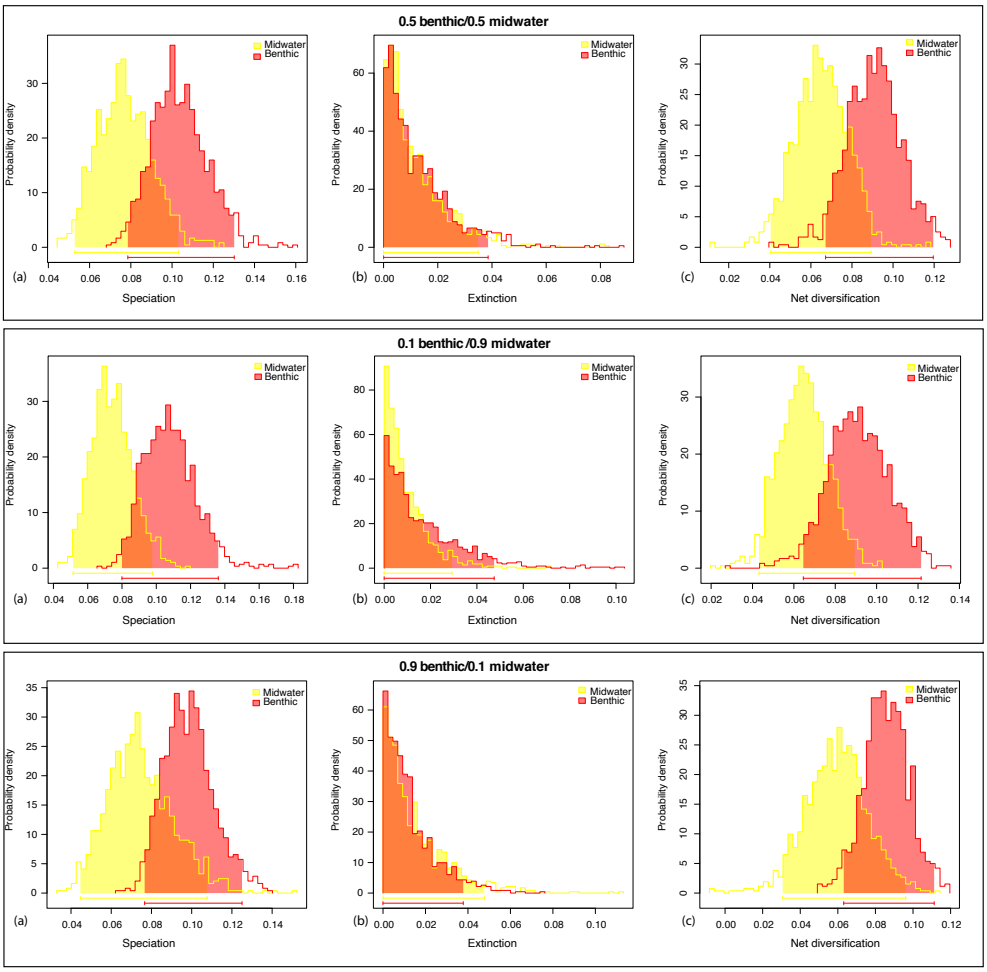

Figure A.29: Marginal distribution of diversification rates obtained using MCMC-based BiSSE analyses applied to the 'master tree', based on the three alternative habitat coding schemes. Estimated (a) speciation, (b) extinction, and (c) net-diversification (speciation minus extinction) parameters for benthic (yellow) and midwater (red) lineages.

References

Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(SUPPL. 2):7–13.

Adams, D. C. and Collyer, M. L. (2019). Phylogenetic comparative methods and the evolution of morphological integration. *Annual Review ofEcology, Evolution, and Systematics*, (25):50:405.

Agassiz, L. (1833). *Recherches sur les poissons fossiles*. Imprimerie de Petipierre, Neuchatel (Suisse).

Alfaro, M. E., Faircloth, B. C., Harrington, R. C., Sorenson, L., Friedman, M., Thacker, C. E., Oliveros, C. H., Černý, D., and Near, T. J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary.

Allen, J. M., Boyd, B., Nguyen, N. P., Vachaspati, P., Warnow, T., Huang, D. I., Grady, P. G., Bell, K. C., Cronk, Q. C., Mugisha, L., Pittendrigh, B. R., Leonardi, M. S., Reed, D. L., and Johnson, K. P. (2017). Phylogenomics from whole genome sequences using aTRAM. *Systematic Biology*, 66(5):786–798.

Arbuckle, K., Bennett, C. M., and Speed, M. P. (2014). A simple measure of the strength of convergent evolution. pages 685–693.

Bannikov, A. F. (2006). Fishes from the Eocene of Bolca, northern Italy, previously classified in the Sparidae, Serranidae and Haemulidae (Perciformes). *Geodiversitas*, 28(2):249–275.

Bannikov, A. F. (2014). The systematic composition of the Eocene actinopterygian fish fauna from Monte Bolca, northern Italy, as known to date. *Miscellanea paleontologica*, 12(January):22–34.

Bellwood, D. R. and Wainwright, P. C. (2002). The History and Biogeography of Fishes on Coral Reefs. In *Coral Reef Fishes*, pages 5–32.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton, J. C., Zhang, F., Buser, T., Campbell, M. a., 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., Betancur-R., R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. a., Ballesteros, J. a., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, Apr 18(APR 2013):1–45.

Betancur-R, R., Ortí, G., and Pyron, R. A. (2015). Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes. *Ecology Letters*.

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., and Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1):162.

Bolger, A., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Carnevale, G., Bannikov, A. F., Marramà, G., and Tyler, J. C. (2014). The Pesciara-Monte Postale Fossil Lagerstatte: 2. Fishes and other vertebrates. In Rendiconti della Società Paleontologica Italiana, editor, *The Bolca Fossil-Lagerstätten: A window into the Eocene World*, number 4, pages 37–63.

Carpenter, K. (1987). Revision of the Indo-Pacific fish family Caesionidae (Lutjanoidea), with descriptions of five new species. *Indo-Pacific Fishes*, 15(56).

Carpenter, K. (2001). Caesionidae: Fusiliers. In *The Living Marine Resources of the Western Central Pacific*, pages 2919–2941. Rome, fao edition.

Carpenter, K. E. (1988). Vol.8. Fusiliers fishes of the world. In *FAO Species Catalogue*. Rome.

Carpenter, K. E. (1990). A Phylogenetic Analysis of the Caesionidae (Perciformes : Lutjanoidea). *Copeia*, (3):692–717.

Carpenter, K. E. (1993). Optimal cladistic and quantitative evolutionary classifications as illustrated by fusilier fishes (Teleostei: Caesionidae). *Syst Biol*, 42(2):142–154.

Clavel, J., Escarguel, G., and Merceron, G. (2015). mvmorph : an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution*, 6(11):1311–1319.

Coates, A. and Obando, J. (1996). The geologic evolution of tge Central American Isthmus. In *Evolution and Environment in Tropical America*, pages 21–56. University of Chicago Press.

Davis, A. M. and Betancur-R., R. (2017). Widespread ecomorphological convergence in multiple fish families spanning the marine-freshwater interface. *Proceedings. Biological sciences*, 284(1854):20170565.

Dupin, J., Matzke, N. J., Särkinen, T., Knapp, S., Olmstead, R. G., Bohs, L., and Smith, S. D. (2017). Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, 44(4):887–899.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459):1427–1434.

Felsenstein, J., Ackerly, D. D., and Mcpeek, M. A. (2012). A comparative method for both discrete and continuous characters using the threshold model. *American Naturalist*, 179(2):154–56.

Fitzjohn, R. G. (2012). Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3(6):1084–1092.

Frédérich, B. and Santini, F. (2017). Macroevolutionary analysis of the tempo of diversification in snappers and fusiliers (Percomorpha: Lutjanidae). *Belgian Journal of Zoology*, 147(1):17–35.

Fricke, R., Eschmeyer, W. N., van der Laan, R., Fricke, R., and van der Laan, R. (2016). Catalog of fishes: genera, species, references.

Friedman, M. and Carnevale, G. (2018). The Bolca Lagerstätten: Shallow marine life in the eocene. *Journal of the Geological Society*, 175(4):569–579.

Froese, R. and Pauly, D. (2019). FishBase.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152.

Gold, J. R., Voelker, G., and Renshaw, M. A. (2011). Phylogenetic relationships of tropical western Atlantic snappers in subfamily Lutjaninae (Lutjanidae: Perciformes) inferred from mitochondrial DNA sequences.

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., Betancur, R., Li, C., Becker, L., Bellora, N., Zhao, X., Li, X., Wang, M., Fang, C., Xie, B., Zhoui, Z., Huang, H., Chen, S., Venkatesh, B., and Shi, Q. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences*, 115(24):201719358.

Hughes, L. C., Ortí, G., Saad, H., Li, C., White, W. T., Baldwin, C. C., Crandall, K. A., Arcila, D., and Betancur-R., R. (2020). Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *bioRxiv*, page 2020.02.18.949735.

Ingram, T. and Mahler, D. L. (2013). SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425.

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*.

Johnson, D. (1980). *The limits and relationships of the Lutjanidae and associated families.* Berkeley: University of California Press, California.

Jordan, D. S. and Evermann, B. W. (1898). *The fishes of North and Middle America: A descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama, pt. 2.* i-xxx edition.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4):772–780.

Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 7(7):811–824.

Lessios, H. A. and Robertson, D. R. (2006). Crossing the impassable: Genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society B: Biological Sciences*, 273(1598):2201–2208.

