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

Melissa Rincon-Sandoval[a,b,1], Emanuell Duarte-Ribeiro[a,1,2], Aaron M. Davis[c], Aintzane Santaquiteria[a], Lily C. Hughes[d,e], Carole C. Baldwin[e], Luisángely Soto-Torres[f], Arturo Acero P.[b], H. J. Walker Jr.[g], Kent E. Carpenter[h], Marcus Sheaves[i], Guillermo Ortí[d,e], Dahiana Arcila[a,j], Ricardo Betancur-R[a,1,2]

[a]Department of Biology, The University of Oklahoma, 730 Van Vleet Oval, Room 314, Norman, OK 73019, USA; [b]Universidad Nacional de Colombia sede Caribe, CECIMAR, Santa Marta, Magdalena, Colombia; [c]Centre for Tropical Water and Aquatic Ecosystem Research (TropWATER), and School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia; [d]Department of Biological Sciences, The George Washington University, Washington, DC 20052; [e]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560; [f]Universidad de Puerto Rico, San Juan, PR, 00931; [g]Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0244 USA; [h]Biological Sciences, Old Dominion University, Norfolk, VA 23529; [i]Marine Data Technology Hub, James Cook University, Townsville, Queensland 4811; [j]Sam Noble Oklahoma Museum of Natural History, Norman, OK, USA.

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 (e.g., sticklebacks, cichlids, whitefishes) recurrently 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 110 (~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 truncate 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 are pervasive among different lineages and across vastly different evolutionary scales, achieving predictable high-fitness solutions for similar environmental challenges, ultimately demonstrating strong determinism in fish body-shape evolution.

phylogenomics | macroevolution | habitat transitions | benthic-pelagic axis | Lutjanidae

Significance

Body shape is a strong predictor of habitat occupation in fishes, which changes 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 streamlined body plans. Recurrent evolution of this pattern across distantly related groups suggests that limited sets of high-fitness solutions exist due to environmental constraints. We provide rigorous tests about these observations showing that similar constraints operate at deeper evolutionary scales in a clade (Lutjanidae) of primarily benthic fish dwellers that repeatedly transitioned into midwater habitats in all major oceans throughout its 45-million-year history. Midwater species strongly converge in body shape, emphasizing evolutionary determinism in form and function along the benthic-pelagic axis.

Author contributions: R.B.R., M.R.S, A.D. and E.D.R. designed the study. H.J.W., K.E.C, C.B., M.S., D.A., R.B.R, and A.A.P. collected and curated tissue samples. D.A., L.C.H., G.O., and C.B. organized the taxonomic sampling and extracted the DNA. L.C.H. and G.O. conducted bioinformatic analyses for data assembly. A.S. conducted ancestral ranges reconstructions; and M.R.S. and A.S. estimated divergence times. E.D.R., M.R.S., and L.S.T. conducted all other analyses. E.D.R., M.R.S., R.B.R. and A.S. made the figures. R.B.R., M.R.S., E.D.R. and G.O. wrote the manuscript; all other authors contributed to the writing. The authors declare no conflict of interest.

Data deposition: Raw sequencing reads are available at NCBI Sequence Read Archive BioProject number (PRJNA630817).

A Figshare digital repository contains alignments, trees and code used for data analysis 10.6084/m9.figshare.13000100.

[1]These authors contributed equally to the work.

[2]To whom correspondence should be addressed. Email: ricardo.betancur@ou.edu, emanuell.ribeiro@ou.edu.

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 (1, 2). 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 (3–5). Nevertheless, the deterministic nature of the processes leading to convergent evolution is still contentious (6, 7).

An emblematic example of evolutionary convergence comes from aquatic environments, 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 (8). 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 (9–13), 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 (14, 15). Quantitative assessments at microevolutionary scales have documented similar cases of resource partitioning on sympatric populations of cichlids and European whitefishes, among others (11, 16–19). 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 (12, 20).

Here, we assess the role of convergent evolution associated with transitions along the benthic-pelagic axis in a clade of tropical and subtropical marine fishes—the snappers and fusiliers in the family Lutjanidae—that bridges both ends of the evolutionary continuum. Previous studies have shown that lutjanids include a number of independent lineages that have undergone niche partitioning along the water column, and that this ecological divergence is seemingly associated with different configurations in feeding ecology and body elongation (21–23). 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. 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 where the family is distributed, providing strong evidence that functional constraints in open water habitats shape phenotypic evolution in fishes in particular, and aquatic vertebrates more generally.

Community ecology studies have demonstrated that marine biodiversity is higher in benthic than pelagic environments (24), 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 (25, 26). This hypothesis makes two predictions: (i) habitat transitions from benthic to midwater systems are expected to be unidirectional or asymmetric, and (ii) the microhabitat homogeneity of pelagic systems provides fewer opportunities for diversification than benthic environments, where multiple niches may cooccur.

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 using independent genomic subsets, 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 oceanic basins. 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.

Results

Phylogenomic inference and tree uncertainty in comparative analysis. Extended results are reported in the *SI Appendix*. Using exon capture approaches (27, 28), we assembled two main phylogenomic data matrices: (a) an expanded supermatrix 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 110 out of *ca.* 136 species; 37% missing cells); and (b) a reduced (phylogenomic-only) matrix obtained with a matrix reduction algorithm, used to assess the sensitivity of phylogenomic results to missing data (1,047 exons and 448,410 nucleotide sites for 84 species; 16% missing cells). 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 (22, 27, 29–32). All analyses invariably resolved seven major clades (Fig. 1; *SI Appendix,* Figs. S2-S5), confirming that the family Lutjanidae, as defined by many studies (e.g., 23, 33), is non-monophyletic with fusiliers (Caesionidae) deeply nested within the broader snapper clade (34). 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 (largely non-overlapping) gene subsets derived from the expanded matrix, each with a sufficient number of genes to overcome sampling error (*SI Appendix,* Dataset S4). 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 (*SI Appendix,* Table S5), generally agreed with those from previous multi-locus studies for the family (27, 29–32, 35; *SI Appendix,* Figs. S12-S14, Table S5, Dataset S5). 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, Ma (~64.2 Ma, 95% HPD: 57.6-68.6 Ma).

The geography of habitat transitions. To assess the geographic prevalence of evolutionary transitions in Lutjanidae, we performed ancestral habitat and ancestral area reconstructions. To infer the history of habitat transitions, we first assigned species into two major habitat categories (benthic and midwater dwellers) and accounted for uncertainty in habitat coding for 13 species using three different probability schemes (see Materials and Methods; *SI Appendix*). Because the implementation of different schemes had an effect on the SIMMAP reconstructions (Fig. 1; *SI Appendix*, Figs. S8-S11), the most likely tip states inferred with each scheme (averaged over the 28 trees in each case) were used for all other downstream analyses that required *a priori* habitat categorization of tips (e.g., trait evolution and convergence, state-dependent diversification). Results of these alternative analyses are reported in combination in the main text and individually in the *SI Appendix*.

To reconstruct ancestral areas, we built a presence-absence matrix of species distribution using alternative biogeographic schemes (36, 37; *SI Appendix,* Dataset S3). Inferences of ancestral ranges using BioGeoBEARS (38) indicate an Indo-west Pacific Ocean origin for lutjanids, with subsequent independent colonization events of the Atlantic (six times) and the eastern Pacific (nine times) via multiple routes (Fig. 1; *SI Appendix,* Figs. S12-S14; see

Supplementary Results for an expanded account on the biogeography). By merging results of ancestral habitat and ancestral range reconstructions, we find support for benthic habitats as the most likely ancestral condition, with independent and recurrent invasions of the water column by benthic lineages at least once within each of the three major oceanic basins (Fig. 1; SI Appendix; Figs. S8-S11). While the Indo-Pacific features more transitions than other basins, our reconstructions highlight the deterministic nature and ubiquity of the transitions (Fig. 1; *SI Appendix*, Fig. S8).

Ecomorphological convergence. To test whether invasions of the water column are associated with a set of convergent high-fitness solutions (e.g., 39, 40), 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 (*SI Appendix*, Fig. S1, Table S7). Traitgram-informed morphospaces (Fig. 2) show that different lutjanid midwater lineages independently evolved slender bodies and furcate caudal fins, an indication of strong ecologically-driven morphological convergence. This pattern is further confirmed based on the threshold model (41), where the full-body shape dataset reveals a substantial correlation between the two habitat states and PC1 (r^2=0.57-0.67), which captured differences in body elongation and caudal fin shape. 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=0.42-0.57) and fins-only (r^2=0.56-0.69) datasets, where only PC1 exhibits significant correlations. We 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=0.07-0.24; for the full-body shape dataset). These results suggest that ecomorphological convergence is less clearly associated with PC2-PC4 axes than it is to the main PC1 axis (*SI Appendix*, Fig. S15).

We used a series of complementary approaches to further assess the scope and strength of convergence. We first compared the relative fit of a set of models of trait evolution in a multivariate framework (mvMORPH [42]), the results of which show split support for the two multi-selective-regime models (BMM and OUM; see Methods), with distinct selective regimes corresponding to the two different habitat categories (Fig. 3a,b; *SI Appendix*, Figs. S16-S18). We then estimated the difference in trait distance between tips in the trees and the maximum distance between those taxa through their evolutionary history (*convevol*, C1-C4 metrics [43]), and

quantified phenotypic similarity based on phylogenetic relatedness (Wheatsheaf index or *w* [44]). 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= 37-45%; *SI Appendix*, Table S8). Likewise, results using the Wheatsheaf index (w= 1.3–1.4; *SI Appendix*, Figs. S19-S20) identified significantly stronger convergence in midwater species than would be expected from a random distribution of trait values simulated under a Brownian Motion (BM) model (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 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= 0.83–0.88; p>0.95), suggesting that morphological diversity is high among benthic dwellers, whereas strong convergent evolution is mostly restricted to midwater lutjanids. Finally, we assessed the optimal number of selective regimes under an Ornstein-Uhlenbeck process without *a priori* designation of habitats (ℓ1ou and SURFACE, [45, 46]). The ℓ1ou (multivariate) and SURFACE (univariate) analyses also identified multiple instances of convergence across lineages with adaptive peaks between clades with similar body plans (deep or slender bodies). For most datasets, the number of non-convergent (adaptive) peak shifts was higher than the number of convergent peaks (*SI Appendix*, Table S9, Dataset S6), and ℓ1ou simulations revealed a significantly greater number of convergent shifts than would be expected by chance (*SI Appendix*, Figs. S21-S23). SURFACE analyses identified a greater number of convergent regimes (*SI Appendix*, Fig. S24) than ℓ1ou for most datasets. Taken together, our results suggest the overall convergence of many lineages to multiple, shared adaptive peaks in body shape ecomorphology (*SI Appendix*, Fig. S25).

Transition rates and diversification in benthic and midwater lineages. We gauged the preference for different habitat states and their effect on rates of habitat transitions (Fig. 3f) and lineage diversification (Fig. 3e), providing a test for the prediction that the adoption of the midwater lifestyle may result in an evolutionary ratchet. For 20 out of the 28 trees, model fitting comparisons supported a state-dependent model (Fig. 3c, d) that incorporates a hidden state (*SI Appendix*, Tables S2-S4) associated with benthic lineages (HiSSE benthic; *SI Appendix*, Fig.

S27a). 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 obtained with HiSSE were consistent with those using the non-parametric FiSSE and parametric BiSSE approaches (*SI Appendix*, Fig. S29, Tables S10-S12), 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 (*SI Appendix,* Tables S10-S12). The HiSSE analyses using a model that accounts for habitat dependent diversification (HiSSE benthic) identified asymmetric transition rates, favoring the expectations that benthic-to-midwater transitions (mean q=0.013) are more frequent than midwater-to-benthic transitions (mean q=0.003; Figs. 1,3f; *SI Appendix,* Fig. S27b).

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 (7, 13, 22, 39, 40, 47–49)—a strong match between form and function (50). This deterministic process is ubiquitous at both temporal and spatial scales, with transitions taking place in lutjanid lineages of different ages and within all major marine biogeographic regions. Within each of the three major oceanic realms, benthic lutjanid lineages invaded the water column at least once. Furthermore, 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*). Snappers 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 (11, 14–19), and more ancient animal lineages such as sharks and aquatic tetrapods (8).

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. Convergent morphologies among midwater species strongly suggest that lineages with independent evolutionary histories but similar habitat preferences are drawn towards 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 (51). Similar outcomes are observed at shallower evolutionary scales in European whitefishes (19) and cichlids in Lakes Apoyo and Xiloá in Nicaragua (18), 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 (9, 12, 13, 51, 52). 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 are often placed in their own family, Caesionidae (53–60). In his revision of lutjanid relationships, Johnson (23) noted that fusiliers feature unique traits among midwater lutjanids, including “an innovative restructuring of the functional complex of the upper jaw (permitting extreme protrusibility for planktivorous feeding) and an alteration of the basic body configuration (providing greater and more rapid swimming ability).” Remarkably, some adaptive landscape analyses that detected a single adaptive shift in Lutjanidae (*SI Appendix*, Figs, S14-S17), 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 fishes. A prime example includes the midwater Boga in the Caribbean, formerly listed as *Inermia vittata* in the family Emmelichthydae, but recently shown to be a derived grunt (Haemulidae; (61). A more extreme case comprises the picarels, previously placed in Centracanthidae, a family that is polyphyletically nested within benthic porgies in the family

Sparidae (62). 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 Creole-fish (formerly *Paranthias colonus*, now *C. colonus*; 63). 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 (64–66), 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 (12). 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 (Fig. 1; *SI Appendix*, Figs, S27-S29), which are roughly twice as fast compared to the midwater counterparts (Fig. 3 e, 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 (67). State-dependent diversification analyses provide some support for these ideas, identifying remarkably faster rates of extinction in midwater than benthic lineages (*SI Appendix*, Tables S10-12).

Snappers and fusiliers exhibit strong but imperfect morphological convergence (68) 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 (69), "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; 70). Outside Lutjanidae, remarkable departures from typical streamlined body shapes found in most oceanic pelagic vertebrates include the slow-swimming ocean sunfishes, butterfishes, moonfish, opah, and tripletails, which feature deep and laterally-compressed body plans. Although we did not examine diets and feeding morphology in this study, a key factor that triggers the invasion of the water column is the trophic adaptation to planktivory. Morphological convergence has been reported in many groups that share specialized dietary shifts to planktivory (e.g., butterflyfishes, wrasses, angelfishes, damselfishes, and sea basses;47, 65, 71, 72). Ecological opportunity for the exploitation of different resources has thus repeatedly promoted morphological and behavioral adaptations associated with water-column transitions (65, 73).

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 convergent 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.

Materials and Methods

Taxonomic sampling and genomic data. Extended Materials and Methods are reported in the *SI Appendix*. 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 25 additional ingroup species from GenBank. Our combined dataset contains 110 species plus 14 outgroups (*SI Appendix*, Dataset S1). High quality DNA extractions were sent to Arbor Biosciences for target enrichment and sequencing. Our target capture probes are based on a set of 1,104 single-copy exons optimized for ray-finned fish phylogenetics (27, 28). 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 (74). For each matrix, we determined the best-fitting partitioning schemes and nucleotide substitution models for both genes and codon positions using PartitionFinder2 (75). 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 (76) using the partition output obtained with PartitionFinder2. Species trees were then inferred with ASTRAL-II v4.7.12 (77) using individual RAxML-based gene trees as input.

Accounting for topological and temporal uncertainty. We built 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. 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 (78), which can handle genome-scale datasets in a Bayesian framework (79). Because MCMCTree running time depends more on the number of partitions defined rather than the number of genes included (79), 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 (*SI Appendix*, Table S1).

Reconstruction of ancestral habitats and ancestral ranges. The habitat occupancy dataset (*SI Appendix*, Dataset S2) was compiled by aggregating information from a wide range of sources, including FishBase, the primary literature, and by consulting experts. The reconstructions performed used a broad sampling of 97 haemulid outgroups (13). To account for 13 lutjanid species with uncertain habitat occupancy, we implemented ancestral character reconstructions that take into account tip-state ambiguity based on stochastic character mapping (SIMMAP [80]), as implemented in the R package phytools (81). We coded these ambiguous tips using three alternative probability schemes: 0.1 benthic/0.9 midwater, 0.50 benthic/0.50 midwater, 0.9 benthic/0.1 midwater (*SI Appendix*).

We also classified species according to their geographical ranges. We built a presence/absence matrix of species considering six recognized marine biogeographic regions (36, 37; *SI Appendix*, Dataset S3): 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 (38). 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 *SI Appendix* 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 (Fig. 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 110 species (total 413 individuals; mean 3.72 individuals per species; *SI Appendix*, Dataset S1). We generated three alternative datasets (following 82) based on digitized landmarks: (i) A full-body and fin shape dataset; (ii) a body-only dataset; and (iii) a fins-only dataset (see SI Materials and Methods, Fig. S1). For each dataset, we performed Procrustes superimposition, calculated species-average coordinates, and conducted both standard (PCA) and phylogenetically-corrected (pPCA) principal component analyses (83, 84). Finally, we determined the number of meaningful PC axes using the broken-stick model (85, 86), 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 (42). 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 (87). 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 (44). Finally, we used other data-driven approaches, as implemented in the R package ℓ1ou v1.42 (45) and SURFACE v0.4 (46), 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 (88). We applied HiSSE (Hidden State Speciation and Extinction), an SSE approach that tests the relative fit of a set of alternative branching models while accounting for hidden states. For comparison, and to estimate habitat-dependent evolutionary rates in a Bayesian framework, we also used BiSSE as implemented in the R package diversitree (89). Finally, we used the nonparametric FiSSE approach, which has shown to be robust to phylogenetic pseudo-replication and model misspecification (90). See *SI Appendix* for details.

Acknowledgements

We are thankful for the extensive and insightful comments by D. Johnson and two anonymous reviewers that helped improve the quality of our study. We thank O. Domínguez, Universidad Michoacana de San Nicolás de Hidalgo (UMSNH) for providing samples. M. Stimson and M. Bagger (GWU) conducted DNA extractions. E. Santaquiteria provided artistic illustrations. Bioinformatic analyses were facilitated by the High-Performance Computing facility of UPR-RP (funded by INBRE Grant P20GM103475) and the OU Supercomputing Center for Education & Research (OSCER). This research was supported by National Science Foundation (NSF) grants DEB-1932759 and DEB-1929248 to RBR, DEB-1457426 and DEB-1541554 to GO, DEB-1541552 to CCB, and DEB-2015404 to DA. M.R.S. was supported by a postdoctoral fellowship from Colciencias (Grant 848-2019).

Legends for figures

Figure 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. 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 multi-state habitats; see Dataset S2). Color gradients along branches denote habitat transitions; purple branches indicate lineages with ambiguous habitats based on reconstructions using alternative coding schemes (see also Figs. S8–S11). Colored circles indicate colonization events (inferred with BioGeoBEARS; see also Figs. S8, S12–S14) of the Atlantic (yellow circles) and the tropical eastern Pacific (purple circles), from Indo-Pacific lineages (center of origin; green circle). Arrows in maps depict reconstructed colonization routes by different lineages in three time slices: 50–12 Ma (mean 31 Ma), before the closure of Tethys Seaway; 12–2.8 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. Thickness of 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. Arrows in the central panel show the transitions rates between benthic and pelagic habitats, as estimated with HiSSE (see also Tables S10-S12).