Li, B., Dettaï, A., Cruaud, C., Couloux, A., Desoutter-Meniger, M., and Lecointre, G. (2009). RNF213, a new nuclear marker for acanthomorph phylogeny. *Molecular Phylogenetics and Evolution*, 50(2):345–363.

Li, C., Hofreiter, M., Straube, N., Corrigan, S., and Naylor, G. J. P. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques*, 54(6):321–326.

Li, C., Lu, G., and Ortí, G. (2008). Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Systematic biology*, 57(4):519–539.

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

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63(6):951–970.

Meyer, B., Meusemann, K., and Misof, B. (2011). MARE v0.1.2-rc.

Mirarab, S. and Warnow, T. (2015). ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31:44–52.

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., Ayala, C., Pérez-Angel, L. C., Rodriguez-Parra, L. A., Ramirez, V., and Niño, H. (2015). Middle Miocene closure of the Central American Seaway. *Science*.

Near, T. J., Bossu, C. M., Bradburd, G. S., Carlson, R. L., Harrington, R. C., Hollingsworth, P. R., Keck, B. P., and Etnier, D. A. (2011). Phylogeny and temporal diversification of darters (Percidae: Etheostomatinae). *Systematic Biology*, 60(5):565–595.

Nelson, J., Grande, T., and Wilson, M. (2016). *Fishes of the World*. Hoboken, 5th edition.

O'Dea, A., 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., and Jackson, J. B. (2016). Formation of the Isthmus of Panama.

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., and Benton, M. J. (2012). Best practices for justifying fossil calibrations. *Systematic Biology*, 61(2):346–359.

Peres-Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, 49(4):974–997.

Pfeiffer, W. (1964). THE MORPHOLOGY OF THE OLFACTORY ORGAN OF HOPLOPAGRUS GUENTHERI GILL 1862. *Canadian Journal of Zoology*, 42(2):235–237.

Puri, H. S. and Vernon, R. O. (1959). *Summary of the geology of Florida and a guidebook to the classic exposures*. Talahassee, special pu edition.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., and Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714):392–395.

Rabosky, D. L. and Goldberg, E. E. (2017). FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution; international journal of organic evolution*.

Reis, M. D. and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian Estimation of Divergence Times. *Molecular Biology and Evolution*, 28(7):2161–2172.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution*, 63(12):3258–3268.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223.

Rocha, L. A., Robertson, D. R., Rocha, C. R., Van Tassell, J. L., Craig, M. T., and Bowen, B. W. (2005). Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, 14(13):3921–3928.

Slater, G. S. C. and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*.

Sorbini, L. (1983). *La collezione Baja di pesci e piante fossili di Bolca: con descrizione di nuovi generi e nuove specie*. University of California Press.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690.

Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology*, 57(5):758–71.

Stayton, C. T. (2005). Morphological evolution of the lizard skull: A geometric morphometrics survey. *Journal of Morphology*, 263(1):47–59.

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

Swift, C. C. and Ellwood, B. B. (1972). Hypsocephalus Atlanticus, a new genus and species of Lutjanid fish from marine Eocene limestones of northern Florida "Study of the Permian-Triassic boundary at the Hong Ngai and Lung Pu sections and complete the results from Lung Cam section for improving. In Natural History Museum, L. A. C., editor, *Contributions in Science*, pages 15–30. Los Angeles.

Uyeda, J. C., Caetano, D. S., and Pennell, M. W. (2015). Comparative Analysis of Principal Components Can be Misleading. *Systematic Biology*, 64(4):677–689.

Zerbino, D. R. and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829.

Appendix B

Appendix: Post-Cretaceous bursts of evolution along the benthic-pelagic axis in marine fishes

Trees, tables, datasets and scripts used for comparative analyses are available from the Figshare digital repository (https://doi.org/10.6084/m9.figshare.20102951.v1)

Fossil calibration and divergence time estimation

Our fossil calibration scheme follows Harrington et al. (2016), but with modifications. Lower bounds of clade age were defined via the minimum age of its earliest fossil representative (calibrations 1-16 below); 95% soft upper bounds were empirically estimated based on the maximum ages of the oldest fossil representatives of successive outgroups for each clade (Hedman, 2010). Monophyly of Pleuronectiformes (ingroup) was assumed (constrained) before analyses, following several recent studies (Harrington et al., 2016; Betancur-R. et al., 2013, 2014); placement of outgroup taxa in tree was also constrained based on current knowledge of the Fish Tree of Life (Betancur-R et al., 2013). Sequence of outgroups used by Harrington et al. (2016) to estimate soft upper bounds' 95% confidence intervals: Aulopiformes (hard lower bound: *†Atolvorator longipectoralis*; absolute age estimate: 125 Ma); non-eurypterygian Euteleostei (hard lower bound: *†Leptolepides haerteisi*; absolute age estimate: 150.94 Ma); Otocephala (hard lower bound: *†Tischlingerichthys viholi*; absolute age estimate: 150.94 Ma); Elopomorpha (hard lower bound: *†Anaethalion zapporum*; absolute age estimate: 151.2 Ma); †Ichthyodectiformes (hard lower bound: *†Occithrissops willsoni*; absolute age estimate: 166.1 Ma); *†Leptolepis coryphaenoides* (absolute age estimate: 181.7 Ma); *†Dorsetichthys bechei* (absolute age estimate: 193.81 Ma); †Pholidophoridae (hard lower bound:

†Knerichthys bronni; absolute age estimate: 221.0 Ma); *†Prohalecites porroi* (absolute age estimate: 236.0 Ma); Holostei (hard lower bound: *†Watsonulus eugnathoides*; absolute age estimate: 247.1 Ma). Hard upper bound was defined based on the stem neopterygian *†Discoserra* (absolute age estimate: 322.8 Ma).

(1) Acanthomorphata. MRCA: *Lampris*, *Myripristis*. Hard lower bound: *†Aipichthys minor*. Diagnosis and phylogenetic placement: *†Aipichthys minor* is placed on the lampridiform stem based on 67 morphological characters (Davesne et al., 2014). Stratigraphic horizon and locality: fish beds at Hadjula (Benton et al., 2015). Absolute age estimate: 98.0 Ma. 95% soft upper bound: 143.0 Ma. Prior setting: log-normal distribution, mean=2.161, SD=1.0 (crown calibration).

(2) Percomorphaceae + Holocentriformes. MRCA: *Myripristis*, *Kurtus*. Hard lower bound: *†Stichocentrus liratus*. Diagnosis and phylogenetic placement: enlarged penultimate anal-fin spine of *†Stichocentrus* represents a synapomorphy of holocentroids (Patterson, 1993). Stratigraphic horizon and locality: fish beds at Hadjula (Benton et al., 2015). Absolute age estimate: 98.0 Ma. 95% soft upper bound: 128.8 Ma. Prior setting: log-normal distribution, mean=1.782, SD=1.0 (crown calibration).

(3) Syngnathiformes. MRCA: *Mullus*, *Syngnathus*. Hard lower bound: *†Gasterorhamphosus zuppichini*. Diagnosis and phylogenetic placement: the placement of *†Gasterorhamphosus zuppichini* in Syngnathiformes is supported by the absence of anal-fin spine, enlarged dorsal-fin spine with serrated posterior margin, elongated tubular snout, absence of pleural ribs, enlarged posterodorsal process of cleithrum, rod-like anteroventral process of coracoids and simple pectoral rays (Near et al., 2012). Stratigraphic horizon and locality: "Calcari di Melissano," Porto Selvaggio, Lecce province, Italy. Absolute age estimate: 69.71 Ma. 95% soft upper bound: 98.1 Ma. Prior setting: log-normal distribution, mean=1.6975, SD=1.0 (crown calibration).

(4) Centropomidae (Latinae + Centropominae). MRCA: *Centropomus*, *Lates*. Hard lower bound: *†Eolates gracilis*. Diagnosis and phylogenetic placement: placement of *†Eolates gracilis* in Latiinae is supported by the presence of posterior pad in infraorbital 1 and by having 10+14 vertebrae (Otero and Otero, 2004); *†Eolates gracilis* is the earliest branching lineage of Latiinae (Otero and Otero, 2004). Stratigraphic horizon and locality: early Eocene, upper Ypresian, Monte Bolca, Italy. Absolute age estimate: 49 Ma. 95% soft upper bound: 72.8 Ma. Prior setting: log-normal distribution, mean=1.525, SD=1.0 (crown calibration).

(5) Menidae. MRCA: *Mene*, *Xiphias*. Hard lower bound: *†Mene purydi*. Diagnosis and phylogenetic placement: *†Mene purydi* has several menid synapomorphies, including a cavernous vault formed by the frontal bones; a pronounced supraoccipital crest ornamented with a distinctive, anteriorly-inclined ridge; well-formed sclerotic ossicles; a lateral plateau on the hyomandibula ornamented with a set of narrow striae; close association of the first and second neural arches (Friedman and Johnson, 2005). Stratigraphic horizon and locality:

northwestern Peru (Friedman and Johnson, 2005). Absolute age estimate: 55.20 Ma. 95% soft upper bound: 84.7 Ma. Prior setting: log-normal distribution, mean=1.7395, SD=1.0 (crown calibration).

(6) Echeneidae. MRCA: *Remora*, *Rachycentron*. Hard lower bound: *†Echeneidae undet.* (Friedman et al., 2013). Diagnosis and phylogenetic placement: identified as belonging to Echeneidae based on synapomorphies for the family (e.g., a dorsal adhesion disc and expanded transverse processes of vertebrae (Friedman et al., 2013). Stratigraphic horizon and locality: NP23 - the fish shales of Grube Unterfeld ("Frauenweiler"). Absolute age estimate: 29.62 Ma. 95% soft upper bound: 51.9Ma. Prior setting: log-normal distribution, mean=1.459, SD=1.0 (crown calibration).

(7) Echeneoidei. MRCA: *Echeneis*, *Scomberoides*. Hard lower bound: *†Ductor vestenae*. Diagnosis and phylogenetic placement: *†Ductor* has been diagnosed either as the sister taxon of crown Echeneoidei, or within crown Echeneoidei as sister to Rachycentridae plus Coryphaenidae. We followed Harrington et al. (2016) opinion and interpreted *†Ductor* as the sister taxon of crown Echeneoidei, as it represents a more conservative application of this fossil placement. Stratigraphic horizon and locality: early Eocene, upper Ypresian, Bolca, Italy. Absolute age estimate: 49 Ma. 95% soft upper bound: 59.1 Ma. Prior setting: log-normal distribution, mean=0.668, SD=1.0 (crown calibration).

(8) Scomberoidini. MRCA: *Scomberoides*, *Trachinotus*. Hard lower bound: *†Scomberoides spinosus*. Diagnosis and phylogenetic placement: *†Scomberoides spinosus* has been identified as an Scomberoidini based on two synapomorphies: 26 vertebrae (Smith-Vaniz, 1984), and posterior fin rays of the dorsal and anal fin developed as finlets (Bannikov, 1990). Stratigraphic horizon and locality: upper Maikop at Chernaya Rechka, Caucasus (Bannikov and Parin, 1997). Absolute age estimate: 19.30 Ma. 95% soft upper bound: 50.9 Ma. Prior setting: log-normal distribution, mean=1.8082, SD=1.0 (crown calibration).

(9) Carangini. MRCA: *Seriola*, *Chloroscombrus*. Hard lower bound: *†Eastmanalepes primaevus*. Diagnosis and phylogenetic placement: the presence of scutes along its flank is a synapomorphy of Carangini within Carangidae. Stratigraphic horizon and locality: early Eocene, upper Ypresian, Bolca, Italy. Absolute age estimate: 49 Ma. 95% soft upper bound: 59.1 Ma. Prior setting: log-normal distribution, mean=0.668, SD=1.0 (crown calibration).

(10) Pleuronectiformes. MRCA: *Psettodes*, *Bothus*. Hard lower bound: *†Heteronectes chaneti*. Diagnosis and phylogenetic placement: *Heteronectes* has been placed as the earliest-branching lineage in the flatfish stem, supported by a set of 58 morphological characters (Friedman, 2008). Stratigraphic horizon and locality: early Eocene, upper Ypresian, Bolca, Italy. Absolute age estimate: 49 Ma. 95% soft upper bound: 72.8 Ma. Prior setting: log-normal distribution, mean=1.525, SD=1.0 (stem calibration).

(11) Bothoid. MRCA: *Symphurus*, *Bothus*. Hard lower bound: *†Eobothus minimus*. Diagnosis and phylogenetic placement: Friedman (Friedman, 2008) provides apomorphy-based

evidence for placement of *†Eobothus minimus* in Pleuronectoidei. Eobothus also shows several derived features (e.g., loss of pelvic-fin spine, anteriorly inclined neural spine of second abdominal vertebra) common to the four 'bothoid' families (Scophthalmidae, Bothidae, Pleuronectidae, and Paralichthyidae). Those characters, however, cannot resolve its relative placement among these lineages (Hoshino, 2001). Stratigraphic horizon and locality: early Eocene, upper Ypresian, Bolca, Italy. Absolute age estimate: 49 Ma. 95% soft upper bound: 58.3 Ma. Prior setting: log-normal distribution, mean = 0.648, SD= 1.0 (stem calibration).

(12) Soleidae + Cynoglossidae. MRCA: *Solea*, *Cynoglossus*. Hard lower bound: *†Eobuglossus eocenicus*. Diagnosis and phylogenetic placement: possibly an stem soleid or cynoglossid (Near et al., 2012; Chanet, 1994). Stratigraphic horizon and locality: upper Lutetian, Gebel Turah, Egypt (Chanet, 1994). Absolute age estimate: 41.2 Ma. 95% soft upper bound: 52.8 Ma. Prior setting: log-normal distribution, mean=0.815, SD=1.0 (crown calibration). Comment: Near et al. (2012) placed the calibration one node below (Samaridae + Soleidae + Cynoglossidae) arguing that "Chanet (1994) argues that *†Eobuglossus* can be identified as a soleid on the basis of the geometry of the ascending process of the blind side premaxilla. We are not convinced that the state in this fossil can be meaningfully distinguished from the condition found in cynoglossids." A placement of *†Eobuglossus* in the Soleidae + Cynoglossidae crown reconciles both Chanet (1994) and Near et al. (2012) opinions.

(13) *Scophthalmus*. MRCA: *Scophthalmus*, *Lepidorhombus*. Hard lower bound: *†Scophthalmus stamatini*. Diagnosis and phylogenetic placement: *†Scophthalmus stamatini* presents features common to Scophthalmidae (e.g., pelvic fins with long insertions that extend on to the urohyal); its placement within *Scophthalmus* is based on the presence of 11 pre-caudal vertebrae (Baciu and Chanet, 2002). Stratigraphic horizon and locality: fish shales of the lower dysodils exposed near Piatra Neamt, Romania. Absolute age estimate: 29.62 Ma. 95% soft upper bound: 51.3 Ma. Prior setting: log-normal distribution, mean=1.4314, SD=1.0 (crown calibration).

(14) Pleuronectidae. MRCA: *Hypsopsetta*, *Paralichthys*. Hard lower bound: *†Oligopleuronectes germanicus*. Diagnosis and phylogenetic placement: *†Oligopleuronectes* shares two derived features with pleuronectids: it is right-eyed and bears a lateral process on the eye-side frontal (Sakamoto et al., 2004). Stratigraphic horizon and locality: lower-Oligocene, Frauenweiler clay-pit, Germany. Absolute age estimate: 29.62 Ma. 95% soft upper bound: 45.8 Ma. Prior setting: log-normal distribution, mean=1.139, SD=1.0 (crown calibration).

(15) Bothidae + "Cyclopsettidae". MRCA: *Bothus*, *Cyclopsetta*. Hard lower bound: *†Oligobothus pristinus*. Diagnosis and phylogenetic placement: stem Bothidae based on the presence of myorhabdoi, intermuscular bones with fimbrate proximal and distal ends (Baciu and Chanet, 2002). Stratigraphic horizon and locality: upper Rupelian, Lower Dysodilic shales, Piatra Neamt, Romania (Baciu and Chanet, 2002). Absolute age estimate: 29.62

Ma. 95% soft upper bound: 32.6 Ma. Prior setting: log-normal distribution, mean= 0.685, SD=1.0 (crown calibration).

(16) Bothus. MRCA: *Bothus*, *Asterorhombus*. Hard lower bound: *†Bothus sp.* Diagnosis and phylogenetic placement: *†Bothus sp.* is diagnosed as a belonging in the genus *Bothus* based on the presence of robust, rectangular haemal spines (Chanet and Sorbini, 2001). Stratigraphic horizon and locality: Middle Tsurevsky Member of the Tsurevsky Formation along the bank of the Psheka River in western North Caucasus (Carnevale et al., 2006). Absolute age estimate: 11.056 Ma. 95% soft upper bound: 32.6 Ma. Prior setting: log-normal distribution, mean= 1.4251, SD=1.0 (crown calibration).

Assessing the robustness of lineage diversification analyses based on multiple taxonomic sampling schemes and other alternative statistics

We assessed whether our limited taxonomic sampling has affected the lineage diversification analyses by performing a sequence of posterior-predictive simulations to test the absolute fit of candidate models to our data. Posterior-predictive simulations (as implemented in TESS) follow a series of steps that start with a MCMC simulation to estimate the posterior probability distribution of diversification parameters of the candidate models based on our observed dataset. The rate parameters (e.g., speciation, extinction) from the joint posterior densities, as well as priors initially used in the rate parameter estimations (e.g., sampling fraction, sampling strategy), are then used to parameterize models and simulate trees. Once these trees are simulated, the summary statistics calculated for the observed dataset are compared to the posterior-predictive distribution.

We compared the cladogenesis patterns observed in the Lineage Through Time (LTT) plots generated using the empirical tree against patterns generated using trees simulated under the three candidate models we initially used for model fitting in TESS: Constant BD, Decreasing BD, and Episodic BD (see methods for a more detailed description of the competing models). Comparisons between the LTT accumulation generated using the empirical tree show a clear deviation to the pattern expected under a Constant BD model (Figure B.5), with significantly negative values of γ (-2.99, P = 0.002). No significant deviation was observed between the LTT accumulation curve generated using the empirical tree and the patterns generated under Decreasing BD and Episodic BD, which indicates agreement with the model fitting analysis and supports a time-heterogeneous process as an explanation for the lineage diversification in Carangaria. The two time-heterogeneous models present similar patterns of lineage accumulation and resemble the pattern obtained with the empirical tree—a steep curve representing an initial burst of diversity accumulation followed by a period of lower rates of diversification. Therefore, both can be used to simulate trees with

cladogenesis patterns similar to the ones observed in the empirical tree, providing a good absolute fit to our dataset. Note that the final selection between these two models follows the model-fitting test presented in the main text.