Figure 2. Traitgram-informed morphospaces for lutjanids illustrating ecomorphological partitioning and convergence across benthic and midwater lineages, as estimated using the full-body dataset. Contour lines represent the two-dimensional density distributions of the species presenting each of the two habitat states. Traitgrams overlain along PC axes depict the phylogeny in Fig. 1, including ancestral habitat reconstructions estimated with SIMMAP (a, PC1 vs. PC2; b, PC3 vs. PC4). Color gradients along branches denote habitat transitions; purple branches and data points indicate lineages with ambiguous habitats based on alternative coding schemes. Branches shifting from red to blue along PC1 extremes highlight convergent evolution in midwater lineages. Parenthetical values indicate the total variance explained by each PC axis.

Figure 3. Model-fitting comparisons and lineage diversification parameters estimated by accounting for phylogenetic uncertainty (28 trees) and habitat coding ambiguity (13 tips with uncertain or multi-state habitats). (*a*, *b*), comparisons of alternative models of morphological evolution using the full body dataset: (*a*) distribution of the AIC values for the three alternative models of continuous trait evolution (BM, OU, BMM, and OUM), and (*b*) AICw of each alternative model and tree. (*c*, *d*) Comparisons for alternative models of lineage diversification: (*c*) distribution of AIC values for seven alternative SSE models (Tables S2-S4); (*d*) AIC weights (AICw) for each SSE model based on each of the 28 trees. (*e*,*f*) Estimated lineage diversification parameters: (*e*) net-diversification values for the three habitat states, and (*f*) transition rates (Q) between benthic and midwater states.

1. S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History*, W. W. Norton & Company, Ed. (1989).
2. J. B. Losos, Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards. *Science (80-.).* **279**, 2115–2118 (1998).
3. A. A. Agrawal, Toward a Predictive Framework for Convergent Evolution: Integrating Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life. *Am. Nat.* **190**, S1–S12 (2017).
4. J. B. Losos, Convergence, adaptation, and constraint. *Evolution (N. Y).* **65**, 1827–1840 (2011).
5. D. B. Wake, M. H. Wake, C. D. Specht, Homoplasy: From Detecting Pattern to Determining Process and Mechanism of Evolution. *Science (80-.).* **331**, 1032–1035 (2011).
6. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution:

Replaying life's tape. *Science (80-.).* **362** (2018).

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

8. G. R. McGhee, *Convergent Evolution: Limited Forms Most Beautiful* (The MIT Press, 2011).

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

10. T. Claverie, P. C. Wainwright, A Morphospace for Reef Fishes: Elongation Is the Dominant Axis of Body Shape Evolution. *PLoS One* **9**, 1–11 (2014).

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

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

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

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

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

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

17. W. J. Cooper, *et al.*, Bentho-pelagic divergence of cichlid feeding architecture was prodigious and consistent during multiple adaptive radiations within African Rift-Lakes. *PLoS One* **5** (2010).

18. K. R. Elmer, *et al.*, Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat. Commun.* **5** (2014).

19. K. Præbel, *et al.*, Ecological speciation in postglacial European whitefish: rapid adaptive

radiations into the littoral, pelagic, and profundal lake habitats. *Ecol. Evol.* **3**, 4970–4986 (2013).

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

21. W. P. Davis, R. S. Birdsong, Coral reef fishes which forage in the water column. *Helgoländer Wissenschaftliche Meeresuntersuchungen* **24**, 292–306 (1973).

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

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

24. J. S. Gray, Marine biodiversity: Patterns, threats and conservation needs. *Biodivers. Conserv.* **6**, 153–175 (1997).

25. M. A. Balisi, B. Van Valkenburgh, Iterative evolution of large-bodied hypercarnivory in canids benefits species but not clades. *Commun. Biol.* **3**, 1–9 (2020).

26. M. Juhn, B. Van Valkenburgh, M. Alfaro, "Exploring Macroevolutionary Ratchets as a potential driver of clade decline" (2017) https:/doi.org/10.1130/abs/2018AM-321109.

27. L. C. Hughes, *et al.*, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci.* **115**, 6249–6254 (2018).

28. L. C. Hughes, *et al.*, Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Mol. Ecol. Resour.* (2020) https:/doi.org/10.1101/2020.02.18.949735.

29. R. Betancur-R, *et al.*, The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr. Tree Life* **1** (2013).

30. R. R. Betancur, *et al.*, Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17** (2017).

31. D. L. Rabosky, *et al.*, An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).

32. T. J. Near, *et al.*, Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12738–12743 (2013).

33. J. Nelson, T. Grande, M. Wilson, *Fishes of the World*, John Wiley & Sons, Ed., 5th Ed. (2016).

34. D. Johnson, Percomorph Phylogeny - Progress and Problems. *Bull. Mar. Sci.* **52**, 3–28 (1993).
35. M. E. Alfaro, Resolving the ray-finned fish tree of life. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6107–6109 (2018).
36. M. Kulbicki, *et al.*, Global Biogeography of Reef Fishes: A Hierarchical Quantitative Delineation of Regions. *PLoS One* **8** (2013).
37. M. D. Spalding, *et al.*, Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *Bioscience* **57**, 573–583 (2007).
38. N. J. Matzke, BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. *R Packag. version 0.2* (2013).
39. K. L. Feilich, G. V. Lauder, Passive mechanical models of fish caudal fins: Effects of shape and stiffness on self-propulsion. *Bioinspiration and Biomimetics* **10**, 036002 (2015).
40. P. W. Webb, Body form, locomotion and foraging in aquatic vertebrates. *Am. Zool.* **24**, 107–120 (1984).
41. J. Felsenstein, Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1427–1434 (2005).
42. J. Clavel, G. Escarguel, G. Merceron, mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015).
43. C. T. Stayton, The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution (N. Y).* **69**, 2140–2153 (2015).
44. K. Arbuckle, C. M. Bennett, M. P. Speed, A simple measure of the strength of convergent evolution. *Methods Ecol. Evol.* **5**, 685–693 (2014).
45. M. Khabbazian, R. Kriebel, K. Rohe, C. Ané, Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* **7**, 811–824 (2016).
46. T. Ingram, D. L. Mahler, SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol. Evol.* **4**, 416–425 (2013).
47. S. T. Friedman, S. A. Price, A. S. Hoey, P. C. Wainwright, Ecomorphological convergence in planktivorous surgeonfishes. *Journal of Evolutionary Biology*. **29**, 965–978 (2016).

48. R. B. Langerhans, D. N. Reznick, “Ecology and Evolution of Swimming Performance in Fishes: Predicting Evolution with Biomechanics” in *Fish Locomotion: An Eco Ethological Perspective*, P. Domenici, B. G. Kapoor, Eds. (Enfield (New Hampshire): Science Publishers, 2010).
49. J. P. Velotta, D. McCormick, A. W. Jones, Reduced Swimming Performance Repeatedly Evolves on Loss of Migration in Landlocked Populations of Alewife. *Physiol. Biochem. Zool.* **91**, 814–825 (2018).
50. A. L. Pigot, *et al.*, Macroevolutionary convergence connects morphological form to ecological function in birds. *Nat. Ecol. Evol.* **4**, 230–239 (2020).
51. S. T. Friedman, *et al.*, Body shape diversification along the benthic-pelagic axis in marine fishes. *Proceedings. Biol. Sci.* **287** (2020).
52. S. A. Price, *et al.*, Building a Body Shape Morphospace of Teleostean Fishes. *Integr. Comp. Biol.* **59**, 716–730 (2019).
53. K. E. Carpenter, *Revision of the Indo-Pacific Fish Family Caesionidae (Lutjanoidea), with Descriptions of Five New Species* (Bishop Museum Press, 1987).
54. K. E. Carpenter, “Volume 5: Bony fishes part 3 (Menidae to Pomacentridae)” in *The Living Marine Resources of the Western Central Pacific*, FAO, (2001), pp. 2919–2941.
55. K. E. Carpenter, “Vol.8. Fusiliers fishes of the world” in *FAO Species Catalogue*, (1988).
56. K. E. Carpenter, A Phylogenetic Analysis of the Caesionidae (Perciformes: Lutjanoidea). *Copeia*, 692–717 (1990).
57. K. E. Carpenter, Optimal cladistic and quantitative evolutionary classifications as illustrated by fusilier fishes (Teleostei: Caesionidae). *Syst Biol* **42**, 142–154 (1993).
58. R. Fricke, W. N. Eschmeyer, R. van der Laan, Eschmeyer’s Catalog of Fishes: Genera, Species, References (2020) (September 20, 2019).
59. R. Froese, D. Pauly, FishBase (2019).
60. T. L. Miller, T. H. Cribb, Phylogenetic relationships of some common Indo-Pacific snappers (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. *Mol. Phylogenet. Evol.* **44**, 450–460 (2007).
61. J. J. Tavera, A. Acero P, E. F. Balart, G. Bernardi, Molecular phylogeny of grunts (Teleostei, Haemulidae), with an emphasis on the ecology, evolution, and speciation

history of New World species. *BMC Evol. Biol.* **12**, 57 (2012).
62. M. D. Sanciangco, K. E. Carpenter, R. Betancur-R, Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). *Mol. Phylogenet. Evol.* **94**, 565–576 (2016).
63. M. T. Craig, P. A. Hastings, A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. *Ichthyol. Res.* **54**, 1–17 (2007).
64. D. R. Bellwood, L. Van Herwerden, N. Konow, Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). *Mol. Phylogenet. Evol.* **33**, 140–155 (2004).
65. S. R. Floeter, M. G. Bender, A. C. Siqueira, P. F. Cowman, Phylogenetic perspectives on reef fish functional traits. *Biol. Rev.* **93**, 131–151 (2018).
66. F. L. Lobato, *et al.*, Diet and diversification in the evolution of coral reef fishes. *PLoS One* **9**, 1–11 (2014).
67. B. Van Valkenburgh, X. Wang, J. Damuth, Cope's rule, hypercarnivory, and extinction in North American canids. *Science (80-.).* **306**, 101–104 (2004).
68. D. C. Collar, J. S. Reece, M. E. Alfaro, P. C. Wainwright, R. S. Mehta, Imperfect morphological convergence: Variable changes in cranial structures underlie transitions to durophagy in moray eels. *Am. Nat.* **183**, 168–184 (2014).
69. E. S. Hobson, Feeding relationship of teleostean fishes on coral reefs in Kona, Hawaii. *Fish. Bull.* **72**, 915–1031 (1974).
70. M. E. Alfaro, D. I. Bolnick, P. C. Wainwright, Evolutionary Consequences of Many-to-One Mapping of Jaw Morphology to Mechanics in Labrid Fishes. *Am. Nat.* **165** (2005).
71. B. Frédérich, L. Sorenson, F. Santini, S. Graham J., M. E. Alfaro, Iterative Ecological Radiation and Convergence during the Evolutionary History of Damselfishes (Pomacentridae). *Am. Nat.* **181** (2012).
72. P. F. Cowman, D. R. Bellwood, L. Van Herwerden, Dating the evolutionary origins of wrasse lineages (Labridae) and the rise of trophic novelty on coral reefs. *Mol. Phylogenet. Evol.* **52**, 621–631 (2009).
73. W. J. Cooper, C. B. Carter, A. J. Conith, A. N. Rice, M. W. Westneat, The evolution of jaw protrusion mechanics is tightly coupled to bentho-pelagic divergence in damselfishes (Pomacentridae). *J. Exp. Biol.* **220**, 652–666 (2017).

74. B. Meyer, K. Meusemann, B. Misof, MARE v0.1.2-rc (2011).
75. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
76. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
77. S. Mirarab, T. Warnow, ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, 44–52 (2015).
78. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
79. M. dos Reis, Z. Yang, "Bayesian Molecular Clock Dating Using Genome-Scale Datasets" in *Evolutionary Genomics: Statistical and Computational Methods*, Second Edi, (Springer Protocols, 2019).
80. J. P. Bollback, SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7** (2006).
81. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
82. A. B. George, M. W. Westneat, Functional morphology of endurance swimming performance and gait transition strategies in balistoid fishes. *J. Exp. Biol.* **222** (2019).
83. D. Adams, M. Collyer, A. Kaliontzopoulou, Geomorph: Software for geometric morphometric analyses. R package version 3.1.0 (2019).
84. L. J. Revell, Size-correction and principal components for interspecific comparative studies. *Evolution (N. Y).* **63**, 3258–3268 (2009).
85. D. A. Jackson, No Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecol. Ecol. Soc. Am.* **74**, 2204–2214 (1993).
86. P. R. Peres-Neto, D. A. Jackson, K. M. Somers, How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput. Stat. Data Anal.* **49**, 974–997 (2005).
87. J. Felsenstein, A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. *Am. Nat.* **179**, 154–56 (2012).
88. W. P. Maddison, P. E. Midford, S. P. Otto, Estimating a binary character's effect on

speciation and extinction. *Syst. Biol.* **56**, 701–710 (2007).

89. R. G. Fitzjohn, Diversitree: Comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**, 1084–1092 (2012).

90. D. L. Rabosky, E. E. Goldberg, FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution (N. Y).* **71**, 1432–1442 (2017).

Supplementary Information for

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

Melissa Rincon-Sandoval[a,b,1], Emanuell Duarte-Ribeiro[a,1,2], Aaron M. Davis[c], Aintzane Santaquiteria[a], Lily C. Hughes[d,e], Carole C. Baldwin[e], Luisángely Soto-Torres[f], Arturo Acero P.[b], H. J. Walker Jr.[g], Kent E. Carpenter[h], Marcus Sheaves[i], Guillermo Ortí[d,e], Dahiana Arcila[a,j], Ricardo Betancur-R[a,1,2]

[a]Department of Biology, The University of Oklahoma, 730 Van Vleet Oval, Room 314, Norman, OK 73019, USA; [b]Universidad Nacional de Colombia sede Caribe, CECIMAR, Santa Marta, Magdalena, Colombia; [c]Centre for Tropical Water and Aquatic Ecosystem Research (TropWATER), and School of Marine and Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia; [d]Department of Biological Sciences, The George Washington University, Washington, DC 20052; [e]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560; [f]Universidad de Puerto Rico, San Juan, PR, 00931; [g]Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0244 USA; [h]Biological Sciences, Old Dominion University, Norfolk, VA 23529; [i]Marine Data Technology Hub, James Cook University, Townsville, Queensland 4811; [j]Sam Noble Oklahoma Museum of Natural History, Norman, OK, USA.

[1]These authors contributed equally to the work.

[2]To whom correspondence should be addressed. Email: ricardo.betancur@ou.edu, emanuell.ribeiro@ou.edu.

This PDF file includes:

Supplementary Materials and Methods

Supplementary Results

Figures S1 to S29

Tables S1 to S12

Appendix SI

Legends for Datasets S1 to S6

SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S6

Supplementary Materials 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*. (1), 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 (2, 3), 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 (4). Several exons that were not included in Hughes *et al*. (2), 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 (3, 5–7). 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 (3). 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 (8). Reads were mapped against reference sequences used in probe design with BWA-MEM (9) and potential PCR duplicates were removed with Samtools v1.9 (10). Mapped reads were extracted for each locus, and an initial contig for each exon was assembled with Velvet v1.2.10 (11). The longest contig assembled by Velvet for each locus

was then used as a reference for aTRAM 2.0 (12) 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 (13). The coding2genome algorithm in Exonerate (14) was used to find reading frames by aligning it to a percomorph reference sequence that was previously verified by visual inspection (2) 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 (15), with Mafft v7.421 (16) 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 (3).

Phylogenomic analyses of exon markers

We combined genomic data from 85 newly sequenced species with sequences for 25 additional species acquired from GenBank. Including 14 haemulid outgroups, and before eliminating duplicate 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 (17), 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) including haemulid outgroups. 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 110 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 (18, 19) 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 (20).

Phylogenetic dating

The complete matrices and subsets were run in MCMCTree using the approximate likelihood method under the HKY85 model (21). 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 (22).

Fossil calibrations

Based on recommendations by Parham *et al*. (23), 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* (24), †*Ottaviania leptacanthus* (25), †*Veranichthys ventralis* (25), †*Goujetia crassispina* (25), †*Lessinia horrenda* (26), and †*Lessinia* sp. (27, 28). 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 (29). Therefore, the calibration is placed as stem Lutjanidae (one node below). Stratigraphic horizon and locality: early Eocene, upper Ypresian, Monte Bolca, Italy (26). Absolute age estimate: 48.5 Ma (30). 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, 5, 31–35). 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 (see Table S1 for details).

(2) Crown Lutjanidae. MRCA: *Etelis oculatus, Lutjanus lutjanus*. Hard lower bound: *†Hypsocephalus atlanticus* (36). Diagnosis and phylogenetic placement: this fossil was first described in Hoplopagrini (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 premaxillarae. 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* (37). 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, maxillae, and premaxillae 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 (33) used the fossil to calibrate a more nested clade within crown lutjanids (i.e., the "lutjanines" + "caesionines" clade); however, no morphological evidence was provided to support this decision. Stratigraphic horizon and locality: late Eocene, Operculinoides-Asterocyclina Zone in the Crystal River formation in north Florida area (38). Absolute age estimate: 33.9 Ma (36). 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 (39), 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., 40), recent studies have challenged the timing of the final closure of the Isthmus of Panama (40). More specifically, Montes *et al*. (41) 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*. (42), 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-L. campechanus*. MRCA: *Lutjanus peru, Lutjanus campechanus.*

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

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

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

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

Table S1. Priors used for divergence time estimations in MCMCTree.

MRCA	Age (Ma)	Distribution	Calibration type	Parameters
Lutjanus lutjanus, *Pomadasys empherus*	48.5-66	Uniform	Soft upper and hard lower bounds	B(0.485,0.66,1e-300,0.05)
Lutjanus lutjanus, *Lutjanus sebae*	33.9-48.5	Uniform	Soft upper and hard lower bounds	B(0.339,0.485,1e-300,0.05)
Lutjanus peru, *Lutjanus campechanus/purpureus*	2.8 (min.)	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Lutjanus inermis, *Ocyurus chrysurus*	2.8 (min.)	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Lutjanus argentiventris, *Lutjanus alexandrei*	2.8 (min.)	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Lutjanus synagris, *Lutjanus guttatus*	2.8 (min.)	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)
Lutjanus cyanopterus, *Lutjanus novemfasciatus*	2.8 (min.)	Cauchy	Hard lower bound	L(0.028,0.1,1,1e-300)

Habitat reconstructions

We estimated ancestral habitats using stochastic character mapping (SIMMAP; 43) under joint reconstructions, as implemented in the R package phytools (44). Because state reconstructions can be influenced by the selection of outgroups, we replaced our sampling of 14 haemulid outgroups with a much broader sampling of 97 species for that family. For this analysis, we did not include any outgroups outside Lutjaniformes (Lutjanidae + Haemulidae) as different large-scale phylogenetic studies of fishes have produced incongruent results (2, 32, 35), failing to resolve the interrelationships of lutjaniforms among other families in Eupercaria. The haemulid time tree and associated midwater/benthic coding for all species are based on Tavera et al. (45). We bound the haemulid tree to each of the 28 Lutjanidae-only

trees (after pruning the 14 haemulid outgroups from our sampling), keeping the ultrametricity of the trees based on our estimated ages for both crown and total groups.

We identified 13 lutjanid species with uncertain habitat occupancy that either lack sufficient information or that are truly multi-state taxa (Dataset S2). These include, for example, deep-sea apsilines and etelines. To account for these uncertainties, we conducted SIMMAP reconstructions that allow the implementation of tip-state probabilities using phytools. We coded these ambiguous tips using three alternative probability schemes: 0.1 benthic/0.9 midwater, 0.50 benthic/0.50 midwater, 0.9 benthic/0.1 midwater. Because the use of different probability schemes had an important effect on the SIMMAP reconstructions (Figs. S8-S11; see Supplementary Results), the most likely tip states inferred with these alternative schemes (averaged over the 28 trees in each case) were used for all other downstream analyses that required *a priori* habitat categorization of tips (e.g., trait evolution and convergence, state dependent diversification). Given the variety of trees and coding schemes, we used the more general 'all rates different' or ARD model for the 84 SIMMAP reconstructions conducted (3 coding schemes for each of the 28 trees). Finally, for each coding scheme, we estimated the number of transitions between benthic and midwater habitats for lutjanids after pruning the haemulid outgroups from all SIMMAP trees.