To further test whether the limited taxonomic sampling scheme may affect our lineage diversification analyses, we expanded the predictive simulation analyses to four alternative phylogenetic scenarios: (1) a 45-taxa tree from Harrington et al. (2016); (2) a 405-taxa supermatrix tree from Rabosky et al. (2018); (3) a 508-taxa grafted tree obtained by adding two well-sampled carangarian subclades (carangoids from Alfaro et al. (2018) and flatfishes from Betancur-R et al. (2013)) into the Rabosky et al. (2018) backbone tree; and (4) a 1006-taxa imputed tree from Rabosky et al. (2018), which was populated with simulated taxa in place of missing tips. Note that the grafted tree does not intend to provide a new phylogenetic hypothesis for Carangaria; instead, it provides a synthesis of our current knowledge of their phylogeny and divergence times into the extended phylogenetic tree assembled by Rabosky et al. (2018). Additionally, crown ages for major Carangaria groups varies substantially among the independently estimated time trees. To address this issue, we recalibrated those trees by adjusting Carangaria's crown group age to reflect the age obtained in our analysis. In all cases, lineage accumulation scenarios resemble the pattern obtained by the exponentially Decreasing BD model (Figure B.6). Therefore, this model also presents a good absolute fit to the alternative datasets regardless of taxonomic sampling schemes used. Finally, we used the aforementioned, independently estimated phylogenetic hypotheses to generate lineage-through-time (LTT) plots and to estimate the Lineage Diversification Index (LDI) statistic (Rabosky et al., 2018). These statistics provides an alternative avenue to examine whether the cladogenetic history of Carangaria is better explained by pure-birth (LDI 0), early burst (positive LDIs), or recent speciation (negative LDIs) processes. As expected, and regardless of the taxonomic coverage, all alternative sampling schemes present positive LDI values (ranging from 0.24 to 0.83; Figure B.7). These results indicate that an early burst of diversification represents the most likely diversification scenario for the clade, strongly suggesting that our analyses remain unaffected by the use of alternative trees and methods.

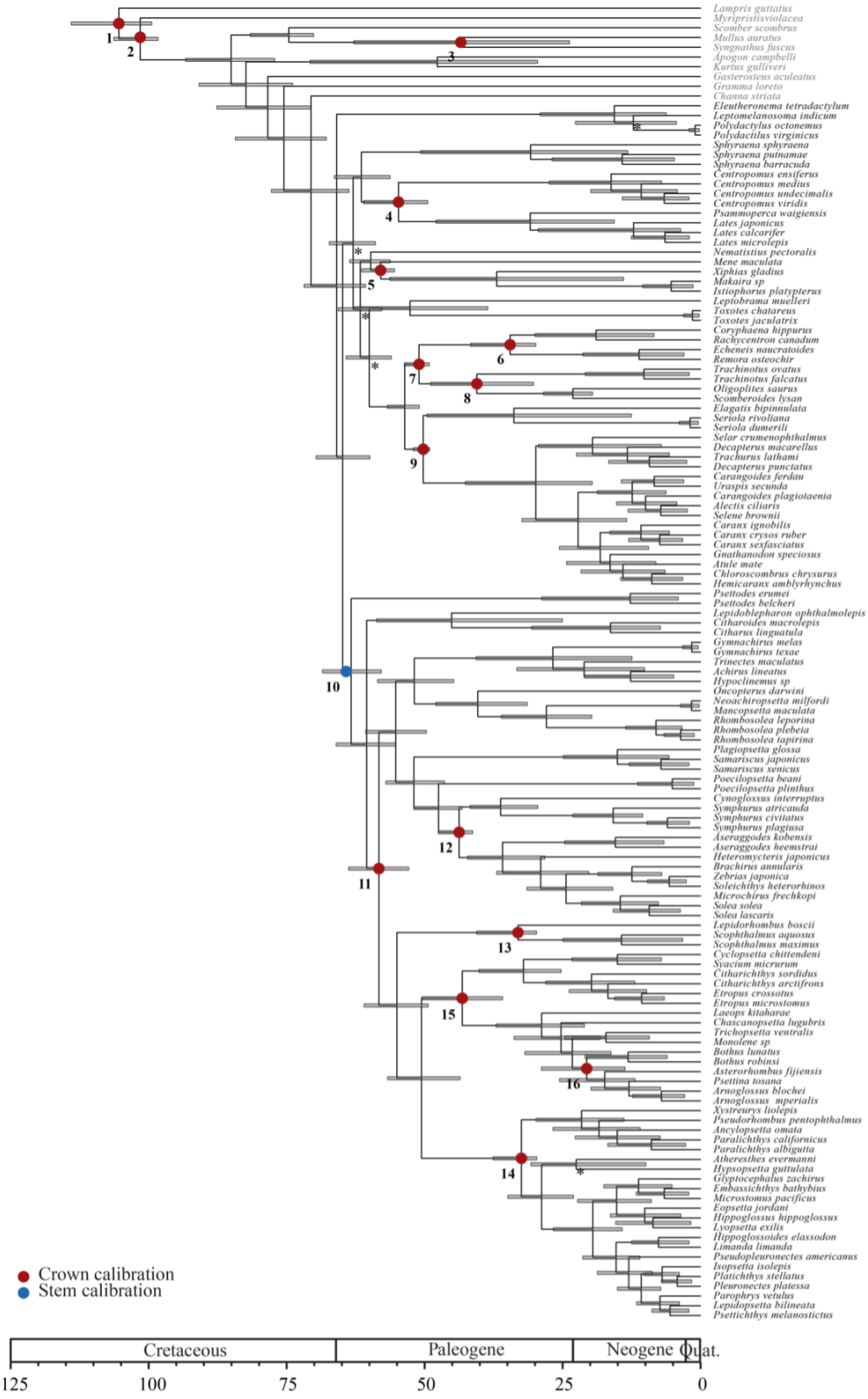

Figure B.1: Maximum Clade Credibility (MCC) tree obtained with BEAST indicating the placement of calibrations used (following the order given in Supplementary Methods). Bars represent the 95% highest posterior credibility intervals of divergence times. Asterisks (*) indicates nodes with low Bayesian posterior probability support (<0.70).

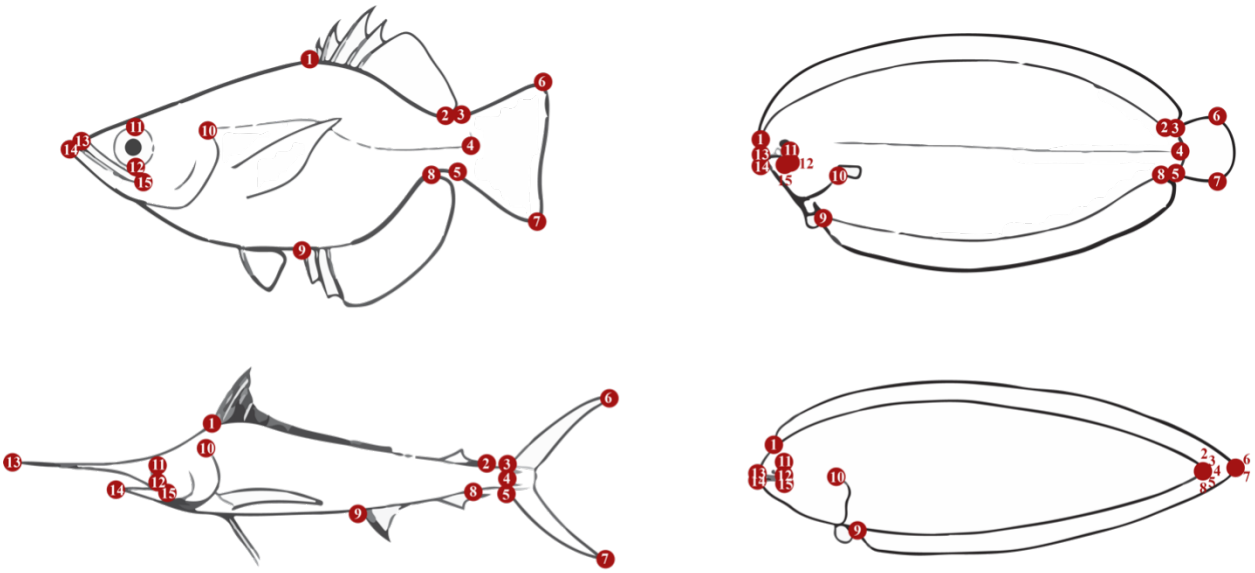

Figure B.2: Landmarks adapted from Chakrabarty (2005) : (1) anterior insertion of dorsal fin, (2) posterior insertion of dorsal fin, (3) dorsal insertion of caudal fin, (4) caudal border of hypural plate aligned with lower lateral line, (5) ventral insertion of caudal fin, (6) posterior end of the most dorsal caudal ray, (7) posterior end of the most ventral caudal ray, (8) posterior insertion of anal fin, (9) anterior insertion of anal fin, (10) dorsal end of the opercula, (11) dorsal margin of the eye, (12) ventral margin of the eye, (13) rostral tip of premaxilla, (14) anterior tip of mandible, and (15) caudal end of maxilla.