Ancestral range reconstructions

We classified species according to their geographical ranges. We built a presence/absence matrix of species considering six recognized marine biogeographic regions (46, 47; Dataset S3): West-Indian Ocean (WIO), Central Indo-Pacific (CIP), Central Pacific (CP), Tropical Eastern Pacific (TEP), Western Atlantic (WA), and Eastern Atlantic (EA). For simplicity, we also 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 (Fig. 1). Note that all lutjanid species in our dataset are currently distributed in a single major basin, except for *Aphareus furca* which occurs in both the Indo-Pacific and TEP. We used the R package BioGeoBEARS (48), 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; 49), DIVA (dispersal-vicariance analyses; 50), and BayArea (Bayesian Inference of Historical Biogeography for Discrete Areas; 51), 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 (52). The *w* parameter is used to infer the optimal dispersal multiplier matrix, which acts as an exponent on that matrix using maximum likelihood (53). 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 (54). 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 (55). To account for the final closure of the Panama Isthmus, which may have occurred as early as 2.8 Ma as stated above (42), 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 (56, 57). 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 (Fig. 1). All BioGeoBEARS analyses (with three and six areas) used the 'master tree' inferred with RAxML as input.

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. While some degree of 'fusiformity' cannot be captured with 2D images, the ease of implementing 2D geometric morphometric approaches based on available photographs is cost effective given the scope of this study. Other approaches for generating of 3D images require CT-scan or multi-camera settings (e.g., http://copis.tubri.org), which are costly and far less accessible. We generated three alternative datasets (following 58) based on digitized landmarks that were chosen to capture the disparity of body plans in Lutjanidae while ensuring homology by avoiding distortions arising from specimen preservation: (i) A

full-body shape dataset that comprises a set of 18 functionally homologous landmarks (Fig. S1), 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. Note that while the kinematics and force production of the pectoral fins vary substantially in active vs. more passive fish swimmers (e.g., 59) we did not examine pectoral fins in this study due to technical limitations in capturing their shape from photos in a two-dimensional plane. 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 (60). 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) (61). Finally, we determined the number of meaningful PC axes using the broken-stick model (62, 63), which minimizes loss of signal while avoiding noise from less relevant axes.

Figure S1. 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). Fixed landmarks that are shared between fins-only and body-only datasets: 1–5, 9, 10.

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 (64). 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 (65).

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 (66). We used a Bayesian MCMC function (threshBayes) as implemented in the R package phytools (44). 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 (67). 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 (68). 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 (69) 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) (65). To complement the ℓ1ou analyses, we also used the SURFACE method (70) for data-driven identification of clades featuring convergent evolution.

State-dependent diversification

We implemented hidden state speciation and extinction analyses using the R package HiSSE. We first 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 (Table S2-S4). For comparison, and to estimate habitat-dependent evolutionary rates in a Bayesian framework, we also used the BiSSE (binary state speciation and extinction) approach implemented in the R package diversitree (71). Finally, 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, 72), we also applied the nonparametric FiSSE approach, which has shown to be robust to phylogenetic pseudoreplication and model misspecification (73). 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.

Other analyses

Methodological details for other analyses conducted are reported in the main text.

Supplementary Results

Phylogenomic inference, divergence times and habitat reconstructions. The reduced matrix assembled using the MARE approach (17) comprises 1,047 exons and a total 448,410 DNA sites for 84 species (16% missing cells). The complete concatenated dataset contains 1,115 exons with an expanded data matrix consisting of 474,132 DNA sites for 110 species (37% missing cells).

In agreement with results from previous studies (2, 5, 31–35), the family Lutjanidae (to the exclusion of Caesionidae) was deemed non-monophyletic based on both concatenated and coalescent-based analyses (Fig. S2-S5). 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 (33, 74), with some notable exceptions explained below. All analyses invariably resolved seven major lutjanid clades (Fig. S2-S5): 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*, previously classified as sparoids (29, 75–77). Recognition of these three subfamilies follows Johnson and Carpenter (29, 75, 76, 78). Next, Clade A (iv) and Clade B (v), as defined by Frédérich & Santini's (33), 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 *niger,* and the fusiliers (formerly Caesionidae (29, 79); genera *Pterocaesio*, *Caesio*, *Gymnocaesio,* and *Dipterygonotus*). We identified a substantially different placement for *Lutjanus bohar* in Frédérich & Santini's tree (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 & 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 [80]), 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 (79). 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 (33, 74) call for a revised taxonomy of genera and subfamilies in Lutjanidae.

The relationships estimated with the expanded matrix, in which 110 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 (33, 74).

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; Dataset S4), 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; Fig. S6). 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 Dataset S5 and Fig. S7. Divergence-time estimates are reasonably in good agreement compared to the age of the lutjanid stem, as estimated by multi-locus analyses (2, 5, 32–34; see Table S5 for a comparison). Studies that did not include internal calibrations for lutjanids placed the origin of the crown group in the early Eocene (32, 33, 35). 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).

SIMMAP analyses based on different coding schemes for uncertain tips had an important impact on the ancestral habitat reconstructions (Fig. S8). Differences obtained were most striking in the Apsilinae and Etelinae clades, which together had 10 (out of 13) species with uncertain or ambiguous habitat affiliations. Our results for those clades are rather similar between the “0.1benthic/0.9 midwater” and “0.5 benthic/0.5 midwater” probability schemes, where tips depict a tendency towards midwater habitat occupancy. However, analyses based on the “0.9 benthic/0.1 midwater” probability scheme suggests ancestral benthic habitats at many nodes in the Apsilinae and Etelinae clades (Fig. S8).

For each of the three alternative probability schemes, we looked into the 28 Lutjanidae-only trees to more thoroughly analyze habitat occupancy patterns or discrepancies between trees inferred using RAxML and ASTRAL-II based on either subsets or full datasets. For the “0.5 benthic/0.5 midwater” scheme (Fig. S9), tip probabilities varied considerably between benthic and midwater habitat occupancy. Yet, the results were largely consistent between subsets and expanded trees, with a general tendency towards midwater habitat occupancy. For the “0.1 benthic/0.9 midwater” scheme (Fig. S10), we found a strong consistency regarding midwater habitat occupancy probabilities in all trees. As with the

master tree (Fig. S8), reconstructions based on the "0.9 benthic / 0.1 midwater" scheme (Fig. S11) identified many ancestral benthic nodes in apsilines and etelines.

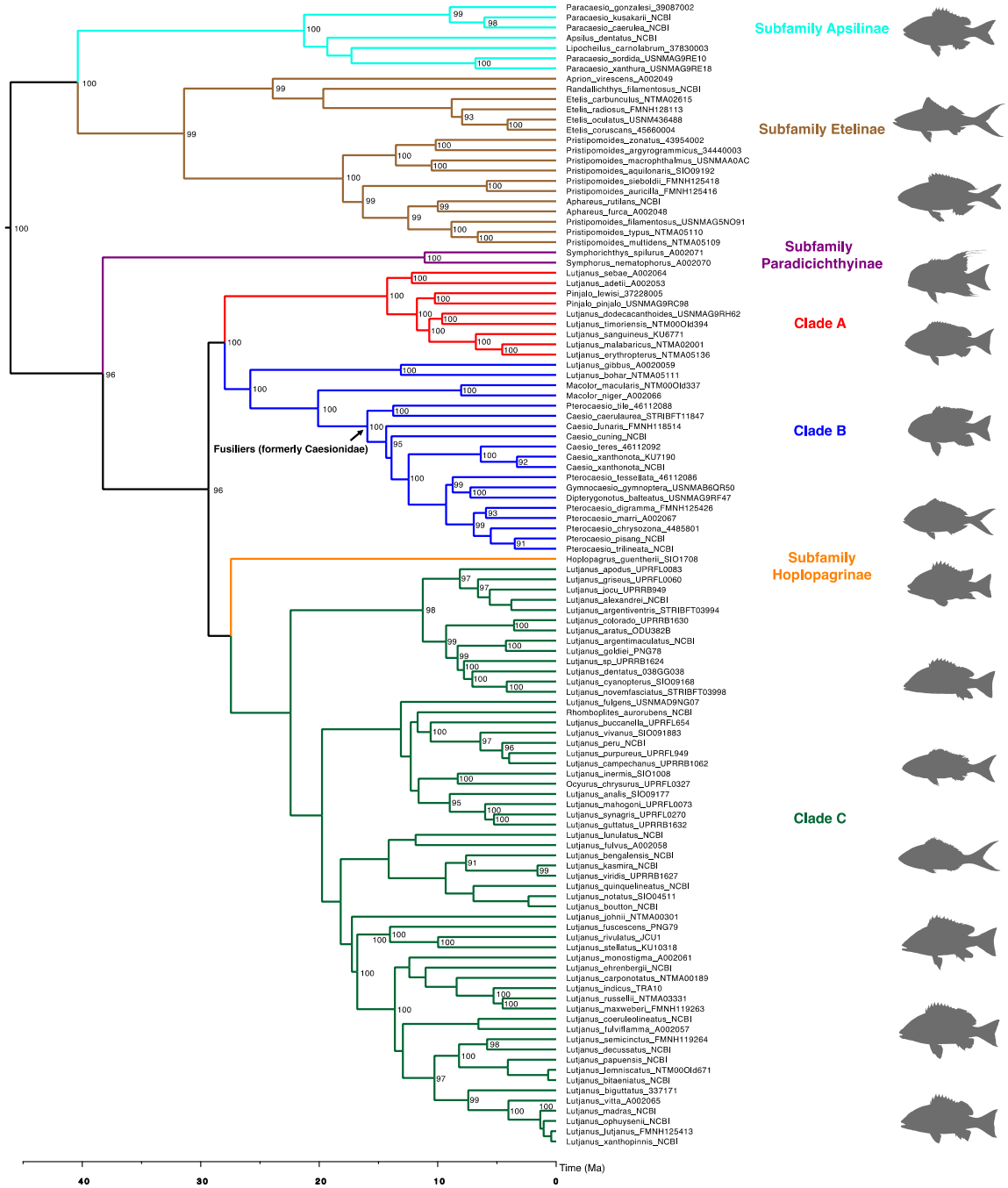


Figure S2. Phylogenetic tree inferred with RAxML for the expanded dataset ('master tree') and time-calibrated using MCMCTree. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support.

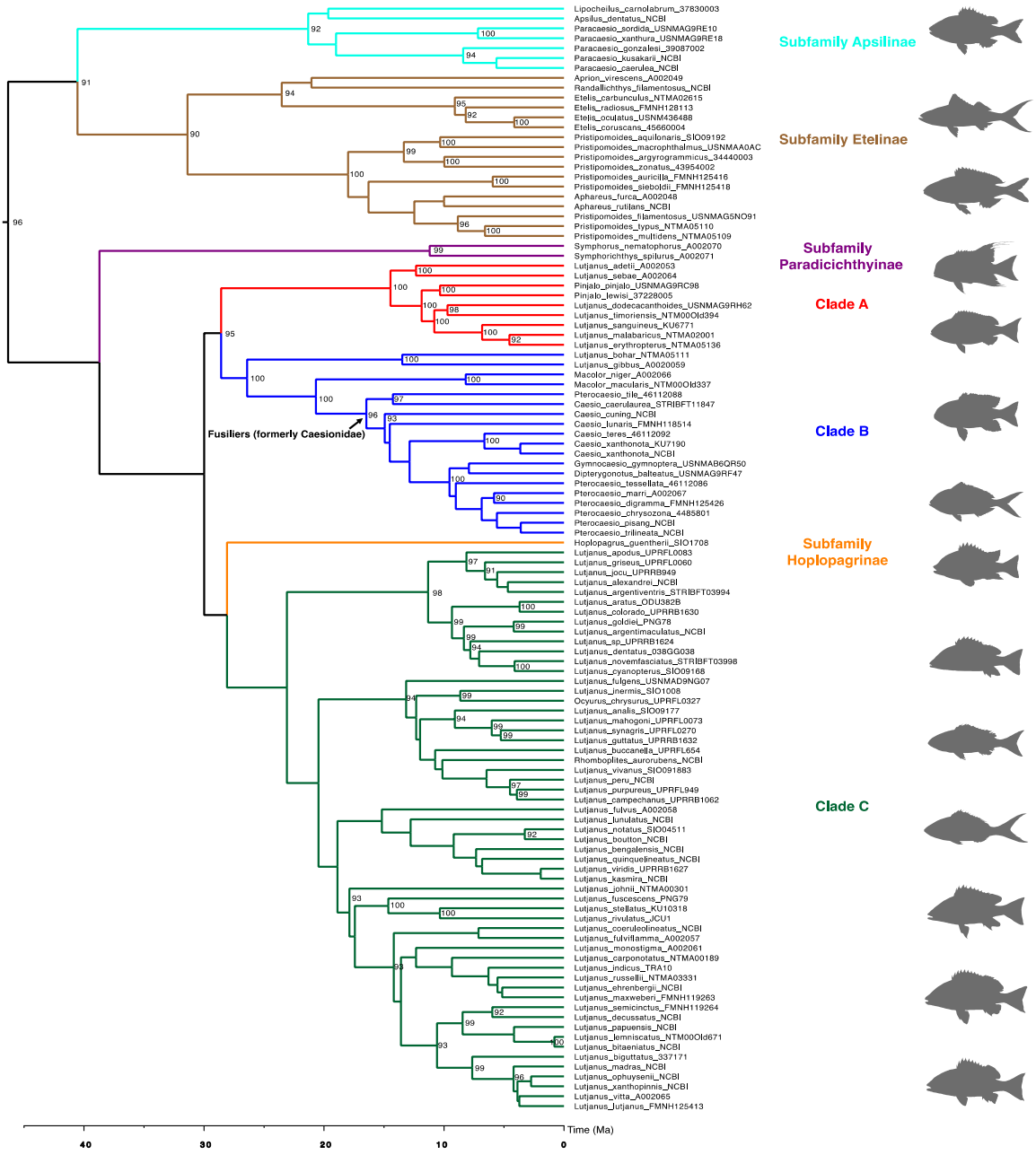


Figure S3. Phylogenetic tree inferred with ASTRAL-II for the expanded dataset and time-calibrated using MCMCTree. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support.

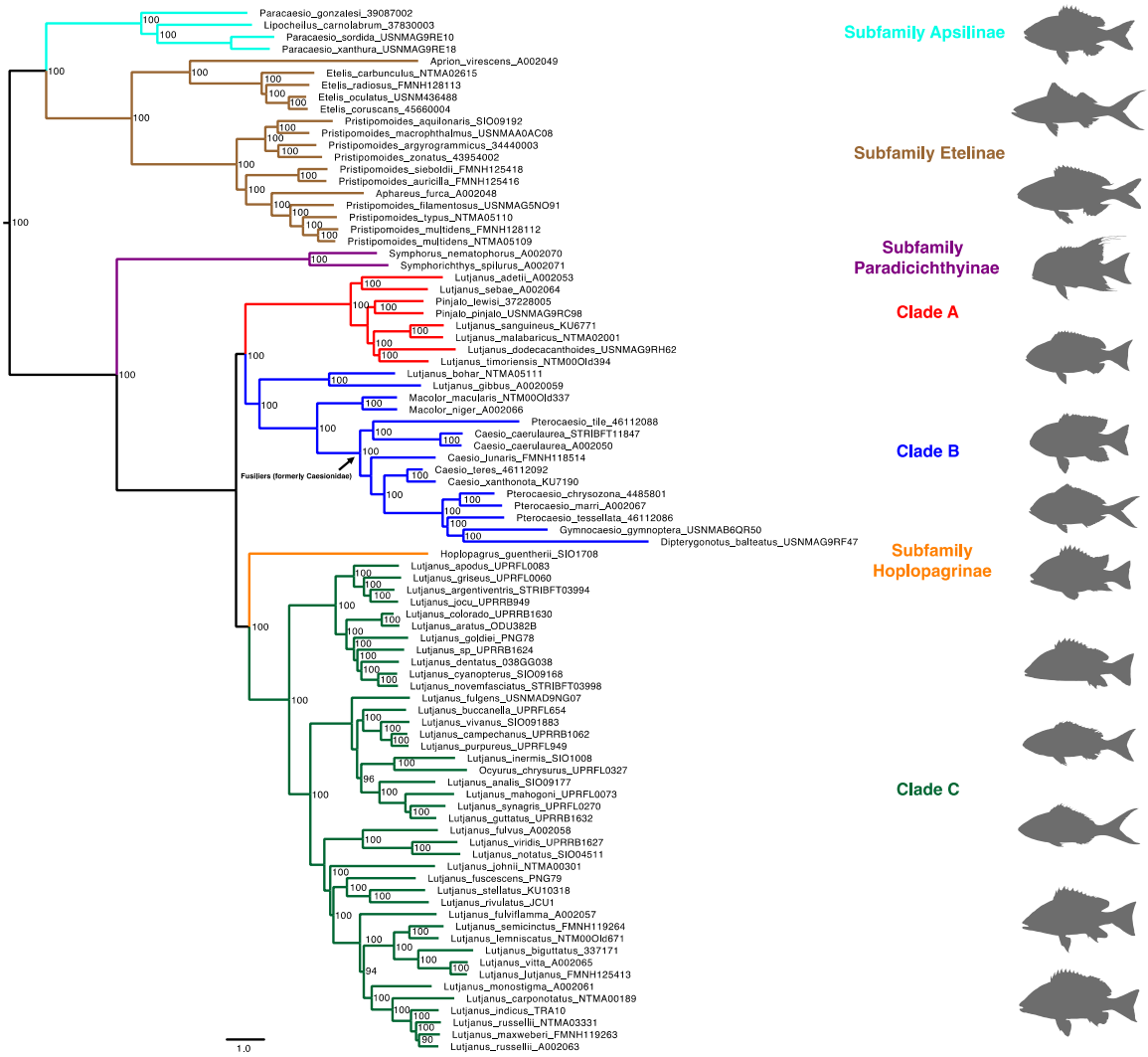


Figure S4. Phylogenetic tree inferred with RAxML for the reduced dataset. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support. The purpose of this inference was to assess sensitivity of phylogenetic results to missing data (16%). Thus, this tree is was not time calibrated.

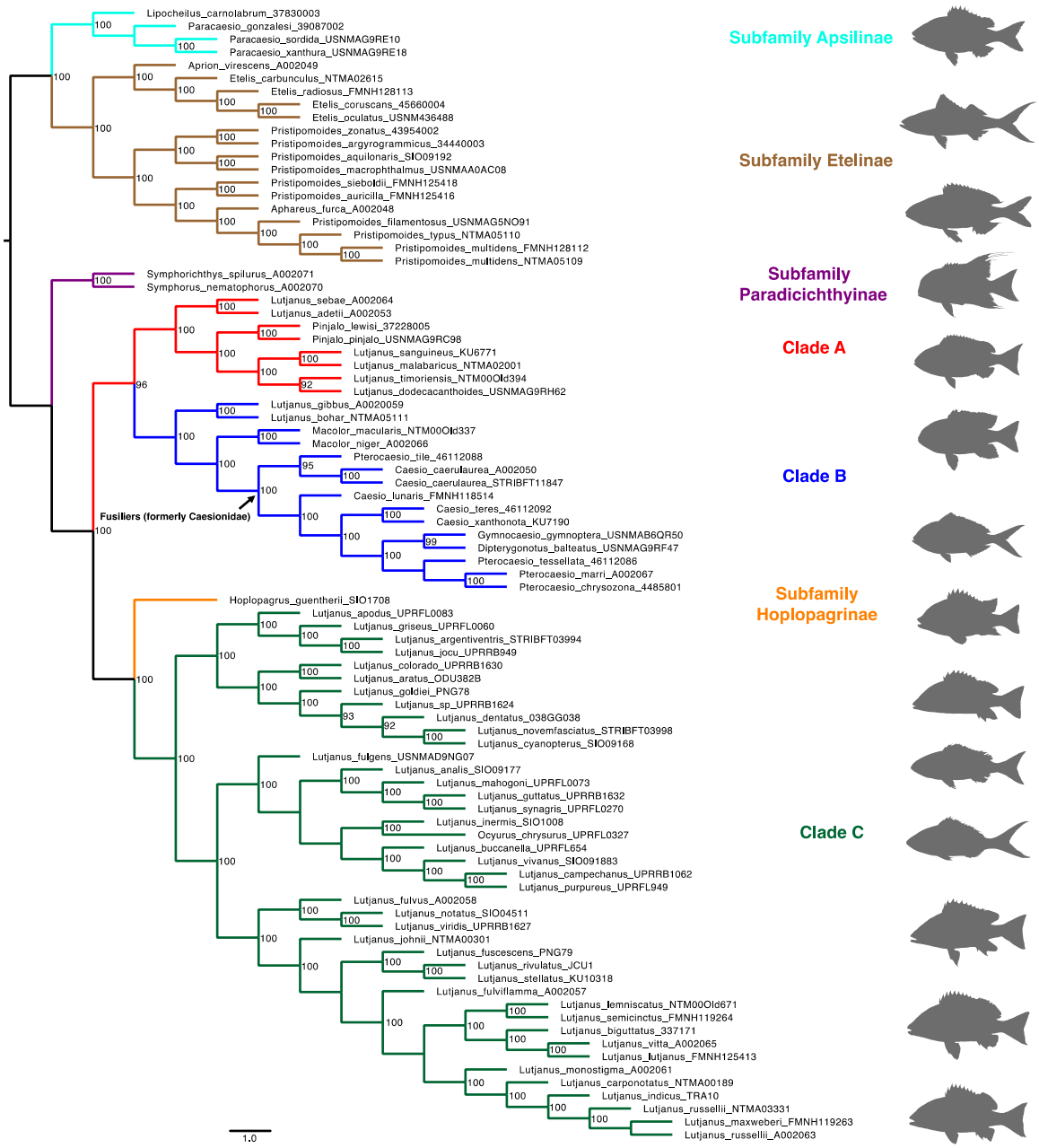


Figure S5. Phylogenetic tree inferred with ASTRAL-II for the reduced dataset. Colors indicate subfamilies and other major clades. Nodal values indicate bootstrap support. The purpose of this inference was to assess sensitivity of phylogenetic results to missing data (16%). Thus, this tree is was not time calibrated.

Table S2. HiSSE alternative models of lineage diversification and model fitting results for the Master tree using the 0.5 benthic/0.5 midwater probability scheme. ε = extinction fraction; τ = net turnover; lnLik = log likelihood; AIC = Akaike Information Criterion.

Model	Parameters				AIC
	States	Hidden states	Free parameters associated to (τ)	Free parameters associated to (ε)	
BiSSE null	0A, 1A	NA	1,1,0,0	1,1,0,0	846.50
BiSSE equal Q	**0A, 1A**	**NA**	**1,1,0,0**	**1,1,0,0**	**844.91**
BiSSE	0A, 1A	NA	1,2,0,0	1,2,0,0	848.85
HiSSE null	0A, 1A	0B, 1B	1,1,2,2	1,1,2,2	846.90
HiSSE full	0A, 1A	0B, 1B	1,2,3,4	1,2,3,4	850.03
HiSSE benthic	0A, 1A	0B	1,2,3,0	1,2,3,0	845.07
HiSSE midwater	0A, 1A	1B	1,2,0,3	1,2,0,3	873.76

Table S3. HiSSE alternative models of lineage diversification and model fitting results for the Master tree using the 0.1 benthic/0.9 midwater probability scheme. ε = extinction fraction; τ = net turnover; lnLik = log likelihood; AIC = Akaike Information Criterion.

Model	Parameters				AIC
	States	Hidden states	Free parameters associated to (τ)	Free parameters associated to (ε)	
BiSSE null	0A, 1A	NA	1,1,0,0	1,1,0,0	849.19
BiSSE equal Q	0A, 1A	NA	1,1,0,0	1,1,0,0	849.43
BiSSE	0A, 1A	NA	1,2,0,0	1,2,0,0	850.81
HiSSE null	0A, 1A	0B, 1B	1,1,2,2	1,1,2,2	851.47
HiSSE full	0A, 1A	0B, 1B	1,2,3,4	1,2,3,4	850.97
HiSSE benthic	**0A, 1A**	**0B**	**1,2,3,0**	**1,2,3,0**	**846.72**
HiSSE midwater	0A, 1A	1B	1,2,0,3	1,2,0,3	874.48

Table S4. HiSSE alternative models of lineage diversification and model fitting results for the Master tree using the 0.9 benthic/0.1 midwater probability scheme. ε = extinction fraction; τ = net turnover; lnLik = log likelihood; AIC = Akaike Information Criterion.

Model	Parameters				AIC
	States	Hidden states	Free parameters associated to (τ)	Free parameters associated to (ε)	
BiSSE null	**0A, 1A**	**NA**	**1,1,0,0**	**1,1,0,0**	**873.53**
BiSSE equal Q	0A, 1A	NA	1,1,0,0	1,1,0,0	873.79
BiSSE	0A, 1A	NA	1,2,0,0	1,2,0,0	876.64
HiSSE null	0A, 1A	0B, 1B	1,1,2,2	1,1,2,2	874.38
HiSSE full	0A, 1A	0B, 1B	1,2,3,4	1,2,3,4	882.74
HiSSE benthic	0A, 1A	0B	1,2,3,0	1,2,3,0	878.66
HiSSE midwater	0A, 1A	1B	1,2,0,3	1,2,0,3	932.35

Table S5. Comparison for stem and crown ages for Lutjanidae based on multiple studies.

Study	Mean crown age	95% HPD	Mean stem (root)	95% HPD
This study	46 Ma	(40-49 Ma)	64 Ma	
Alfaro et al. 2018 (34)	-	-	66 Ma	(60-72 Ma)
Hughes et al. 2018 (2)	-	-	79 Ma	(67-87 Ma)
Rabosky et al. 2018 (35)	52.87 Ma	-	66.324 Ma	
Betancur-R. et al. 2017 (32)	50 Ma	-	62 Ma	-
Frédérich & Santini, 2017 (33)	54 Ma	(45-66 Ma)		
Betancur-R. et al. 2013 (5)	-	-	64.62 Ma	(35.7-86.7 Ma)
Near et al. 2011 (31)	-	-	52 Ma	(47-57 Ma)

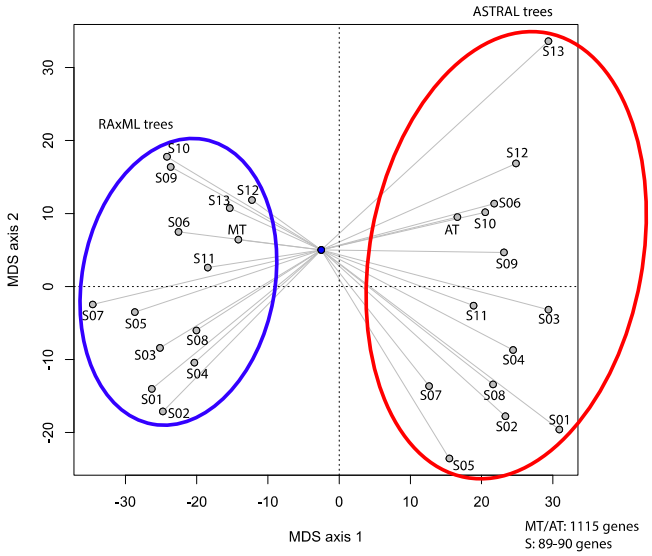


Figure S6. Tree spaces for the twenty-eight trees estimated in this study. MT: 'master tree', AT: alternative ASTRAL-II tree based on the full dataset. The blue dot represents the average (centroid) tree in tree space.

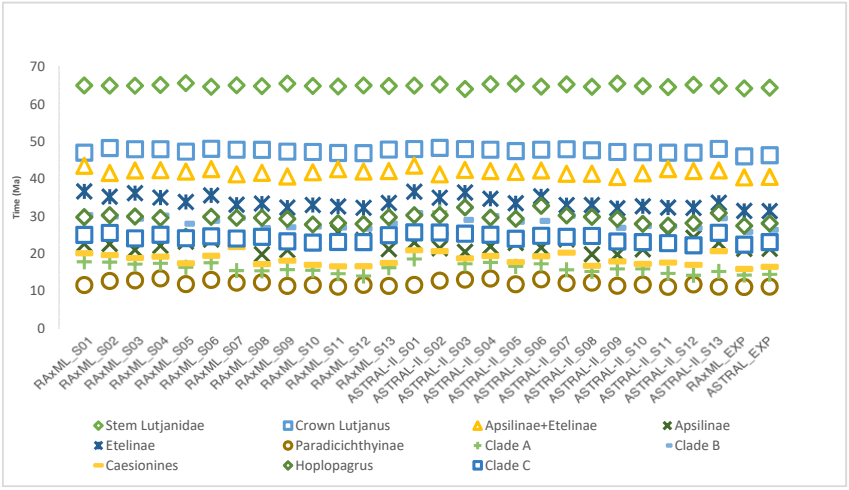


Figure S7. Divergence date uncertainty for major lutjanid clades based on the 28 trees dated with MCMCTree. 'Master tree': RAxML_EXP.

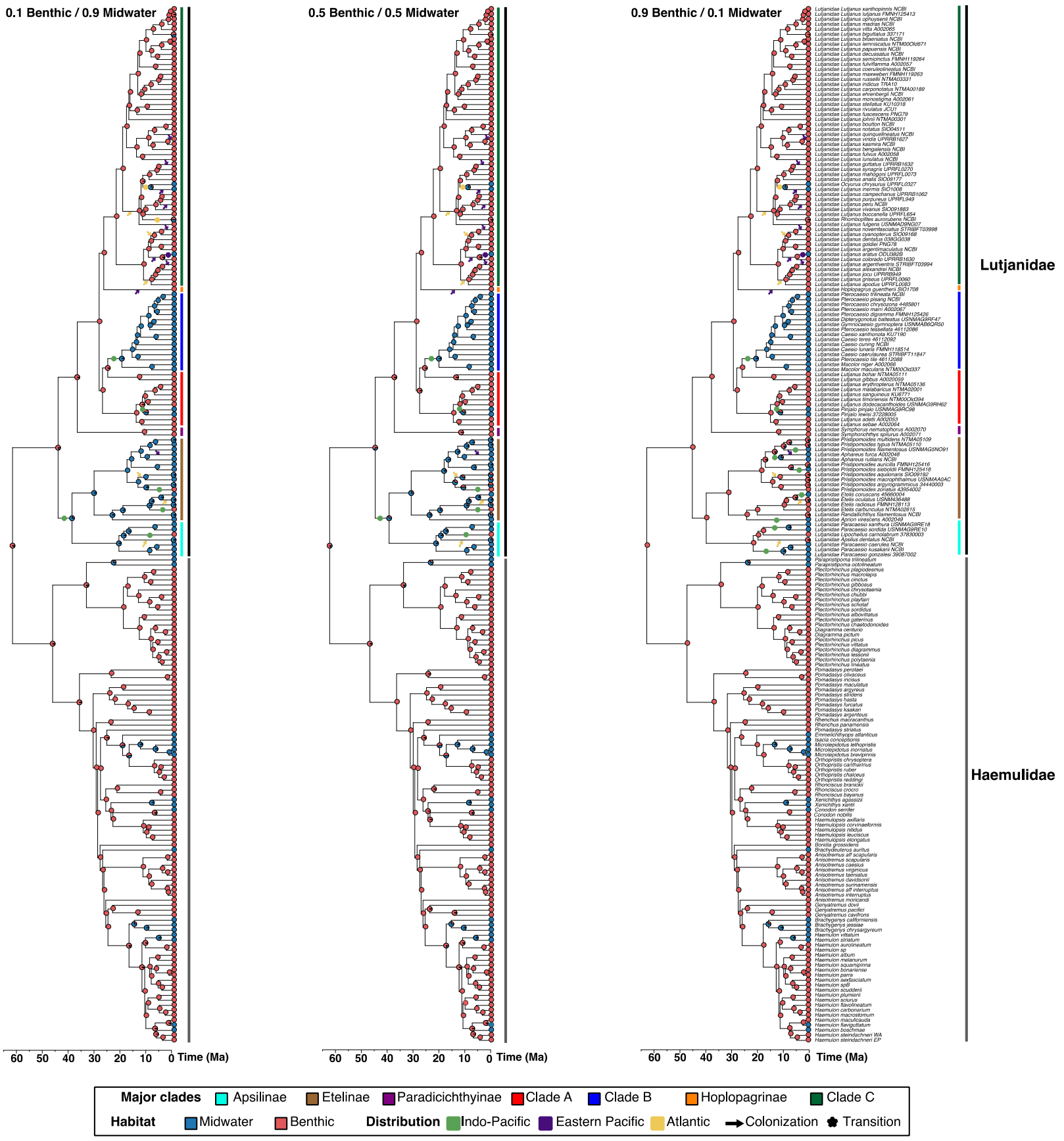


Figure S8. SIMMAP reconstructions based on the 'master tree' (RAxML) and expanded haemulid outgroups (from Tavera et al., 47) following three alternative probability schemes for coding species with ambiguous habitat affiliations: 0.1 benthic/0.9 midwater, 0.50 benthic/0.50 midwater, 0.9 benthic/0.1midwater. This figure shows how different coding schemes affect habitat reconstructions (see also Figs. S9-S11 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 Figs. S12-S14) are indicated with arrows. Stars indicate the most likely ancestral area where SIMMAP analyses identify a habitat transition.

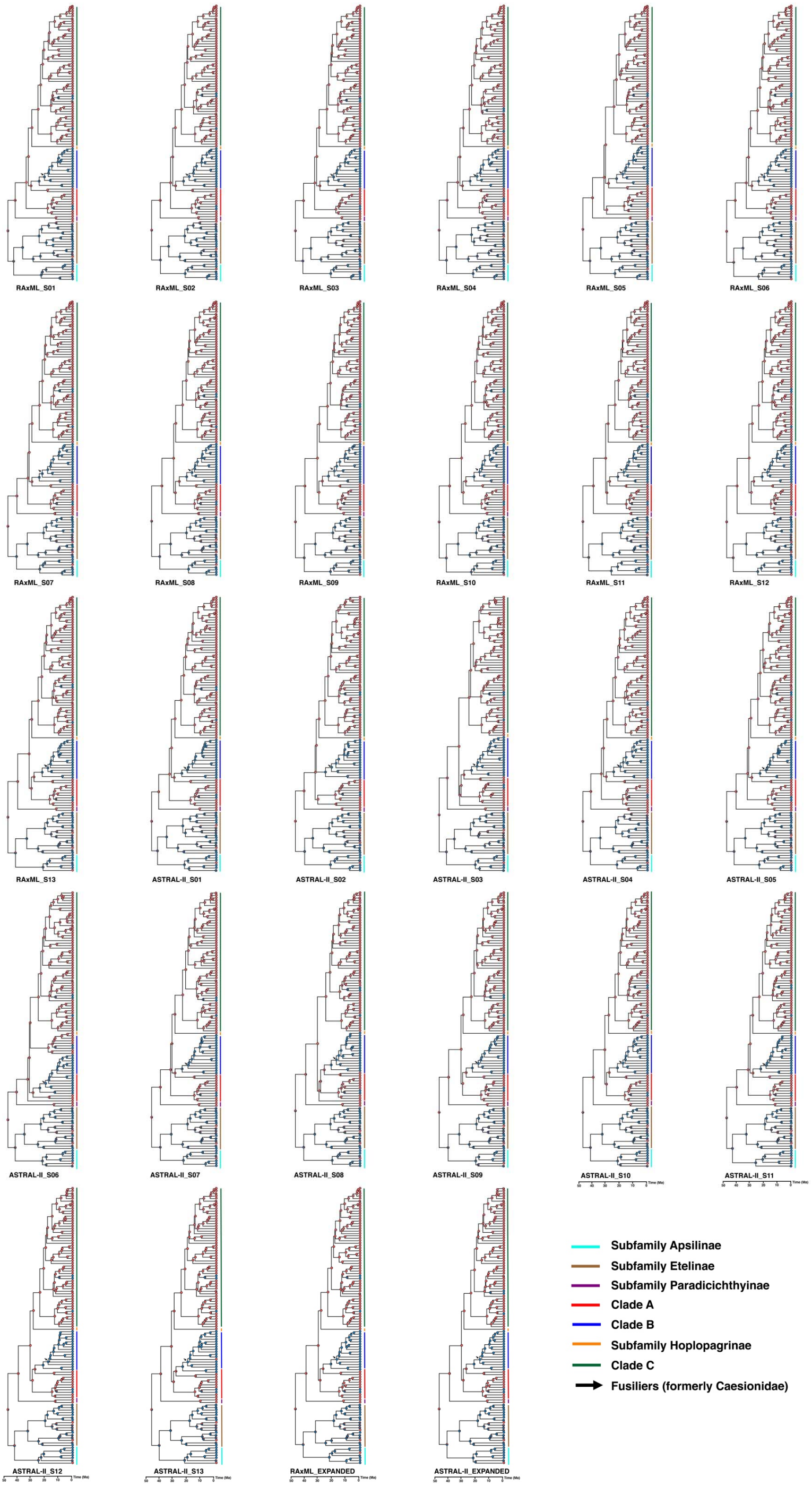


Figure S9. SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.5 benthic/0.5 midwater probability scheme. The reconstructions included all haemulid outgroups (see Fig. S8), but these were pruned out here to facilitate visualization.