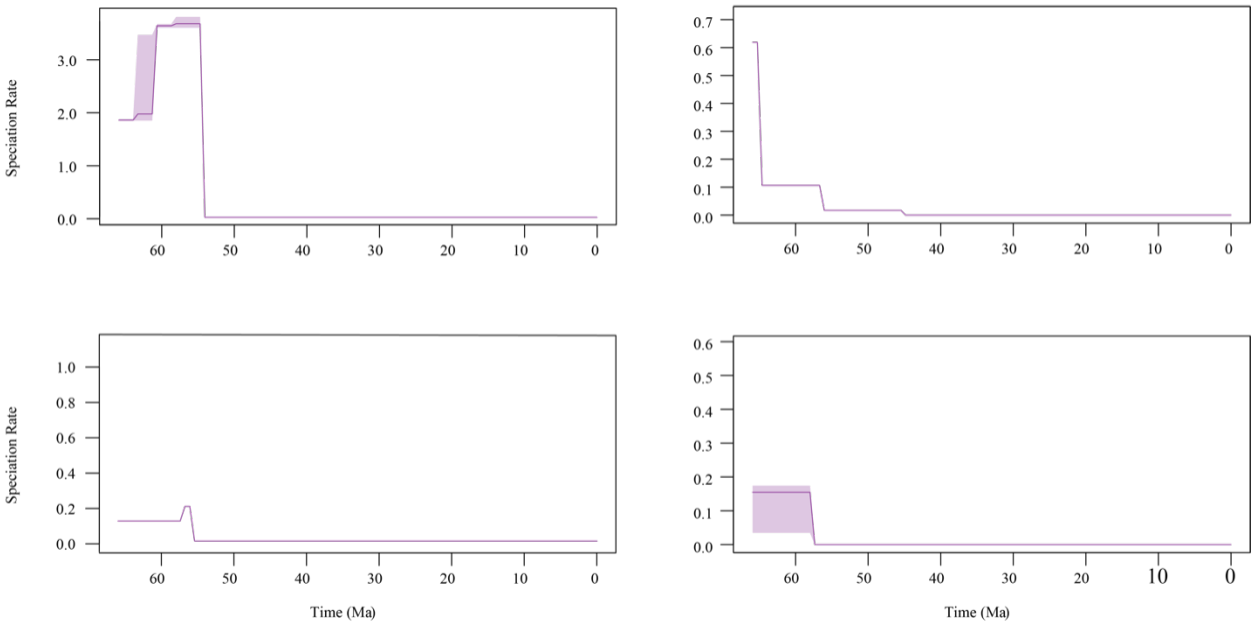

Figure B.3: Rates of speciation through time estimated from the MCC tree using the CoMET function in TESS. Plots represent the inconsistence of CoMET results and its highly sensitivity to the choice of hyper-priors. All analyzes used a minimum threshold value of effective samples (ESS) of 500.

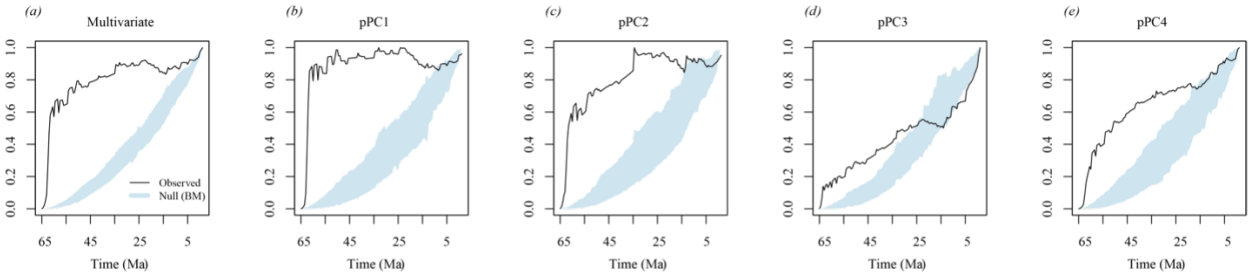

Figure B.4: Accumulation of disparity through time for the first four phylogenetically corrected PC axis (pPCA): (a) multivariate disparity; (b) pPC1, (c) pPC2, (d) pPC3, (e) pPC4. Black lines represent observations; blue shaded areas show the 95% confidence intervals of the constant rate Brownian motion model.

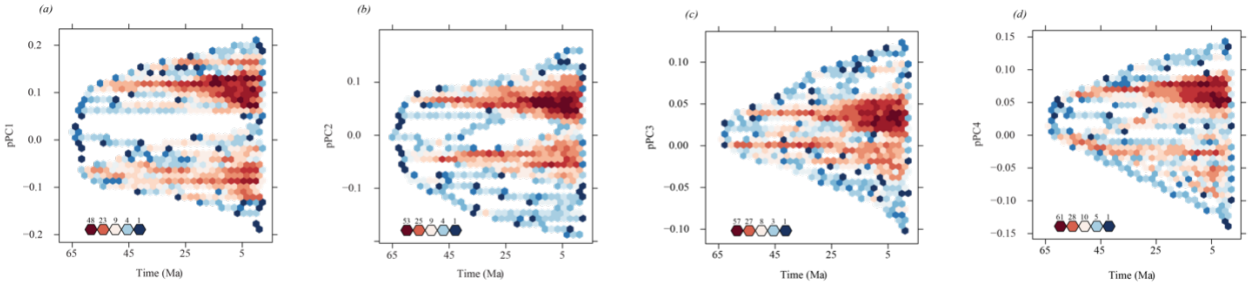

Figure B.5: Traitgram plots representing the process of morphospace filling through time in Carangaria based on the first four phylogenetically corrected PC axes (pPCA): (a) pPC1; (b) pPC2; (c) pPC3; (d) pPC4. Densities were calculated using ancestral state reconstructions derived from univariate rate-heterogeneous models of continuous trait evolution, in 1 Myr time slices.

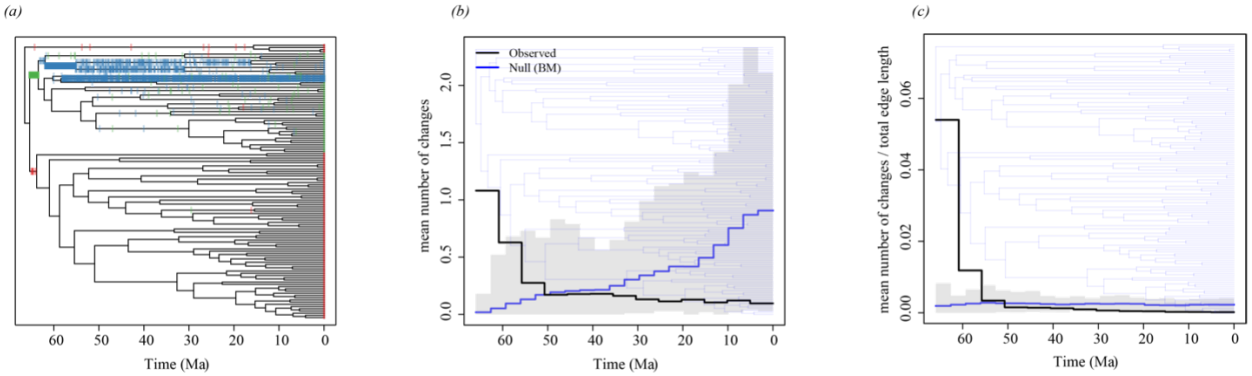

Figure B.6: Habitat transitions though time. (a) Stochastic mapping of ecological transitions calculated from 1000 SIMMAP replicates on the maximum clade credibility tree. (b) Average number of changes for in 5 Myr slices (black line represents observations; blue lines indicate mean values for the null expectations under a constant rate of character evolution; shaded area represents the 95% confidence intervals). (c) Rate of habitat transitions in 5 Myr time slices—mean number of changes per time segment, divided by the total edge length encompassed by the segment.

References

Alfaro, M. E., Faircloth, B. C., Harrington, R. C., Sorenson, L., Friedman, M., Thacker, C. E., Oliveros, C. H., Černý, D., and Near, T. J. (2018). Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary.

Baciu, D. S. and Chanet, B. (2002). Les Poissons plats fossiles (teleostei: Pleuronectiformes) de l'Oligocène de Piatra neamt (Roumanie). *Oryctos*, 4:17–38.

Bannikov, A. F. (1990). Fossil carangids and apolectids of the USSR. *Trudy Paleontologicheskogo Instituta*, (244):1–106.

Bannikov, A. F. and Parin, N. N. (1997). The list of marine fishes from Cenozoic (Upper Paleocene-Middle Miocene) localities in Southern European Russia and adjacent countries. *Journal of Ichthyology*.

Benton, M. J., Donoghue, P. C. J., Asher, R. J., Friedman, M., Near, T. J., and Vinther, J. (2015). Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica*.

Betancur-R, R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Cureton, J. C., Zhang, F., Buser, T., Campbell, M. a., 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., Betancur-R., R., Broughton, R. E., Wiley, E. O., Carpenter, K., López, J. A., Li, C., Holcroft, N. I., Arcila, D., Sanciangco, M., Ii, J. C. C., Zhang, F., Campbell, M. a., Ballesteros, J. a., Roa-varon, A., Willis, S., Borden, W. C., Hough, D. J., and Lu, G. (2013). The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*, Apr 18(APR 2013):1–45.

Betancur-R., R., Henhong, C. L. I., Unroe, T. H. A. M., Allesteros, J. E. A. B., Rtí, G. U. O., Betancur, R., Li, C., Munroe, T. a., Ballesteros, J. a., and Ortí, G. (2013). Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). *Systematic Biology*, 62(5):763–785.

Betancur-R., R., Ortí, G., Betancur-R, R., Ortí, G., Betancur-R., R., and Ortí, G. (2014). Molecular evidence for the monophyly of flatfishes (Carangimorpharia: Pleuronectiformes). *Molecular Phylogenetics and Evolution*, 73(1):18–22.

Carnevale, G., Bannikov, A. F., Landini, W., and Sorbini, C. (2006). Volhynian (early Sarmatian sensu lato) fishes from Tsurevsky, North Caucasus, Russia. *J. Paleontol.*

Chakrabarty, P. (2005). Testing Conjectures about Morphological Diversity in Cichlids of Lakes Malawi and Tanganyika. *Copeia*, 2005(2):359–373.

Chanet, B. (1994). Eobuglossus eocenicus (Woodward , 1910) from the Upper Lutetian of Egypt , one of the oldest soleids [Teleostei , Pleuronectiformi]. *Neues Jahrbuch für Geologie und Paläontologie*.

Chanet, B. and Sorbini, C. (2001). A male fish Bothus podas (Delaroche, 1809)[Pleuronectiformes: Bothidae] in the Pliocene of the Marecchia river (Italy). *BOLLETTINO-SOCIETA PALEONTOLOGICA ITALIANA*, 40(3):345–350.

Davesne, D., Friedman, M., Barriel, V., Lecointre, G., Janvier, P., Gallut, C., and Otero, O. (2014). Early fossils illuminate character evolution and interrelationships of Lampridiformes (Teleostei, Acanthomorpha). *Zoological Journal of the Linnean Society*.

Friedman, M. (2008). The evolutionary origin of flatfish asymmetry. *Nature*, 454(July):209–212.

Friedman, M., Johanson, Z., Harrington, R. C., Near, T. J., and Graham, M. R. (2013). An early fossil remora (Echeneoidea) reveals the evolutionary assembly of the adhesion disc. *Proceedings. Biological sciences / The Royal Society*, 280(1766):20131200.

Friedman, M. and Johnson, G. D. (2005). A New Species of Mene (Perciformes: Menidae) from the Paleocene of South America, with Notes on Paleoenvironment and a Brief Review of Menid Fishes. *Journal of Vertebrate Paleontology*, 25(4):770–783.

Harrington, R. C., Faircloth, B. C., Eytan, R. I., Smith, W. L., Near, T. J., Alfaro, M. E., and Friedman, M. A. (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evolutionary Biology*, pages 1–14.

Hedman, M. M. (2010). Constraints on clade ages from fossil outgroups. *Paleobiology*, 36(01):16–31.

Hoshino, K. (2001). Monophyly of the Citharidae (Pleuronectoidei: Pleuronectiformes: Teleostei) with considerations of pleuronectoid phylogeny. *Ichthyological Research*.

Near, T. J., Eytan, R. I., Dornburg, a., Kuhn, K. L., Moore, J. a., Davis, M. P., Wainwright, P. C., Friedman, M., and Smith, W. L. (2012). Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences*, 109(34):13698–13703.

Otero, O. and Otero, O. (2004). Anatomy, systematics and phylogeny of both Recent and fossil latid shes (Teleostei, Perciformes, Latidae). *Society*, pages 81–133.

Patterson, C. (1993). An overview of the early fossil record of acanthomorphs. *Bulletin of Marine Science*, 52(1):29–59.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., and Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714):392–395.

Sakamoto, K., Uyeno, T., and Micklich, N. (2004). Oligopleuronectes germanicus gen. et. sp. nov., an Oligocene pleuronectid from Frauenweiler, S. Germany. *Bulletin of the National Science Museum, Tokyo*, C 30:89–94.

Smith-Vaniz, W. F. (1984). Carangidae: relationships. In *Ontogeny and Systematics of Fishes*, pages 522–530.

Appendix C

Appendix: Phylogenomic and comparative genomic analyses support a single evolutionary origin of flatfish asymmetry

Trees, tables, datasets and scripts used for comparative analyses are available from the Figshare digital repository (https://doi.org/10.6084/m9.figshare.20102951.v1)

Note 1 - Phylogenomic data

Reanalyzed and newly generated data

We assessed flatfish phylogenetic relationships by analyzing three genome-scale datasets: (1) (Lü et al., 2021, hereafter LEA) dataset (18 species and 1693 exon markers; see the Figshare digital repository), (2) a non-coding ultraconserved elements (UCE) dataset (53 species and 596 markers; available in the Dryad digital repository doi:10.5061/dryad.2fj55)Harrington et al. (2016), and (3) a newly generated exonic dataset that included 990 exonic markers and covered 389 carangarian species. Flatfish diversity in the new dataset is represented by a sample of 207 out of ca. 600 species, including representatives from all 15 valid families. See the electronic Figshare digital repository for detailed taxonomic sampling information. We archived the sequence data generated for this manuscript as raw reads in the NCBI Sequence Repository (PRJNA862198). Exon alignments are also available in the Figshare digital repository (dx.doi.org/10.6084/m9.figshare.20372565).

Library preparation was performed by Arbor Biosciences using the dual round ('touch-down') capture protocol. Exon capture probes were designed based on alignments of 1,104

single-copy exons for seven species strategically selected to cover Carangaria diversity (i.e., *Trinectes inscriptus, Carangoides armatus, Cynoglossus maculipinnis, Remora remora, Microstomus kitt, Rachycentron canadum, Samariscus triocellatus*)(for detailed information on probe design see Hughes et al., 2018). Legacy markers widely used in fish phylogenetics were also added to the initial probe set: TBR1, MYH6, KIAA1239, PLAGL2, PTCHD1, RIPK4, SH3PX3, SIDKEY, SREB2, ZIC1, SVEP1, GPR61, SLC10A3, UBE3A, and UBE3A-like. A total of 412 samples were sequenced at the University of Chicago (https://fgf.uchicago.edu) using one-lane Illumina HiSeq 4000 with paired-end 100 bp reads.

Assembly

We trimmed Fastq files for adapter contamination and low-quality base calls using Trimmomatic v0.365. We then mapped all reads against the reference sequences used for probe design with BWA-MEM and removed potential PCR duplicates using Samtools v1.96. We isolated mapped reads for each locus and assembled a preliminary contig for each exon using Velvet v1.2.107. Next, we used the longest assembled contig for each locus as a reference to obtain longer contigs with aTRAM 2.08. We ran aTRAM for a maximum of five iterations, using Velvet as the assembler method. Next, we removed redundant contigs with CD-Hit-EST using a 99% similarity threshold. We then aligned exons using TranslatorX, with Mafft v7.42111 as the underlying aligner algorithm. Three of the 412 newly sequenced species were excluded due low-quality sequencing results, leaving 409 newly sequenced taxa for use in downstream analysis.

Quality control (QC)

Given the pervasiveness of contamination in phylogenomic datasets, in addition to the raw data quality assessments, we implemented a multistep quality control (QC) pipeline to remove cross-contaminated sequences or misidentified species using an array of sequence- and tree-based approaches. To reduce the adverse effects of missing data in phylogenetic reconstructions, we discarded exon markers containing fewer than 75% of the total number of species. After this step, 114 of exon markers were excluded from the analysis due to high levels of missing data leaving a total of 990 markers. Next, we cross-validated species identifications by blasting COI sequences against the BOLD and NCBI public repositories using the bold_identification python script. We re-examined vouchers presenting inconsistencies between morphological and molecular identifications, which ultimately led to 15 updates in species identifications and 20 complete sample removal (sequences belonging to non-carangarian species). We then conducted reciprocal BLAST searches for all exon alignments. Only sequences in which the best-hit results corresponded with species from the same family-level clade were retained for the QC next step. Finally, we conducted branch-

length correlation (BLC) analyses. BLC works under the assumption that contaminated, mislabeled, or paralog sequences will exhibit remarkably long branches when gene-trees are constrained to a reference phylogeny. We used a preliminary concatenation-based phylogeny as a reference topology to estimate constrained gene trees for each exon alignment separately. Concatenation-based and constrained gene trees were then estimated using IQ-TREE. Gene trees with a branch-length ratio ¿5 were flagged for visual inspection, and species presenting remarkably long branches were removed from that specific exon alignment. After this final QC step, the final exonic dataset included 990 exonic markers for 389 carangarians, including 207 flatfish species. Details of the taxa and sequences removed are available in the Figshare digital repository.

Note 2 – Phylogenomic inference

Background on how GC bias affects phylogenetic inference

Evolutionary non-homogeneous processes such as base compositional non-stationarity (BCNS) have been recognized as a major source of phylogenetic error driving topological conflicts in many groups. Some of these conflicts are likely results of model misspecification, as probabilistic phylogenetic reconstruction methods are typically based on models of sequence evolution that assume homogeneous base compositions across lineages in the tree (these include the models applied by LEA). These assumptions, however, are often violated, and recent work has demonstrated that variations in GC content are especially severe among flatfish lineages and are likely a major source of phylogenetic conflict in Carangaria. This type of bias can be overcome using nucleotide substitution models that permit evolutionary rates to vary among lineages. Here we applied the GHOST model Crotty et al. (2020) of sequence evolution as implemented in IQ-TREE, which has been shown to accurately resolve tree topologies from heterotachously-evolved sequences by allowing substitution rates, branch lengths, and base frequencies to vary among lineages. Further sources of phylogenetic bias related to BCNS are discussed in the phylogenetic results section (see below).

Subset assembly

GHOST is an edge-unlinked mixture model consisting of independent site classes k, each having a separate set of model parameters and edge lengths on the same tree topology. The GHOST model is, therefore, very parameter-rich, with the total number of free parameters rapidly increasing as a function of the number of k and the number of taxa. Consequently, tree searching using our large phylogenomic data set became statistically intractable under the GHOST model due to the large number of parameters. We addressed this limitation using a diversified sampling strategy to assemble three taxa-reduced subsets (20-, 30-,

and 46 taxa) while minimizing the loss of phylogenetic diversity within Carangaria. This sampling strategy is especially important to reduce the well-known effects of long branch attraction (LBA) in in phylogenetic inference, a phylogenetic bias that is especially severe in heterotachously-evolved sequences (see below). Detailed taxonomic sampling and alignment summary statistics including the percentage of missing data, GC content, proportion of variable sites, and alignment length is available in the Figshare digital repository.