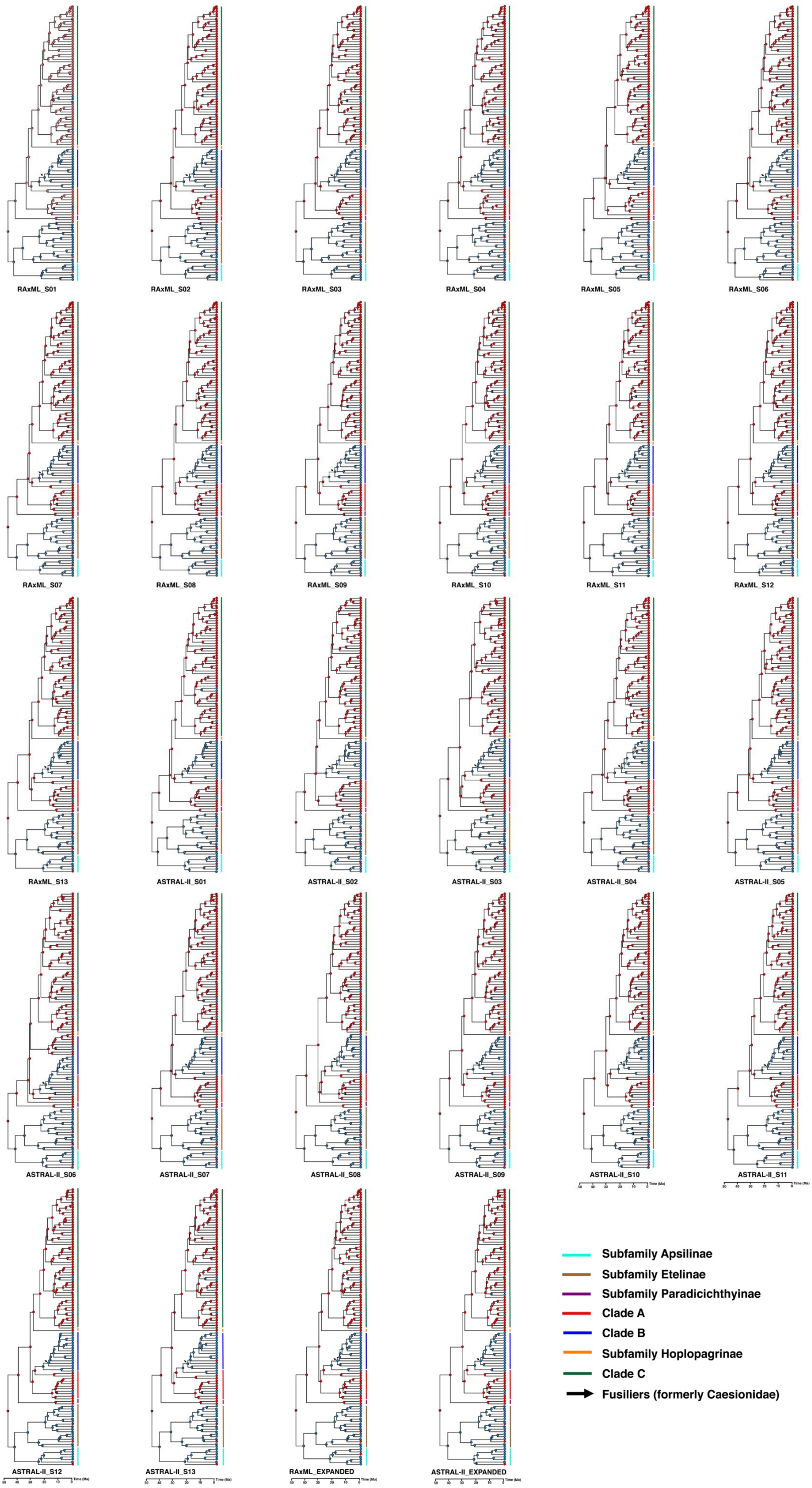


Figure S10. SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.1 benthic/0.9 midwater probability scheme. The reconstructions included all haemulid outgroups (see Fig. S8), but these were pruned out here to facilitate visualization.

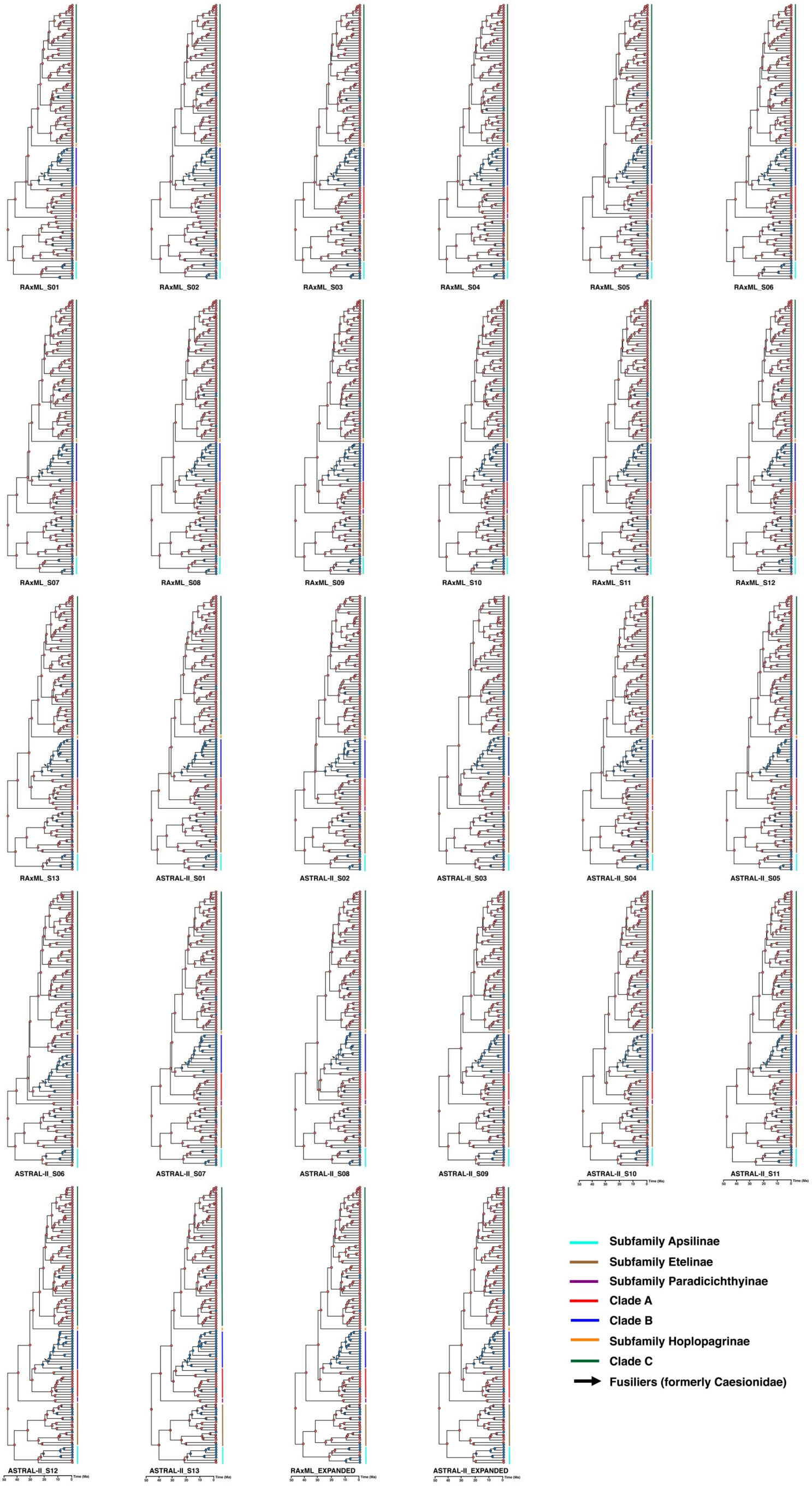


Figure S11. SIMMAP reconstructions for the 28 Lutjanidae-only trees following 0.9 benthic/0.1 midwater probability scheme. The reconstructions included all haemulid outgroups (see Fig. S8), but these were pruned out here to facilitate visualization.

Biogeographic analyses. The best-supported biogeographic model for lutjanids based on six areas was the BayAREA + j + w (AICw= 0.61; Table S6, Fig. S12), whereas for three areas the DEC + j + w model had a better fit (AICw= 0.43; Fig. S13). However, because support for BayAREA + j + w with three areas was also substantial (AICw= 0.2), we thus report all results (six and three areas) based on the BayAREA + j + w model (Fig. S14). 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 lutjanid lineages occurred within the Indo-Pacific. Lutjanines, 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 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.

Irrespective of the number of areas used (three or six), our biogeographic reconstructions suggest that lutjanids colonized the TEP nine different times and the Atlantic six times (Fig. 1). Note, however, that all these reconstructions are based on the 'master tree' only, and thus different topologies might potentially result in different number of colonization events. At least four lineages colonized the TEP by dispersing eastwards across the Eastern Pacific Barrier (including one event in *Aphareus furca* that is phylogeographic in scope), while five lineages colonized it from the WA through the Central American Seaway before the closure of the Isthmus of Panama. 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. Lineages that colonized the Atlantic used different routes. 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 of the Tethys Seaway, likely requiring lineages to colonize the Atlantic via South Africa or crossing the Eastern Pacific Barrier and the Central American Seaway. Regarding the remaining lineages that colonized the WA, at least one of them also crossed the Tethys Seaway while the other two took two alternative routes, via either the southern African coast or through the Eastern Pacific Barrier. The route that the MRCA of *Pristipomoides aquilonaris* and *P. macrophthalmus* took to colonize the WA is ambiguous as it could have occurred through any of the three routes mentioned. 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).

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 (Figs. 1, S8). While some areas feature more transitions than others (e.g., Indo-Pacific vs. Eastern Pacific; Fig. 1, S8), the ubiquitous nature of habitat transitions in lutjanids is a remarkable result of this study that highlights the deterministic character of these changes.

Table S6. Summary statistics of the 12 biogeographic models implemented in BioGeoBEARS for the six- and three-areas schemes. LnL = LogLikelihood; AICc = corrected Akaike Information Criterion.

	Six-areas scheme			**Three-areas scheme**		
Models	LnL	AICc	AICc weight	LnL	AICc	AICc weight
DEC	-275.3	554.8	4.10E-12	-104.7	213.6	1.50E-16
DEC+J	-273.4	553.1	9.40E-12	-71.01	148.2	0.023
DEC+W	-281.2	568.6	4.00E-15	-100.1	206.4	5.30E-15
DEC+J+W	-272.5	553.3	8.50E-12	-67.02	142.4	0.43
DIVALIKE	-276.6	557.3	1.20E-12	-99.45	203	3.00E-14
DIVALIKE+J	-275.9	558.1	7.80E-13	-72.04	150.3	0.0083
DIVALIKE+W	-284.9	575.9	1.00E-16	-95.02	196.3	8.70E-13
DIVALIKE+J+W	-275.3	559	4.90E-13	-67.27	142.9	0.33
BAYAREALIKE	-309.6	623.3	5.50E-27	-155.8	315.7	1.00E-38
BAYAREALIKE+J	-249	504.2	0.39	-72.46	151.1	0.0054
BAYAREALIKE+W	-308.5	623.2	5.70E-27	-150.4	307.1	7.60E-37
BAYAREALIKE+J+W	-247.5	503.3	0.61	-67.77	143.9	0.2

Geometric morphometric analyses. The number of meaningful PC axes varied among the three morphometric datasets (Fig. S15; Table S7). 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 (Fig. 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 convergences. This pattern is further confirmed based on the threshold model (81), where the full-body shape dataset reveals substantial correlation between the two habitat states and PC1 (r^2=0.57-0.67; based on different habitat coding schemes), which captured differences in body elongation and caudal fin shape. 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=0.42-0.57) and fins-only (r^2=0.56-0.69) 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=0.07-0.24; for the full-body shape dataset). These results suggest that ecomorphological convergence is less clearly associated with PC2-PC4 axes than it is to the main PC1 axis.

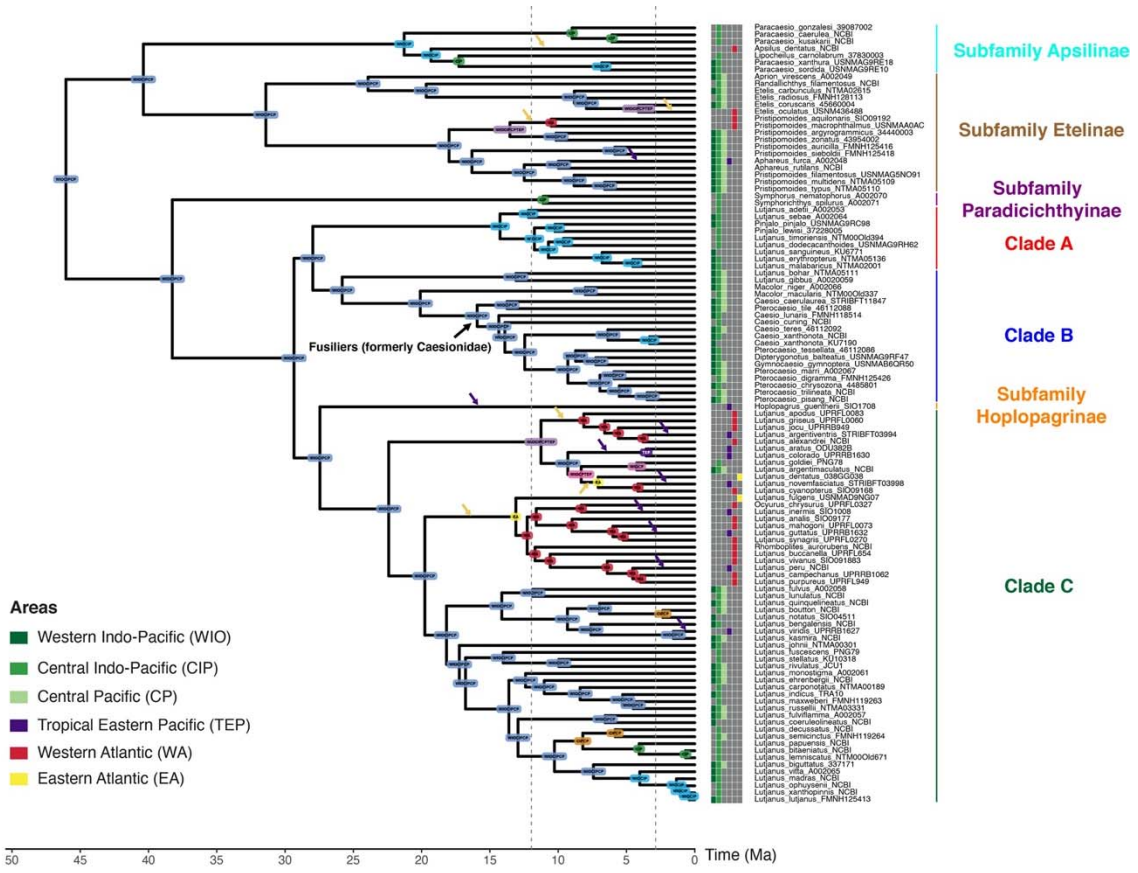


Figure S12. Ancestral area reconstructions (BioGeoBEARS) for Lutjanidae using the best-supported biogeographical model for six areas (BAYAREALIKE+j+w) applied to the 'master tree.' Boxes represent the geographic distribution of extant species. Dotted lines represent the time constraints that correspond 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.

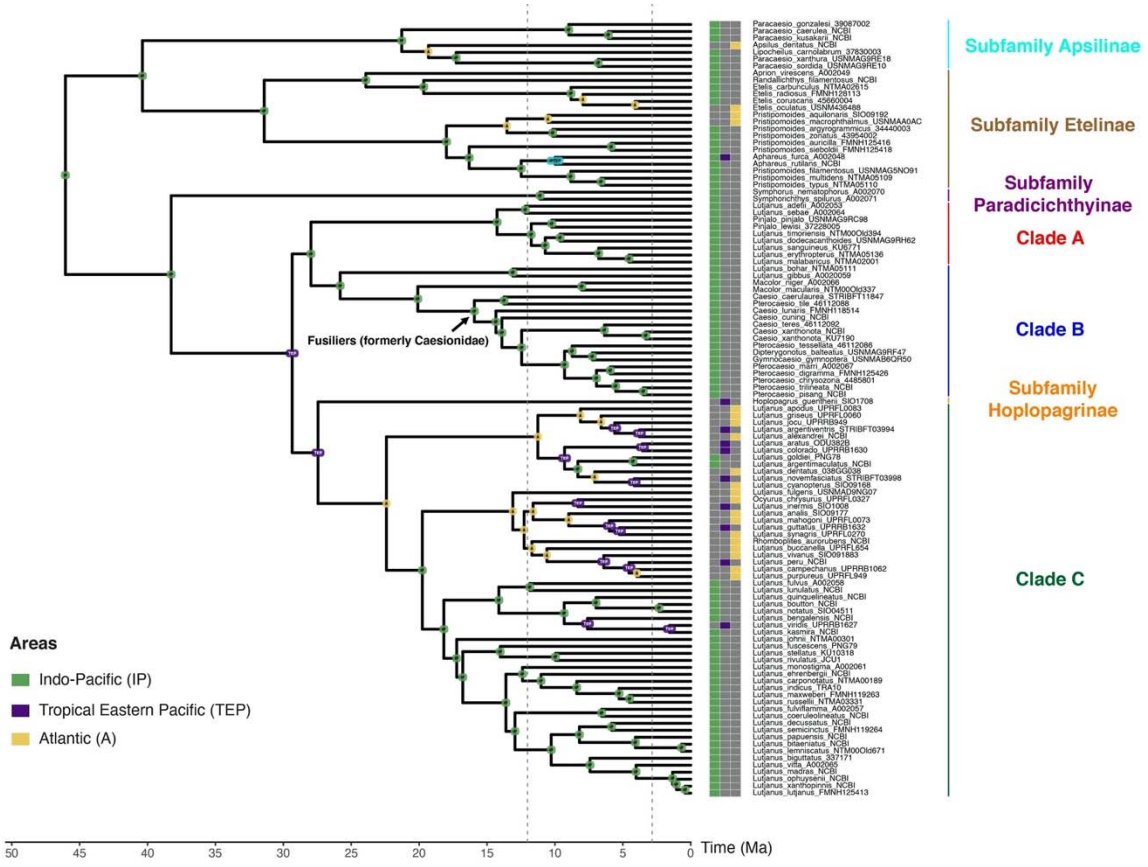


Figure S13. Ancestral area reconstructions (BioGeoBEARS) for Lutjanidae using the best-supported biogeographical model for three areas (DEC +j+w) applied to the 'master tree.' Boxes represent the geographic distribution of extant species. Dotted lines represent time constraints that correspond to two 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).

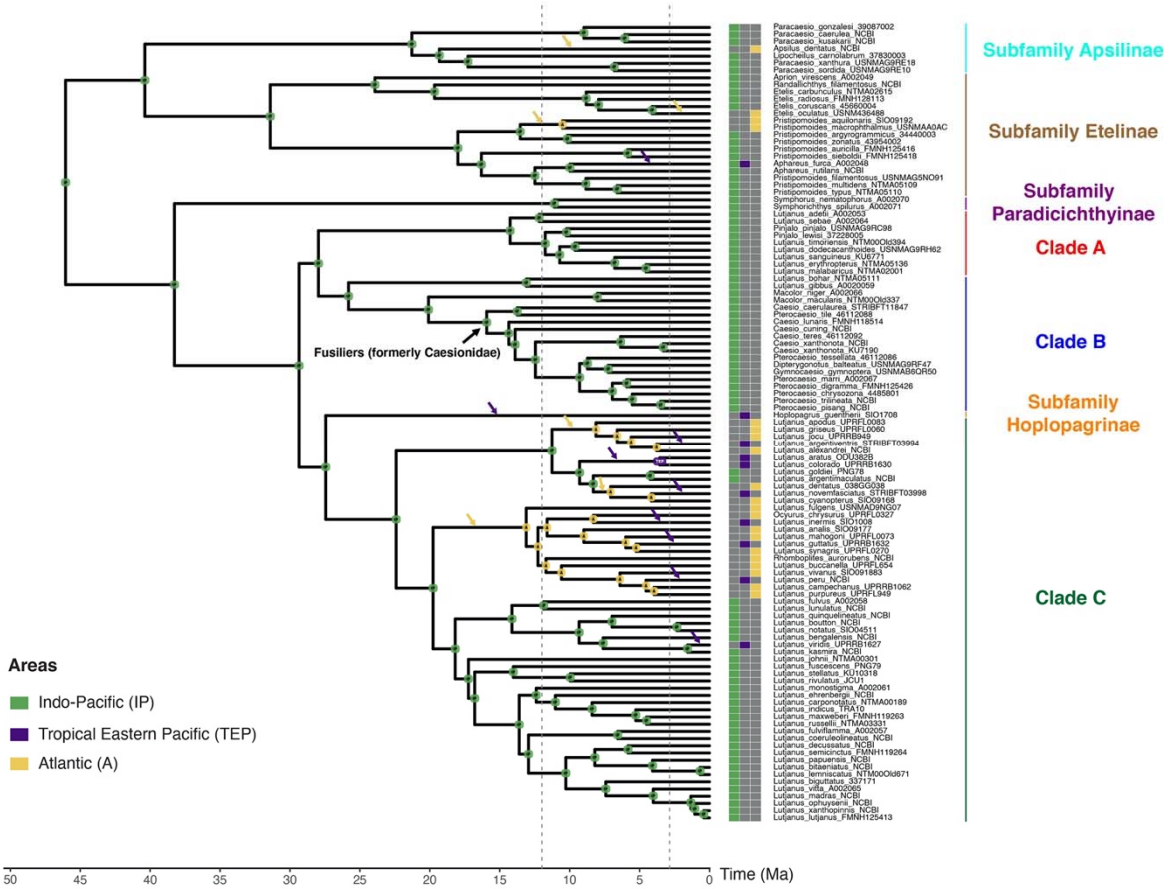


Figure S14. Ancestral area reconstructions (BioGeoBEARS) for Lutjanidae based on the three-areas scheme applied to the 'master tree' but using the best-fit model for six areas (BAYAREALIKE+j+w; see Fig. S8). Boxes represent the geographic distribution of extant species. Dotted lines represent time constraints 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.

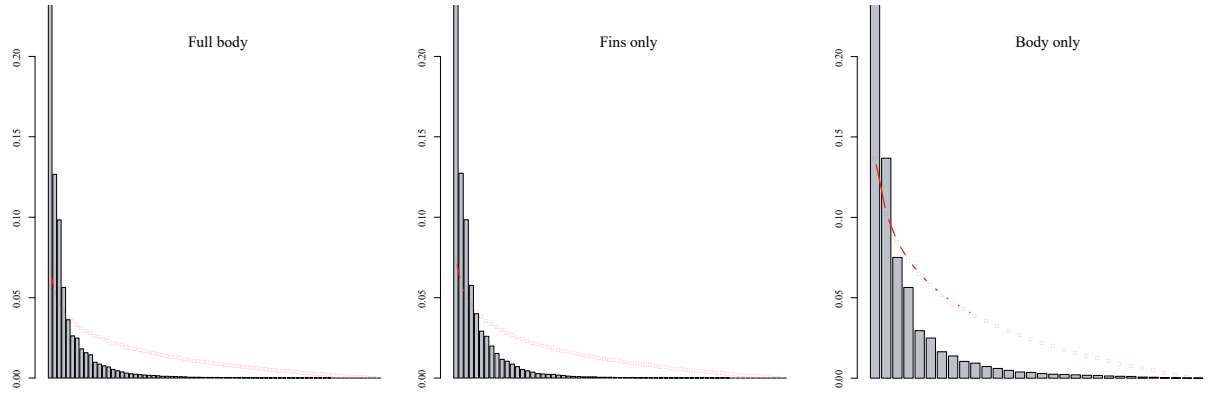


Figure S15. Plots of the broken-stick method showing PC axis variation (x axis) for full body-shape, fins-only, and body-only datasets. Plots represent (red dashed lines) the broken stick

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.

Table S7. Proportion of variance and standard deviation for all major PC axes (up to 95%) and datasets analyzed.

Dataset		PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Full body shape	Standard deviation	0.067	0.034	0.032	0.022	0.018	0.015	0.015	0.013	0.012	0.011	0.009	0.009	0.008
	Proportion of Variance	0.507	0.127	0.098	0.056	0.036	0.026	0.025	0.018	0.016	0.014	0.008	0.009	0.007
	Cumulative Proportion	0.507	0.634	0.732	0.788	0.825	0.851	0.876	0.894	0.910	0.924	0.934	0.943	0.950
Body only	Standard deviation	0.031	0.015	0.011	0.010	0.007	0.006	0.005	0.005	0.004	0.004	-	-	-
	Proportion of Variance	0.009	0.007	0.006	0.005	0.004	0.004	0.003	0.003	0.002	0.009	-	-	-
	Cumulative Proportion	0.583	0.720	0.795	0.851	0.881	0.906	0.926	0.936	0.946	0.956	-	-	-
Fins only	Standard deviation	0.063	0.031	0.028	0.021	0.018	0.015	0.014	0.012	0.011	0.009	0.009	0.008	-
	Proportion of Variance	0.511	0.127	0.098	0.056	0.036	0.026	0.025	0.018	0.016	0.012	0.008	0.009	-
	Cumulative Proportion	0.511	0.639	0.737	0.793	0.829	0.855	0.880	0.898	0.914	0.926	0.934	0.943	-

Convergence analyses. Snappers and fusiliers display considerable morphological diversity in body shape concerning body depth and fin shape (Fig. 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).

Trait evolution in benthic and midwater lineages. 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 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).

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 (Table S8). 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= 37-45%). 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 ≥0.38; Table S8).

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*= 1.3–1.4; Figs. S19-S20) 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*= 0.83–0.88; *p* >0.95). These results support the idea that morphological diversity is high among benthic species, and strong convergent evolution is largely restricted to midwater lutjanids.

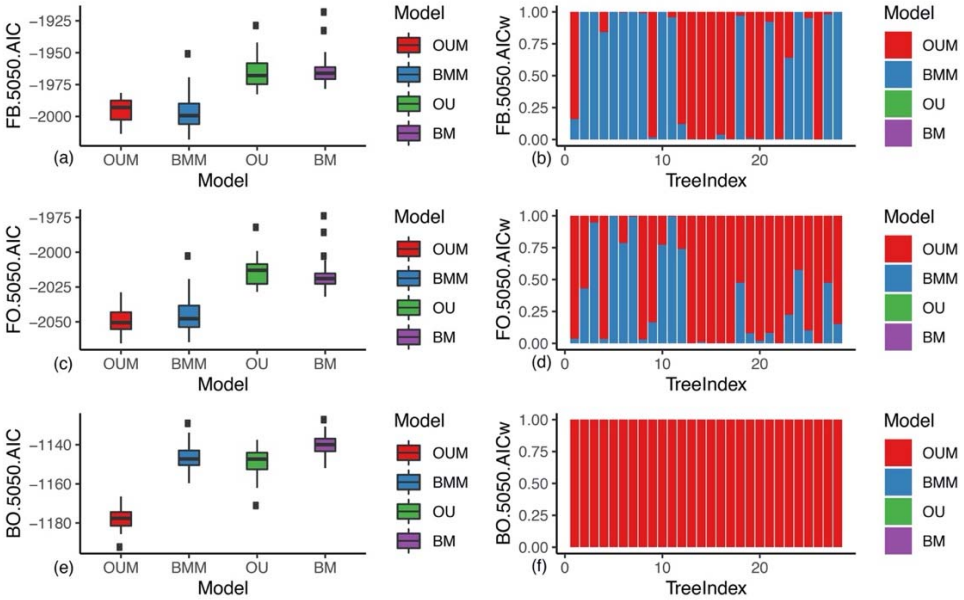


Figure S16. Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.5 benthic/0.5 midwater probability scheme. Distribution of the Akaike Information Criterion (AIC) and Akaike weight (AICw) values for the three alternative models of continuous 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.

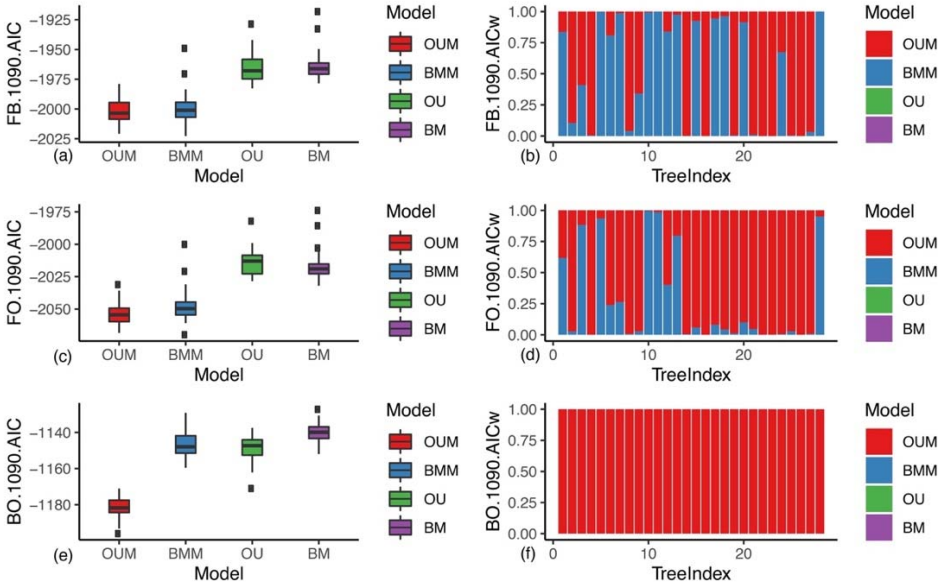


Figure S17. Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.1 benthic/0.9 midwater probability scheme. Distribution of the Akaike Information Criterion (AIC) and Akaike weight (AICw) values for the three alternative models of continuous 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.

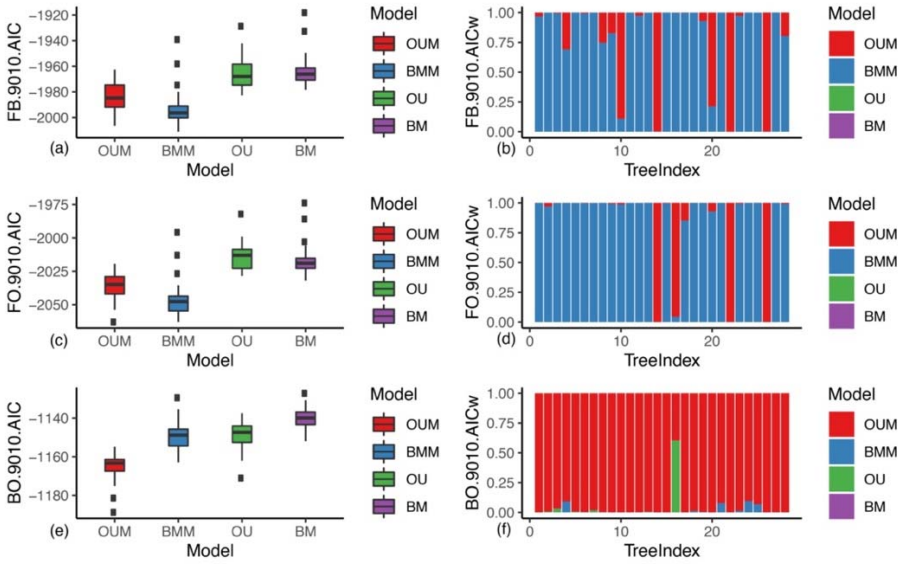


Figure S18. Model-fitting comparisons for alternative models of morphological evolution based on a set of 28 phylogenetic trees 0.9 benthic/0.1 midwater probability scheme. Distribution of the

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.

Table S8. C1-C5 convergence measures and p-values for convevol analyses ran using the full body-shape, body-only, and fins-only datasets. Asterisks represents statistically significant values. * = p< 0.001.

Habitat coding	Morphometric dataset	C1	C2	C3	C4	C5
0.5/0.5	**FB**	0.40 (p<0.01)	0.06 (p<0.01)	0.17 (p<0.01)	0.01 (p=0.01)	6 (p=0.48)
	FO	0.39 (p<0.01)	0.06 (p<0.01)	0.17 (p<0.01)	0.01 (p=0.01)	4 (p=0.67)
	BO	0.43 (p<0.01)	0.03 (p<0.01)	0.21 (p<0.01)	0.01 (p=0.01)	7 (p=0.38)
0.9/0.1	**FB**	0.38 (p<0.01)	0.06 (p<0.01)	0.16 (p<0.01)	0.01 (p=0.05)	3 (p=0.84)
	FO	0.37 (p<0.01)	0.06 (p<0.01)	0.16 (p<0.01)	0.01 (p=0.03)	3 (p=0.81)
	BO	0.43 (p<0.01)	0.03 (p<0.01)	0.20 (p<0.01)	0.01 (p=0.03)	7 (p=0.40)
0.1/0.9	**FB**	0.41 (p<0.01)	0.07 (p<0.01)	0.18 (p<0.01)	0.01 (p<0.01)	6 (p=0.48)
	FO	0.41 (p<0.01)	0.06 (p<0.01)	0.18 (p<0.01)	0.01 (p=0.03)	4 (p=0.68)
	BO	0.45 (p<0.01)	0.03 (p<0.01)	0.22 (p<0.01)	0.01 (p=0.05)	7 (p=0.42)

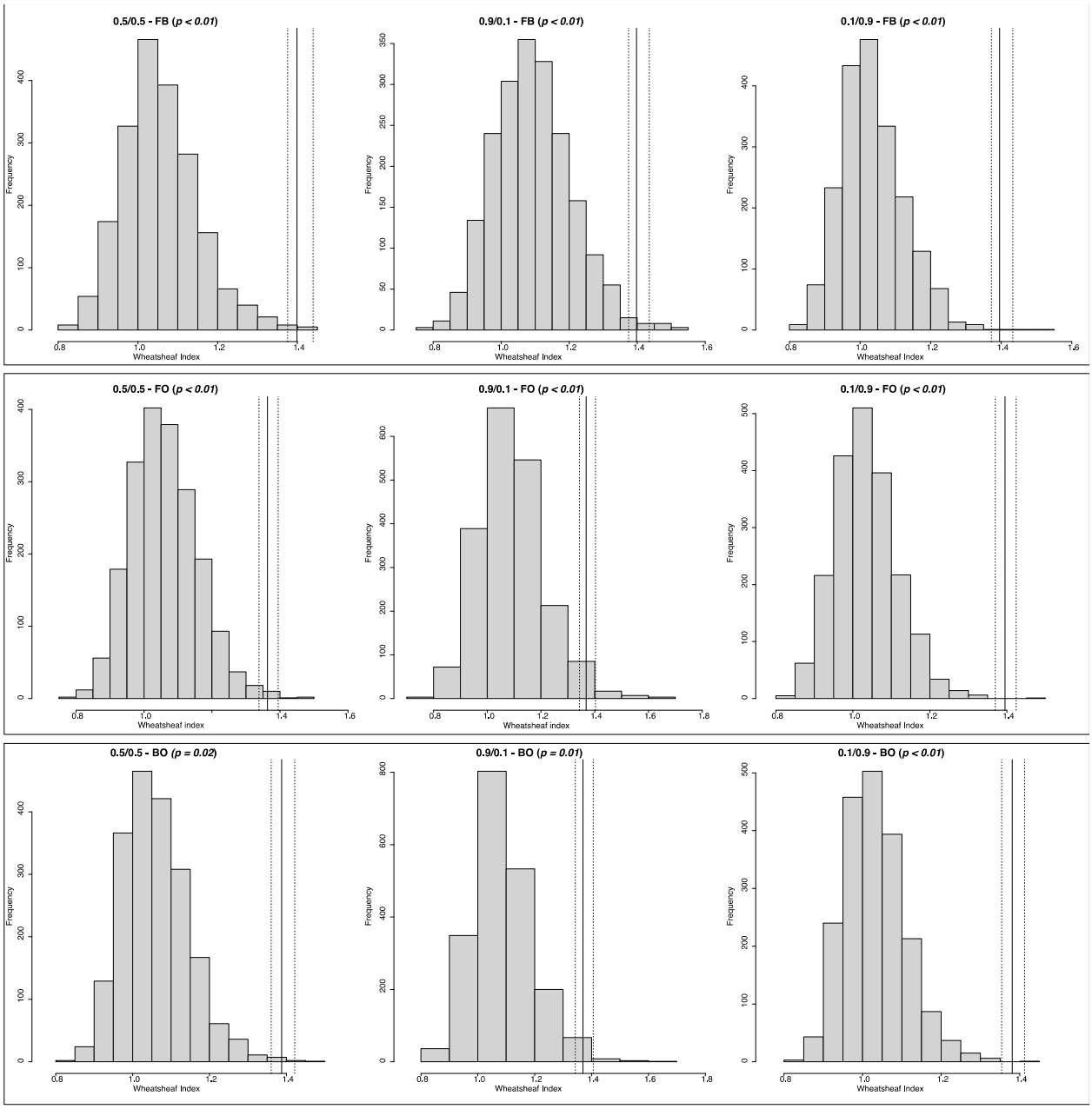


Figure S19. Histograms representing the distribution of bootstrapped Wheatsheaf index values for all morphometric datasets based on PCA and the three different habitat coding schemes. Black thick lines represent the calculated Wheatsheaf index. Dashed lines show 95% confidence interval.

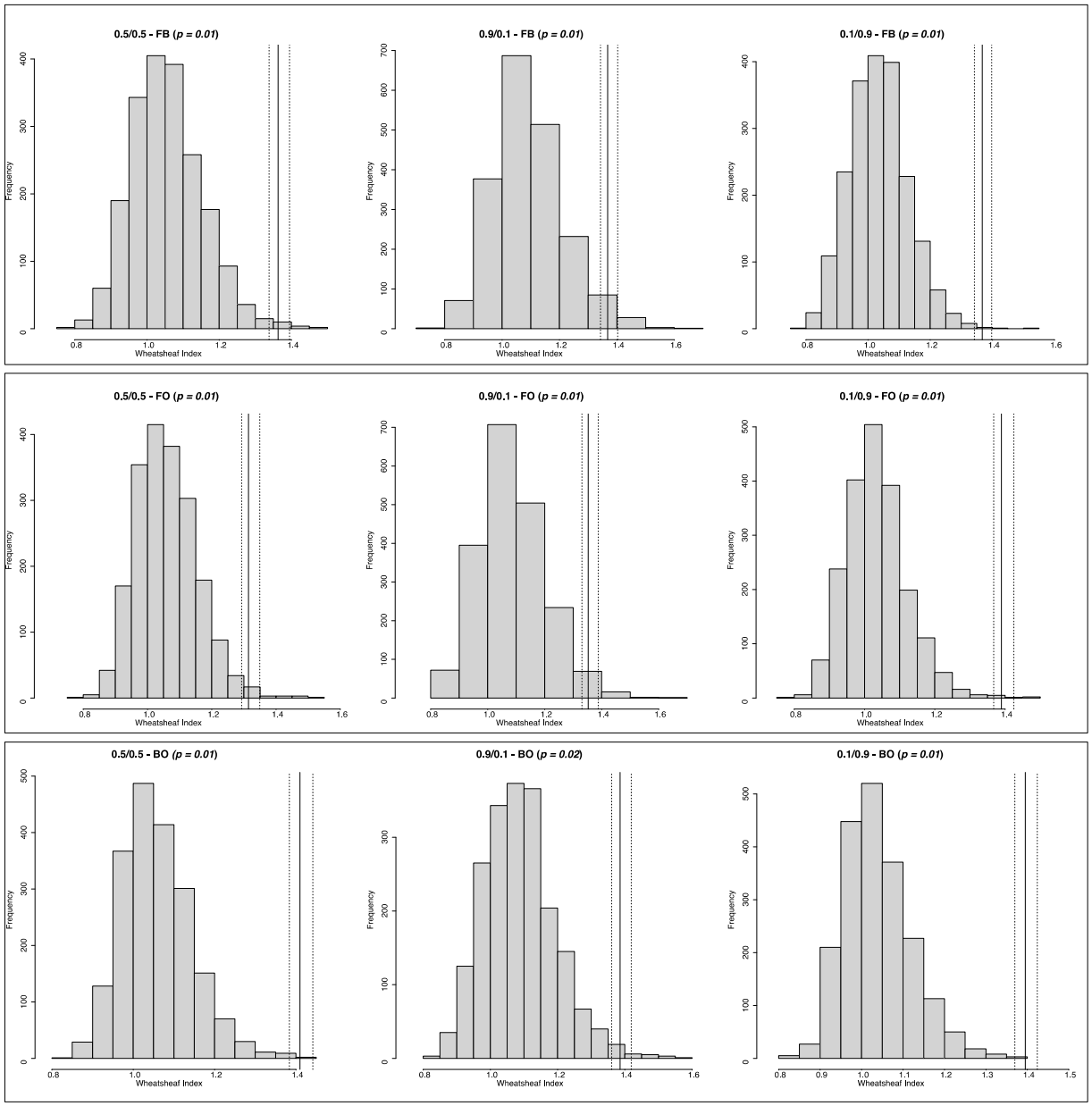


Figure S20. Histograms representing the distribution of bootstrapped Wheatsheaf index values for all morphometric datasets based on pPCA and the three different habitat coding schemes. Black thick lines represent the calculated Wheatsheaf index. Dashed lines show 95% confidence interval.