Phylogenomic reconstruction methods

For all concatenated datasets and gene alignments, we estimated ML trees in IQTREE using a homogeneous model (HM; GTR+G) and the non-homogeneous (NHM) GHOST model. For non-homogeneous analyses we used four independent k and inferred separate base frequencies for each of them (GTR+FO*H4). For concatenation-based analyses using the homogeneous model, the best-fitting partitioning scheme was determined for the complete dataset and subsets using PartitionFinder2 under a Bayesian Information Criterion (BIC). For gene tree analyses based on the homogeneous model, we used by-codon partitions based on sequence alignments from each locus. We then inferred species trees with ASTRAL-III Zhang et al. (2018) using individual IQTREE-based gene trees (both homogeneous and non-homogeneous) as input. Because gene tree inference using the full dataset is statistically intractable under the GHOST model, ASTRAL analyses using this dataset were limited to gene trees estimated under the homogeneous model. In this case, we ran an additional analysis based on gene trees in which weakly supported nodes (bootstrap values [BS] ¡ 20%) were collapsed into polytomies to reduce the effects of gene tree error (Figure C.1). For comparison, we also used non-parametric bootstrapping to summarize node support across all IQTREE and ASTRAL analyses (Ufboot and multi-locus bootstrapping, respectively).

Phylogenomic reconstruction results

Our results provide overwhelming support for a single flatfish origin (FM hypothesis). The support for the FM hypothesis was unequivocal in analyses based on non-coding ultraconserved elements (UCEs), consistently resolving the FM tree regardless of the phylogenetic reconstruction method and the model of nucleotide substitution (Figure 3.1c). Most of the concatenation-based analyses (6 out of 9 trees) using protein-coding sequences (LEA's + newly sequenced data), however, favored the flatfish polyphyly (FP) topology—likely a result of model misspecification and the failure to accommodate phylogenetic biases related to BCNS. Besides violations of the base composition homogeneity assumption (discussed above), heterotachously-evolved sequences are hypothesized to be indirectly associated with increased effects of incomplete lineage sorting (ILS), an important source of phylogenetic discordance. This is because the higher local recombination rates in GC-rich regions are as-

sociated with an extended retention time of ancestral polymorphism Romiguier et al. (2010). ILS is typically modeled by coalescent-based methods, which may explain why coalescent-based outperformed concatenation-based inferences, with 10 out of 12 ASTRAL analyses resolving the FM hypothesis.

The ASTRAL analysis using LEA's dataset based on gene trees inferred under the HM failed to recover the FM. This dataset lacks critical flatfish taxa such as the earliest-branching Pleuronectoidei lineage (Citharidae) and three out of four Carangaria orders (i.e., Carangiformes, Istiophoriformes, and Centropomiformes). Therefore, LEA's dataset may be subject to long branch attraction (LBA), a statistical bias that describes the tendency for long branches to be grouped together in phylogenetic trees. Although usually associated with parsimony inferences, LBA has been claimed to affect maximum likelihood analyses, especially when using fast-evolving GC-rich genes. Nevertheless, coalescent-based analyses of this same dataset under the NHM model are sufficient to recover the FM topology. The ASTRAL analysis using the newly sequenced complete dataset was limited to gene trees estimated under the HM due to statistical intractability. This analysis also failed to resolve FM, a result likely caused by gene tree estimation error or the limitation in accommodating BCNS. Weakly supported gene tree branches negatively impact species-tree inference, particularly when gene trees are inferred from few informative characters relative to the total number of sampled taxa. Collapsing gene-tree branches with low support or short lengths is a well-known strategy to improve species-tree accuracy by limiting the effects of gene-tree estimation error. This strategy was efficient in reducing noise and successfully recovered the phylogenetic signal supporting the FM in our full dataset (Figure C.1). Results for all datasets and subsets are presented in the Figshare digital repository.

Time calibration

We estimated divergence times for the ASTRAL tree inferred with LEA's dataset using a penalized likelihood method (treePL). Note that this analysis was performed for illustration purposes only; for a comprehensive time-calibration analysis, see Harrington et al. (2016). We applied a secondary calibration strategy using the 95% distribution of node ages recovered by (Harrington et al., 2016) as reference. Calibration points include: (1) Carangaria, MRCA Psettodes erumei and Polydactylus sextarius (min 62 Ma, max 82 Ma); (2) Pleuronectiformes, MRCA Trinectes maculatus and Polydactylus sextarius (min 55 Ma, max 70 Ma). (3) Pleuronectoidei, MRCA Trinectes maculatus and Pseudorhombus dupliciocellatus (min 52.5 Ma, max 65 Ma).

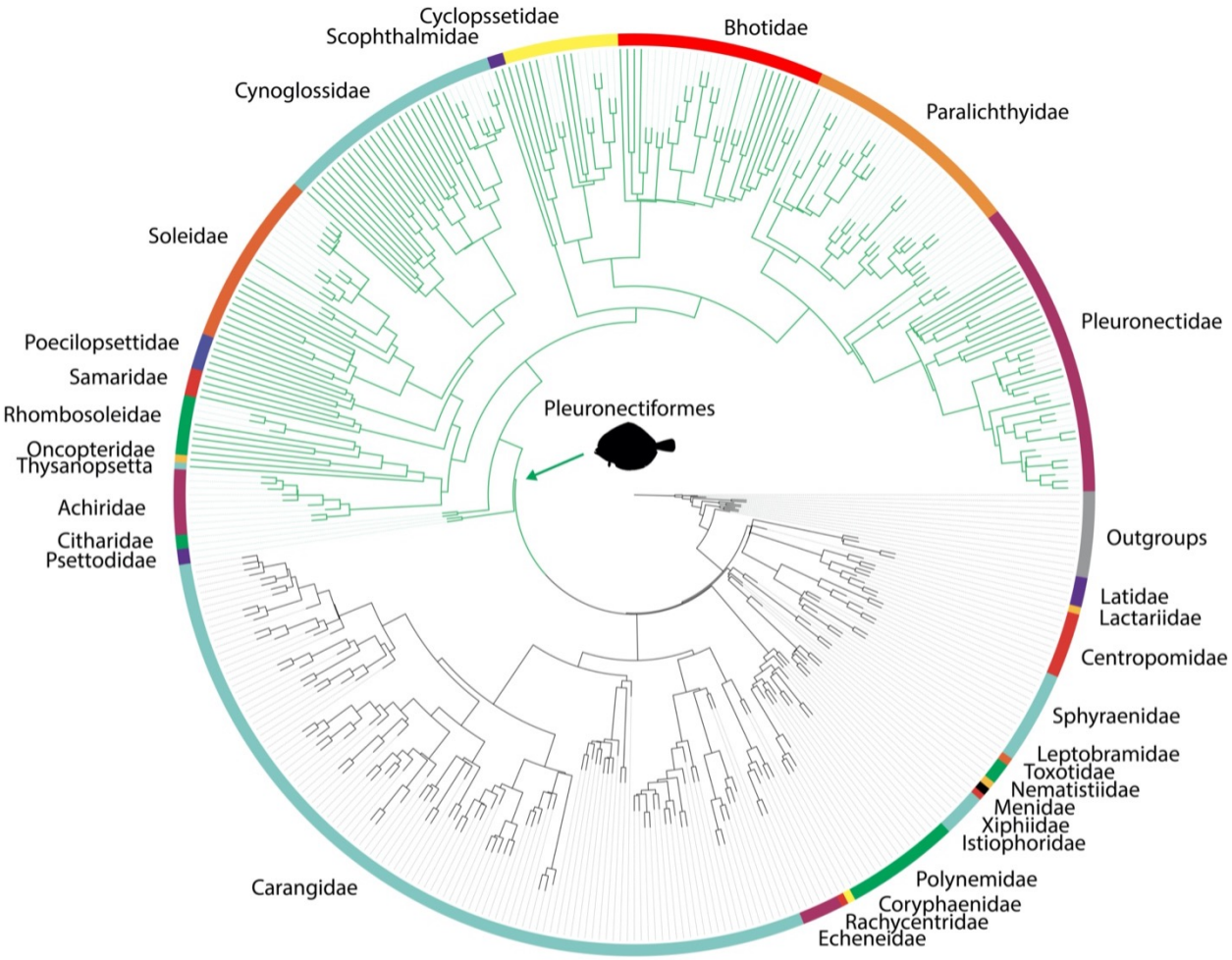

Figure C.1: Species tree for Carangaria estimated with ASTRAL-III using gene trees with weakly supported nodes (bootstrap values [BS] $< 20\%$) collapsed into polytomies to reduce the effects of gene tree error. Gene trees were estimated using IQTree (HM; GTR+G) based on newly sequence data that covers 990 loci and 389 species. Collapsing gene-tree branches with low support helped reduced the negative effects of gene tree estimation error, successfully resolving flatfish monophyly (FM [BS 76%]; green arrow) using the complete dataset. Tree files with tip labels and support values are available from the Figshare digital repository.