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 (69), 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; Fig. S15). 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; Table S9). To properly account for phylogenetic co-variation we report all downstream analyses based on pPCA (61, 82). 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 (Table S9). 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 (76, 83–88). 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. See also Johnson (29): "The family Caesionidae is clearly a classical example of a group with identifiable affinities, which has invaded a very different adaptive zone than that occupied by its closest relatives and has undergone a moderate radiation. This was made possible by an innovative restructuring of the functional complex of the upper jaw (permitting extreme protrusibility for planktivorous feeding) and an alteration of the basic body configuration (providing greater and more rapid swimming ability."

Results for the multivariate ℓ1ou and SURFACE analyses for the twenty-eight trees produced largely congruent results (Fig. S25; Dataset S6, supplementary spreadsheet), which are summarized in Table S9; 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*; Fig. S21) 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 shapes. By contrast, *L. cyanopterus, L. dentatus*, *L. madras, L. carnolabrum,* and *L. biguttatus* tend to converge on deep-bodied phenotypes (Fig. S22). 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 (Fig. S23). 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 (Fig. S24). 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 (Fig. S25) 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 65). 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 S26). 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.

Table S9. ℓ1ou adaptive and convergent regimes estimated using the 'master tree' (RAxML) and an alternative tree (ASTRAL-II) under AICc or pBIC models. FBS: Full body-shape; BO: Body only; FO: Fins only.

	AICc ℓ1ou			pBIC ℓ1ou		
Dataset	**shifts**	**regimes**	**conv. regimes**	**shifts**	**regimes**	**conv. regimes**
Master_tree_FBS_pPC1	14	4	3	1	1	0
Master_tree_FBS_PC1	15	6	5	1	1	0
Alternative_tree_FBS_pPC1	14	4	3	1	1	0
Alternative_tree_FBS_PC1	7	4	2	1	1	0
Master_tree_FBS_pPC1-pPC4	16	9	4	9	8	1
Master_tree_FBS_PC1-PC4	19	13	6	10	9	1
Alternative_tree_FBS_pPC1-pPC4	14	9	3	9	8	1
Alternative_tree_FBS_PC1-PC4	18	11	4	9	8	1
Master_tree_BO_pPC1	13	8	2	1	1	0
Master_tree_BO_PC1	10	5	4	1	1	0
Alternative_tree_BO_pPC1	9	5	1	1	1	0
Alternative_tree_BO_PC1	10	5	4	1	1	0
Master_tree_BO_pPC1-pPC2	13	7	4	5	4	1
Master_tree_BO_PC1-PC2	12	6	3	5	4	1
Alternative_tree_BO_pPC1-pPC2	11	5	3	3	2	1
Alternative_tree_BO_PC1-PC2	10	5	4	1	1	0
Master_tree_FO_pPC1	14	6	4	1	1	0
Master_tree_FO_PC1	17	7	6	1	1	0
Alternative_tree_FO_pPC1	9	4	2	1	1	0

Alternative_tree_FO_PC1	9	4	3	1	1	0
Master_tree_FO_pPC1-pPC4	13	10	3	9	9	0
Master_tree_FO_PC1-PC4	19	11	4	8	8	0
Alternative_tree_FO_pPC1-pPC4	14	9	4	8	8	0
Alternative_tree_FO_PC1-PC4	14	9	4	7	7	0

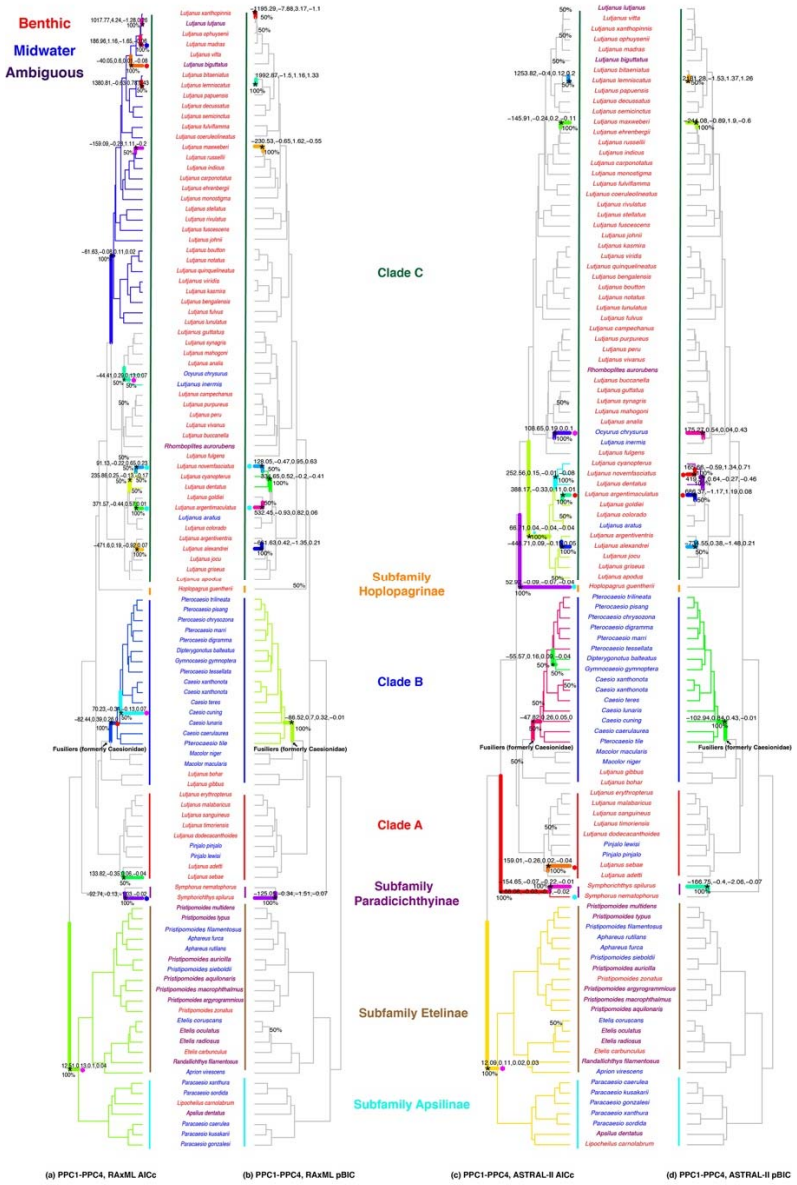


Figure S21. Adaptive and convergent shifts in Lutjanidae for the full-body shape dataset (first four PC axes) with ℓ1ou using the AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

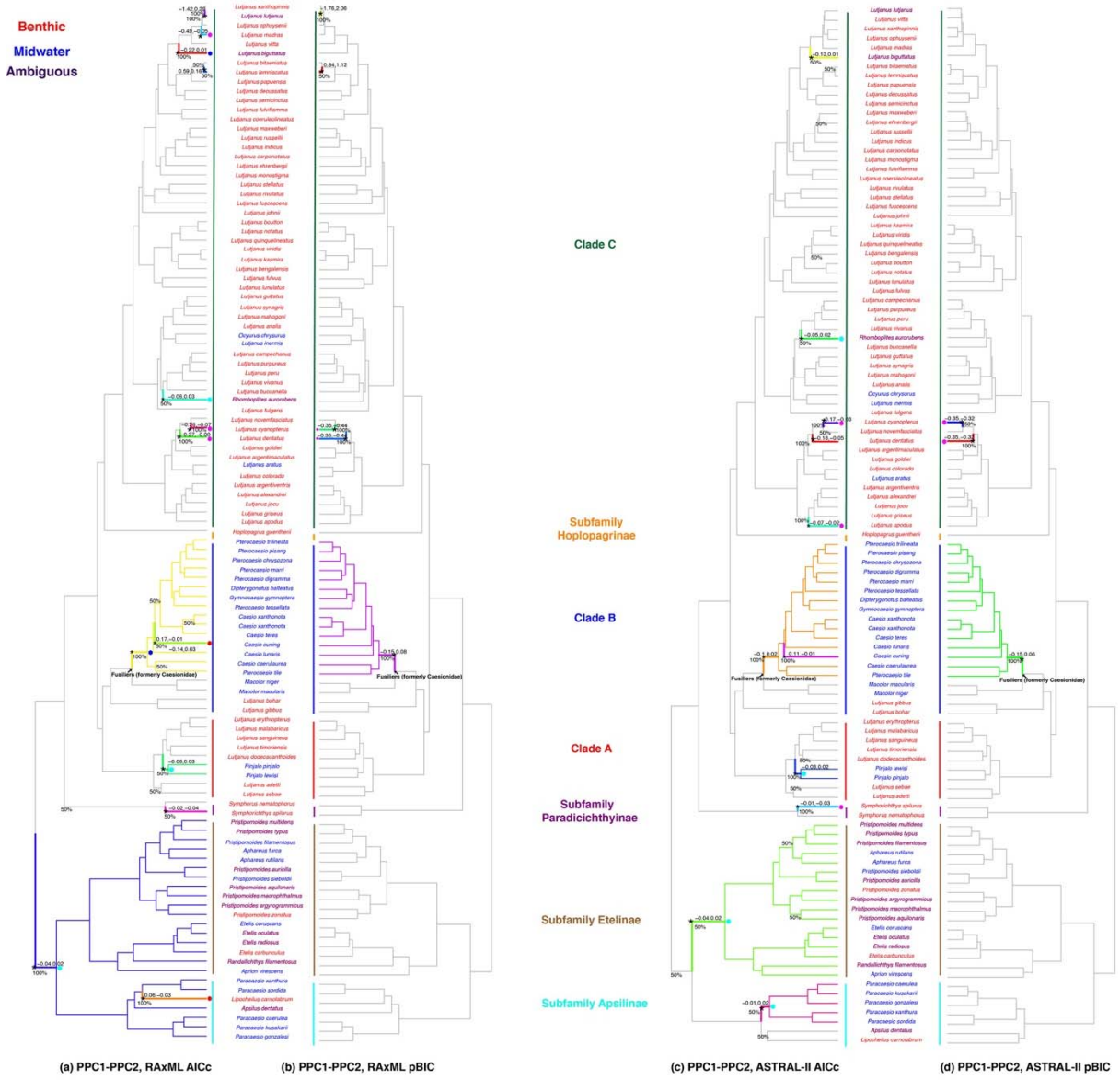


Figure S22. Adaptive and convergent regimes in Lutjanidae for the body-only dataset (first two PC axes), with ℓ1ou using both AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

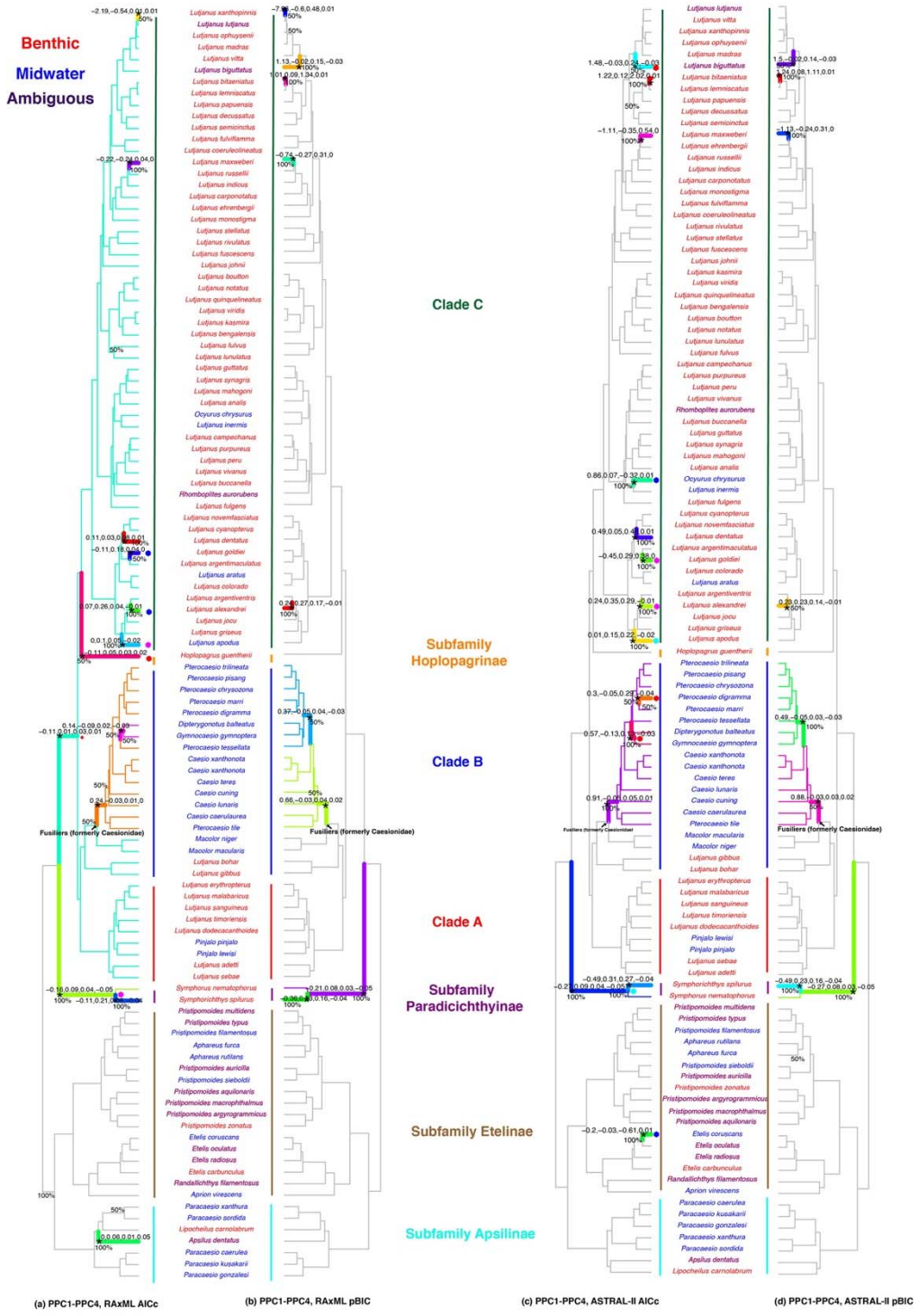


Figure S23. Adaptive and convergent regimes in Lutjanidae for the fins-only dataset (first four PC axes), with ℓ1ou using both AICc and pBIC. Stars indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum (trait optima values for each axis indicated). Colored polygons indicate convergent peaks.

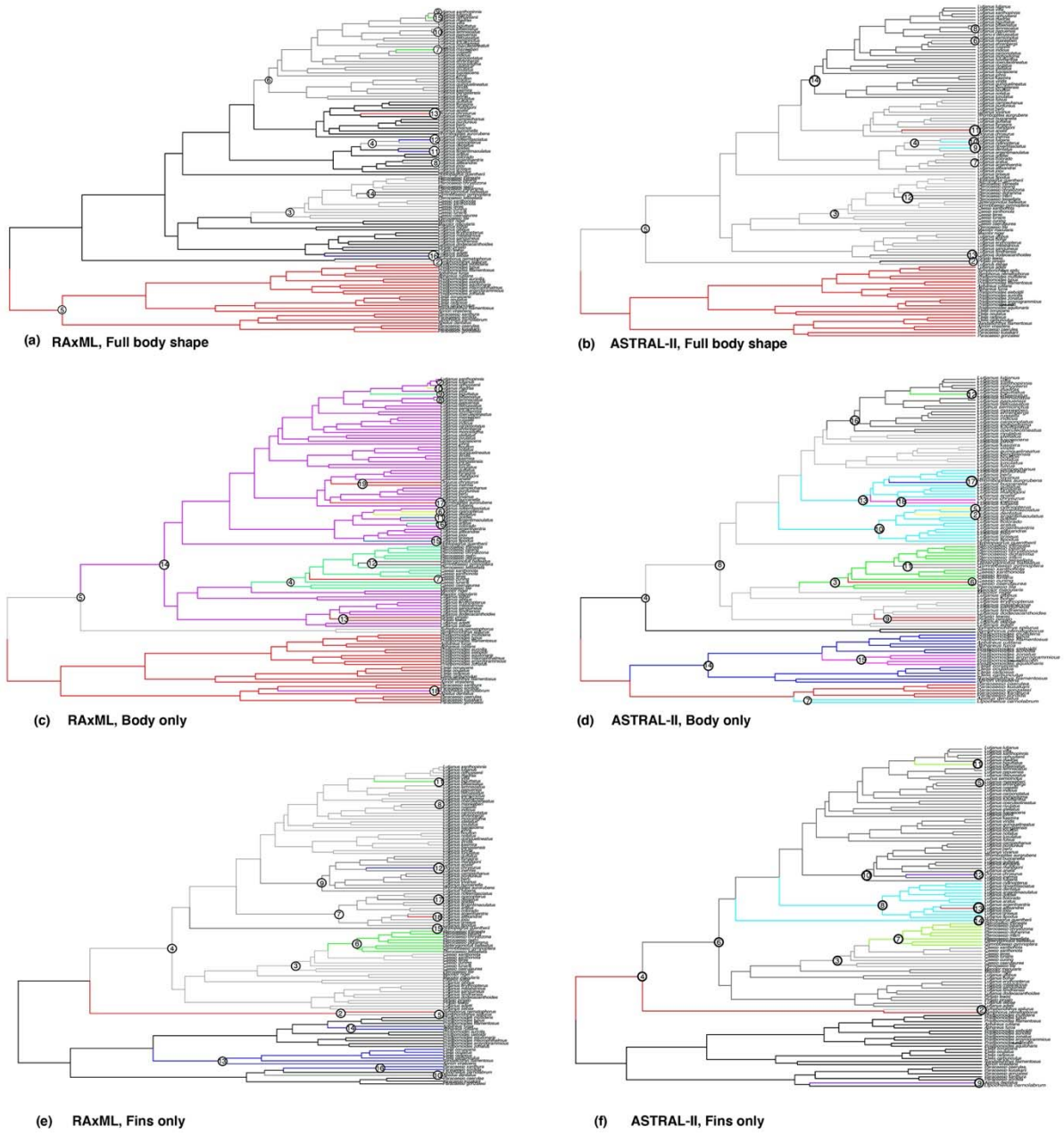


Figure S24. Adaptive and convergent shifts in Lutjanidae for the (a) full-body shape (b) body-only, and (c) fins-only datasets with SURFACE and AICc. Numbers indicate phenotypic shifts from mean trait values, and edges of the same color are inferred to have converged to the same selection optimum value.