Note 3 – Detecting positive selection

To investigate the presence of lineage-specific adaptations under the flatfish monophyly (FM) topology, we tested for positive selection among the 1693 orthologs identified by LEA using the adaptive branch-site random effects likelihood model of codon-substitution aBSREL-Smith et al. (2015) implemented in HyPhy. aBSREL is a recently developed approach that adapts its complexity to the dataset by inferring the optimal number of ω rate categories based on branch lengths. This method also avoids false positive results common to other branch-site models (including PAML, used by LEA) due to rate variation in background lineages. We used two different foreground schemes: (1) stem flatfishes, which aimed to detect genes responsible for the initial break of symmetry in the single branch leading to all extant flatfish species; and (2) crown flatfishes, which aimed to detect genes responsible for further adaptations experienced later in the flatfish radiation (Figure C.2). We identified 67 PSGs in the stem flatfish lineage, and 588 PSGs shared between Psettodoidei and at least one pleuronectoid lineage. Thirty seven of the genes that experienced positive selection in the stem flatfish branch were also identified as PSG in the crown flatfishes including genes recognizably involved in the determination of left/right symmetry during embryonic development (bmp4), sensory perception of sound and otolith morphogenesis (dcdc2b), cell differentiation and proliferation (dlk1, dlk2), eye photoreceptor cell development (dzank1), cardiac ventricle development (fntb), cilium assembly (dcdc2b, dzank1, fhdc2, intu), skeletal muscle tissue development (qkia), proliferation and migration of cranial neural crest cells (tinagl1), and double-strand break repair (nbn). Nevertheless, the involvement of most of these PSGs in the development of the flatfish asymmetric body plan requires further validation. We performed gene ontology (GO) enrichment tests for gene functions and pathways on PSG sets independently for stem and crown flatfishes using R package clusterprofiler Yu et al. (2012). Enrichment tests for gene ontology show no significant overrepresentation in any functional categories. Results for the stem and crown foreground schemes are presented in Tables C.2-3 (see Figshare digital repository).

Note 4 – Evolutionary rates and lineage-specific substitutions

LEA calculated the relative evolutionary rates of each branch in the tree using a two-cluster analysis and Tajima's relative rate test. They estimated the relative evolutionary rates for all the single-copy genes using zebrafish as the outgroup species Lü et al. (2021). The authors state that Psettodoidei presents slightly slower evolutionary rates compared to members of Pleuronectoidei. They used these results to explain why, unlike other flatfishes, species of Psettodes exhibit a 'simply an asymmetric percoid' phenotype. We argue that this result

is contingent on an arbitrary order of species in the Y-axis of their regression plot (LEA fig. 2b). Despite the fact that there is substantial overlap in evolutionary rates between lineages of Pleuronectoidei and Psettodoidei, the authors interpreted this finding as additional evidence of convergence, even though the single-origin hypothesis provides a more parsimonious explanation in this case. LEA also reported several alignments that highlight Pleuronectoidei-specific amino acid substitutions. The outgroup selection is extremely important for these analyses, but surprisingly the authors largely failed to include Psettodoidei or any other non-flatfish carangarian species in their figures. A close inspection of their raw alignments shows that only one of the reported substitutions (wnt9b) is a derived estate exclusive to Pleuronectoidei (Figure C.3). Both missense substitutions in sgca (a gene involved in musculature development), which the authors claimed to be exclusive to Pleuronectoidei, are shared with at least one non-flatfish carangarian, and, therefore, may not explain the thin musculature and development of the flatfish phenotype (Figure C.3). Note that the sole alignment reported in LEA's supplement that does include Psettodes (hoxd12a, a gene involved in dorsal fin development in flatfishes) shows an amino acid substitution shared between Pleuronectoidei and Psettodoidei (Figure C.3). Surprisingly, the authors interpreted this evidence as convergent evolution despite the fact that a single origin provides a more straightforward solution.

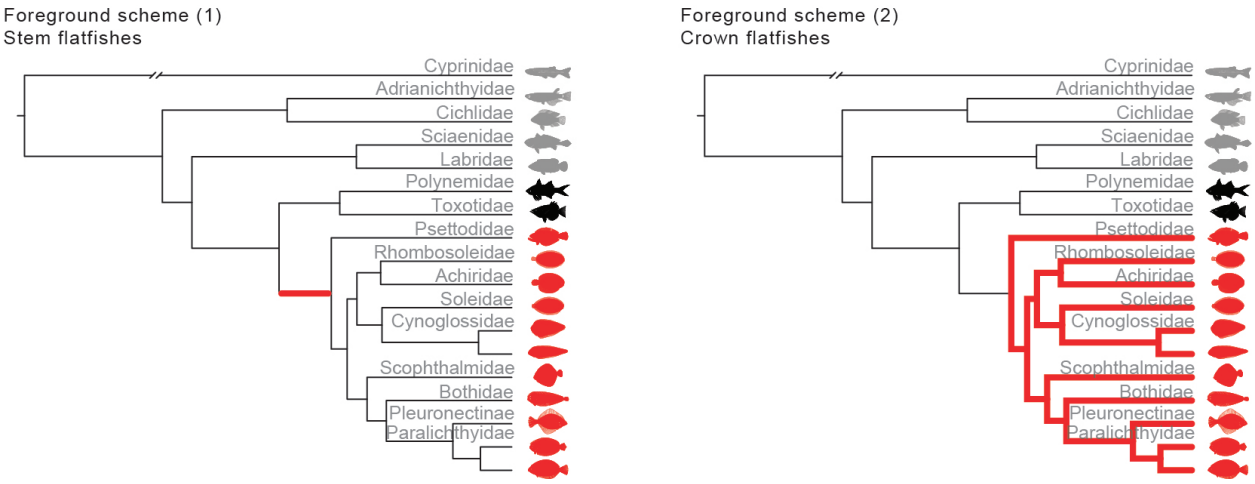

Figure C.2: Different foreground schemes used to detect positively selected genes with aBSREL: (a) stem-flatfishes, which aimed to detect genes responsible for the initial break of symmetry in the single branch leading to all extant flatfish species; and (b) crown flatfishes, which aimed to detect genes responsible for further adaptations experienced later in the flatfish radiation.

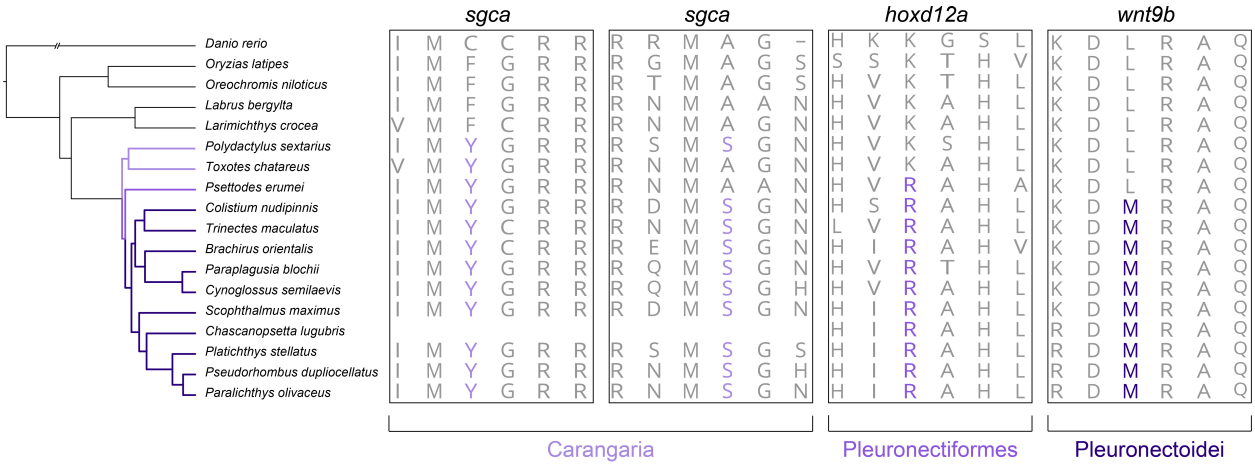

Figure C.3: Amino acid alignments highlighting substitutions (in purple) claimed to be Pleuronectoidei-specific by LEA. By adding sequences from Psettodes and other non-flatfish carangarians into the alignments we show that three out the four Pleuronectoidei-specific substitutions discussed by the authors are in fact shared with Psettodoidei or other non-flatfish carangarian species. These include two missense substitutions in the musculature development related gene sgca, which the authors suggest is related to the formation of the flat phenotype. Both substitutions are shared with Polydactylus sextarius, a symmetrical carangarian species, and, therefore, may not necessarily explain the evolution of the flatfish thin musculature. The hoxd12a substitution is shared between Pleuronectoidei and Psettodoidei. This gene is involved in the development of the flatfish dorsal fin, an important locomotion adaptation that improves their maneuverability in the benthic environment, and that may provide additional evidence for the single origin of the asymmetric body plan.

References

Crotty, S. M., Minh, B. Q., Bean, N. G., Holland, B. R., Tuke, J., Jermiin, L. S., and Haeseler, A. V. (2020). GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. *Systematic Biology*, 69(2):249–264.

Harrington, R. C., Faircloth, B. C., Eytan, R. I., Smith, W. L., Near, T. J., Alfaro, M. E., and Friedman, M. A. (2016). Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. *BMC Evolutionary Biology*, pages 1–14.

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., Betancur, R., Li, C., Becker, L., Bellora, N., Zhao, X., Li, X., Wang, M., Fang, C., Xie, B., Zhoui, Z., Huang, H., Chen, S., Venkatesh, B., and Shi, Q. (2018). Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proceedings of the National Academy of Sciences*, 115(24):201719358.

Lü, Z., Gong, L., Ren, Y., Chen, Y., Wang, Z., Liu, L., Li, H., Chen, X., Li, Z., Luo, H., Jiang, H. H., Zeng, Y., Wang, Y., Wang, K., Zhang, C., Jiang, H. H., Wan, W., Qin, Y., Zhang, J., Zhu, L., Shi, W., He, S., Mao, B., Wang, W., Kong, X., and Li, Y. (2021). Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan. *Nature Genetics*, 53(5):742–751.

Romiguier, J., Ranwez, V., Douzery, E. J., and Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Research*, 20(8):1001–1009.

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5):1342–1353.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, 16(5):284–287.

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*.