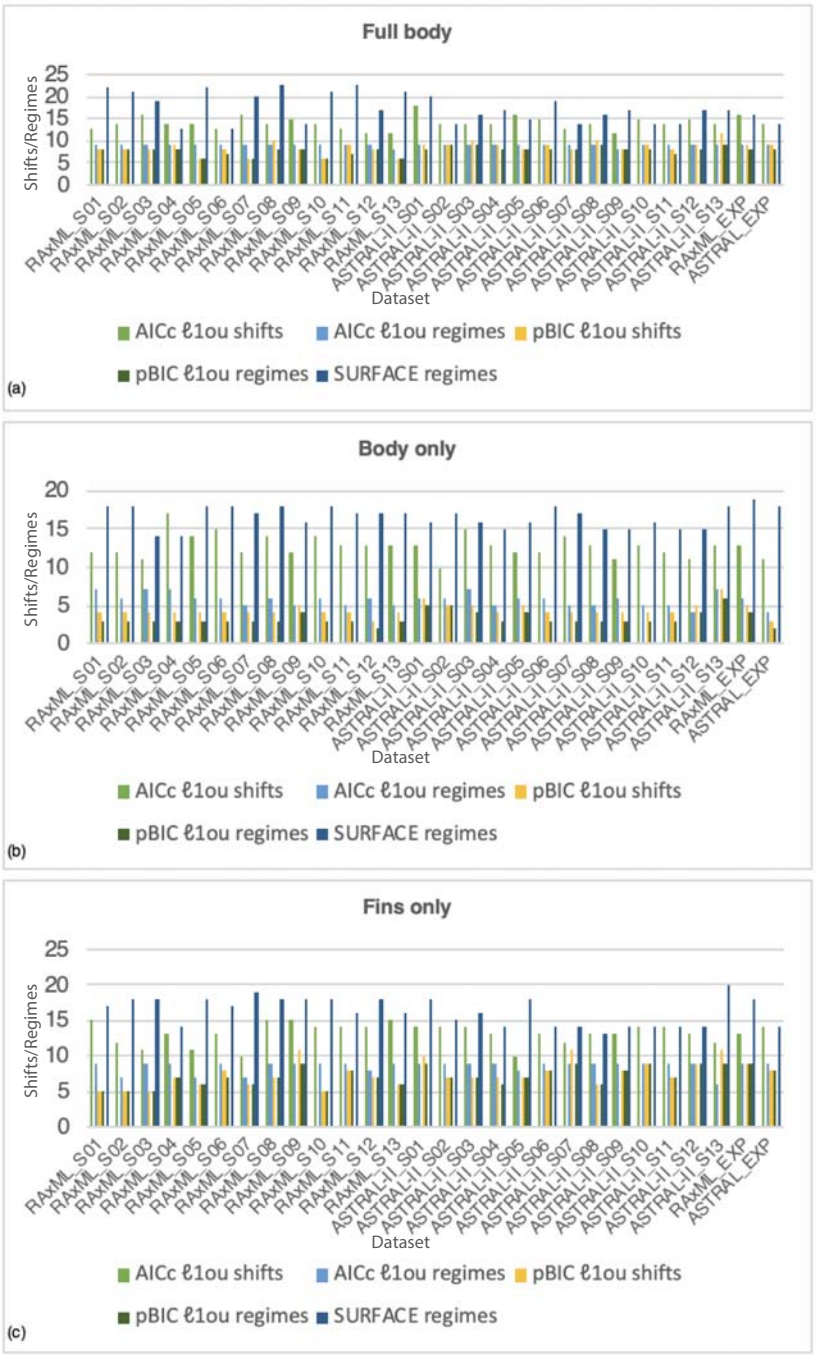


Figure S25. ℓ1ou and SURFACE results for each alternative morphometric dataset: a) full body shape, b) body only, and c) fins only. Each panel shows the number of shifts and regimes for AICc and pBIC models from ℓ1ou analyses, and the number of regimes for SURFACE analyses.

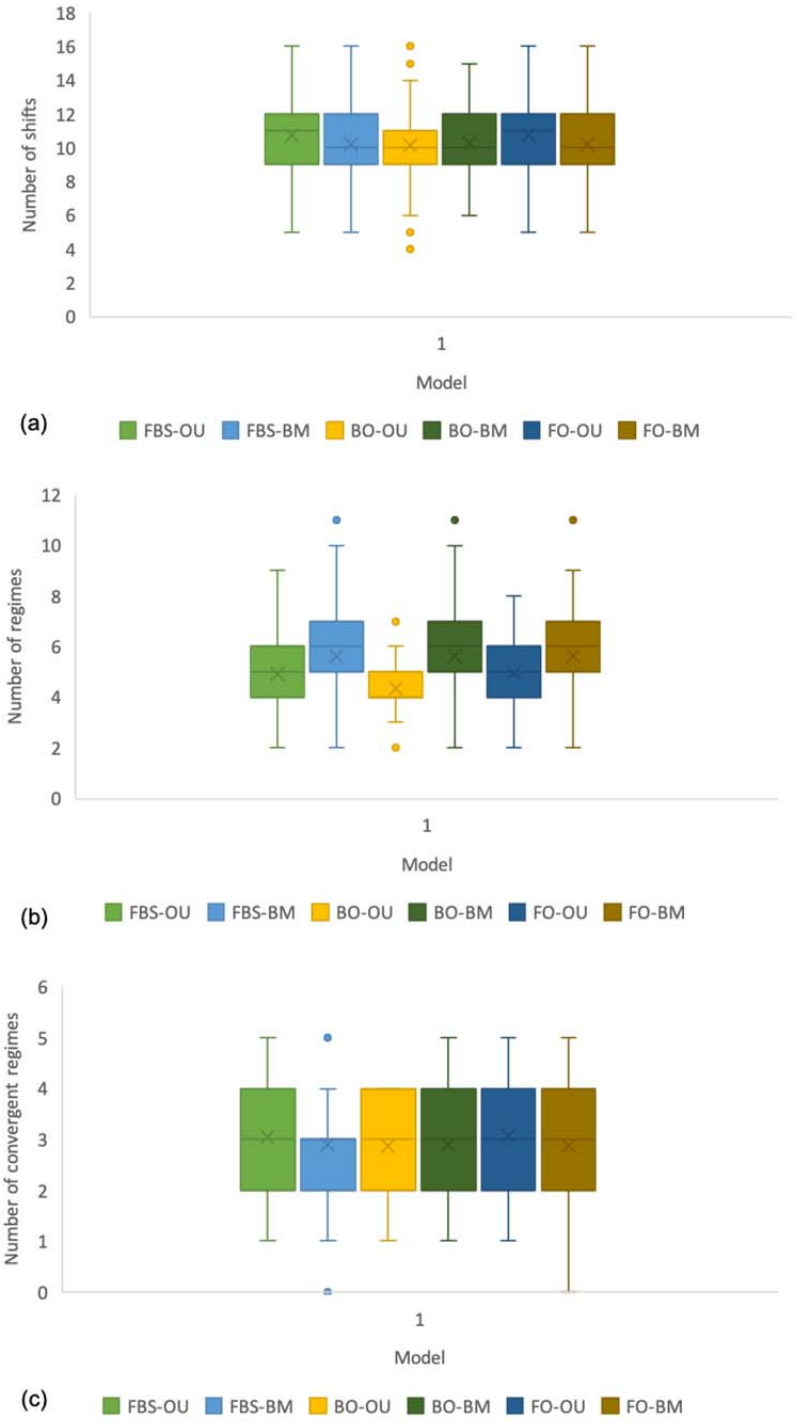


Figure S26. Number of shifts, regimes, and convergent regimes identified with the empirical dataset against 99 simulated BM and OU model distributions for single PC axes. FBS: Full body-shape; BO: Body only; FO: Fins only. a. Number of shifts; b. number of regimes; c. number of convergent regimes.

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; Fig. S27a). While this model was not decisively favored by our data and the support was shared with two alternative null models (Fig. S28b), 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 (Fig. S29, Tables S16-S18), 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 (Tables S16-S18).

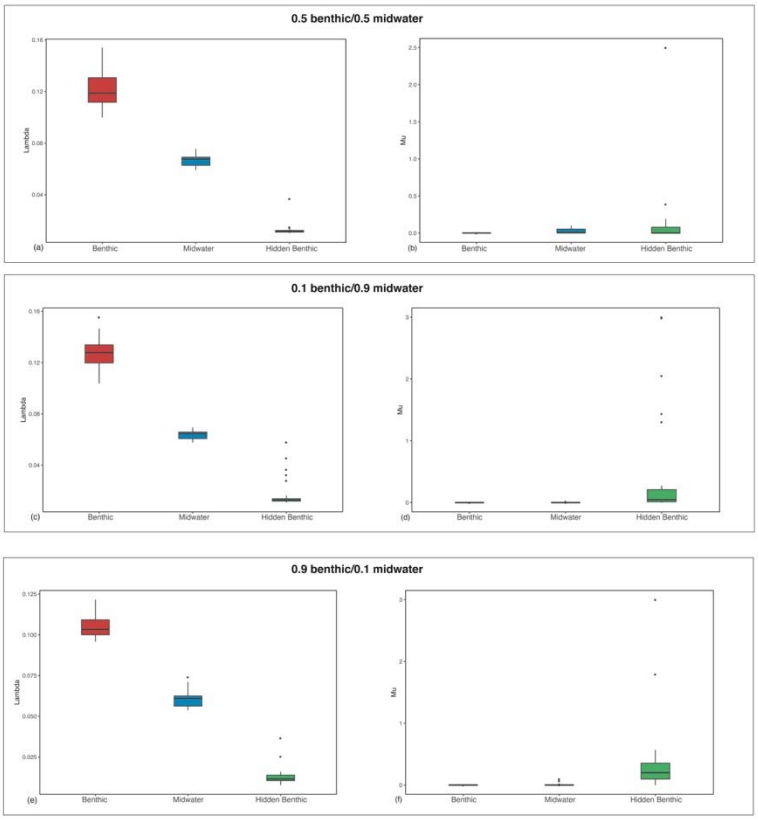


Figure S27. Distribution of diversification rates estimated under the 'HiSSE benthic' model for the 28 trees based on three alternative habitat coding schemes. (*a*) Net-diversification (speciation minus extinction), (*b*) speciation, and (*c*) extinction parameters obtained for midwater (blue) benthic (red) habitats, including a hidden state (pink) associated with benthic lineages.

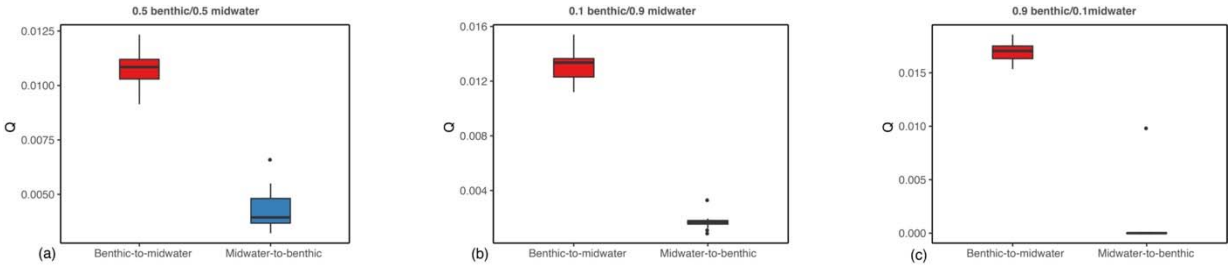


Figure S28. Distribution of transition rates between the different habitat states estimated under the 'HiSSE benthic' model for the 28 trees using the three alternative habitat coding schemes.

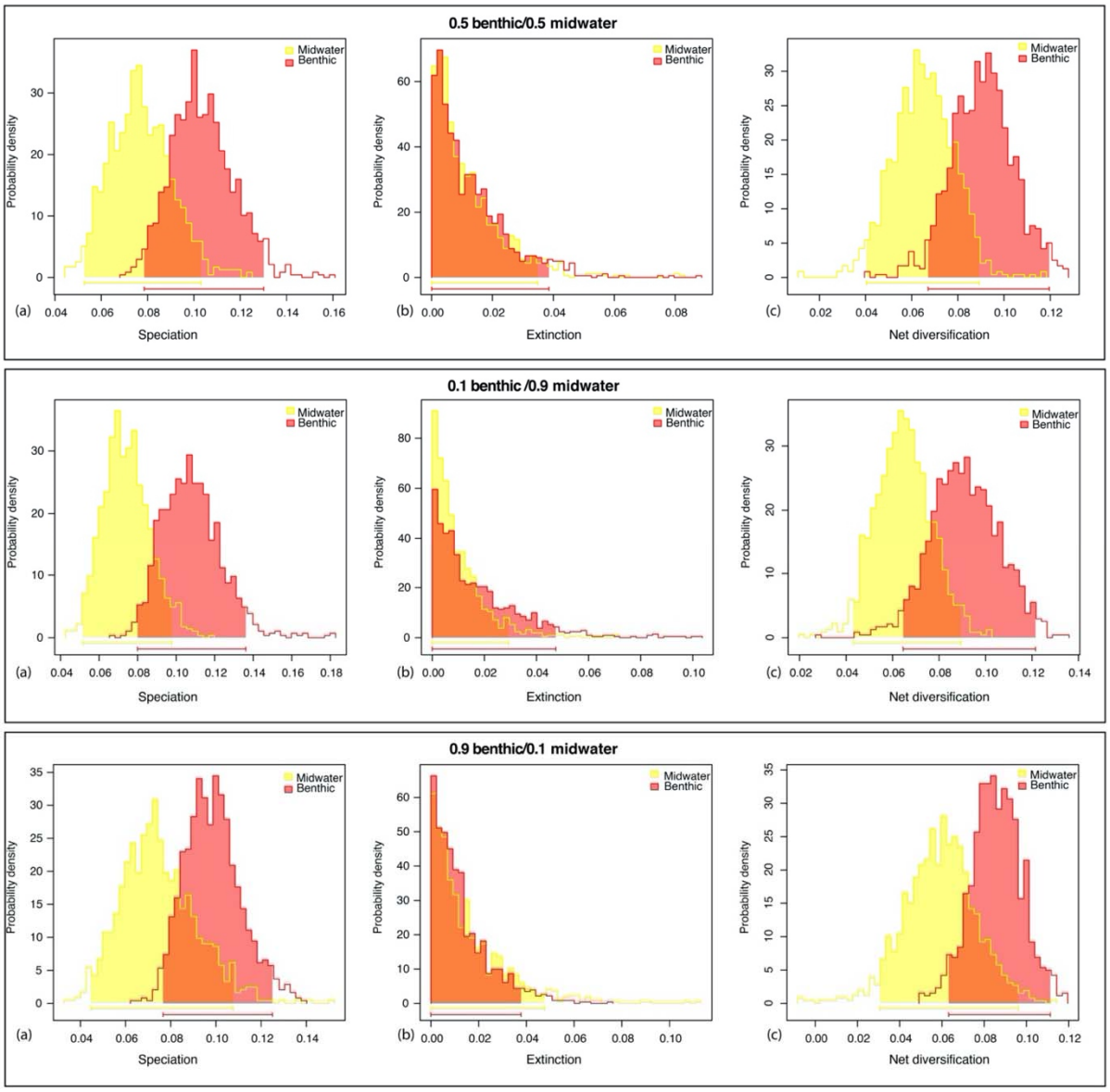


Figure S29. Marginal distribution of diversification rates obtained using MCMC-based BiSSE analyses applied to the 'master tree', based on the three alternative habitat coding schemes.

Estimated (*a*) speciation, (*b*) extinction, and (*c*) net-diversification (speciation − extinction) parameters for benthic (yellow) and midwater (red) lineages.

Table S10. Diversification rates estimated for the different habitat states under the three alternative methods using the 'master tree' and the 0.5/0.5 habitat coding scheme. λ =speciation rates; μ = extinction rates.

	Benthic (λ)	**Benthic (μ)**	**Midwater (λ)**	**Midwater (μ)**	**Hidden benthic (λ)**	**Hidden benthic (μ)**
HiSSE benthic	0.120	2.06E-09	0.075	0.068	0.011	7.71E-09
FiSSE	0.133	NA	0.085	NA	NA	NA
BiSSE MCMC (Mean)	0.114	0.015	0.075	0.013	NA	NA

Table S11. Diversification rates estimated for the different habitat states under the three alternative methods using the 'master tree' and the 0.1/0.9 habitat coding scheme. λ =speciation rates; μ = extinction rates.

	Benthic (λ)	**Benthic (μ)**	**Midwater (λ)**	**Midwater (μ)**	**Hidden benthic (λ)**	**Hidden benthic (μ)**
HiSSE benthic	0.123	2.06E-09	0.060	0.068	0.013	0.016
FiSSE	0.133	NA	0.085	NA	NA	NA
BiSSE MCMC (Mean)	0.108	0.025	0.074	0.015	NA	NA

Table S12. Diversification rates estimated for the different habitat states under the three alternative methods using the 'master tree' and the 0.9/0.1 habitat coding scheme. λ =speciation rates; μ = extinction rates.

	Benthic (λ)	**Benthic (μ)**	**Midwater (λ)**	**Midwater (μ)**	**Hidden benthic (λ)**	**Hidden benthic (μ)**
HiSSE benthic	0.099	2.06E-09	0.055	0.068	0.011	0.14
FiSSE	0.126	NA	0.088	NA	NA	NA
BiSSE MCMC (Mean)	0.101	0.018	0.074	0.024	NA	NA

Appendix I: Files available on Figshare digital repository 10.6084/m9.figshare.13000100

Legends for Datasets:

Dataset S1 (separate file). Samples used in the current study, including associated specimen, tissue number, and image source.

Dataset S2 (separate file). Habitat coding for species in Lutjanidae, including references used. A total of 18 species are interpreted as either having multi-state or uncertain habitats (B/M).

Dataset S3 (separate file). Matrix showing presence (1) or absence (0) of species based on six widely recognized marine biogeographic regions (Spalding et al., 2007; Kulbicki et al. 2013).

Dataset S4 (separate file). Information of the 13 subsets used in this study. Genes and sequence length are specified.

Dataset S5 (separate file). Age estimates for MCMCTrees.

Dataset S6 (separate file). ℓ1ou and SURFACE results.

SI References

1. C. Li, M. Hofreiter, N. Straube, S. Corrigan, G. J. P. Naylor, Capturing protein-coding genes across highly divergent species. *Biotechniques* **54**, 321–326 (2013).
2. L. C. Hughes, *et al.*, Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6249–6254 (2018).
3. L. Hughes, *et al.*, Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. *Mol. Ecol. Resour.* (2020) https:/doi.org/10.1101/2020.02.18.949735.
4. A. Smit, R. Hubley, P. Greeb, RepeatMasker Open-4.0. *http://www.repeatmasker.orh* (2013).
5. R. Betancur-R, *et al.*, The Tree of Life and a New Classification of Bony Fishes. *PLoS Curr. Tree Life* **1** (2013).
6. C. Li, R. Betancur-R., W. Leo Smith, G. Ortí, Monophyly and interrelationships of Snook and Barramundi (Centropomidae sensu Greenwood) and five new markers for fish phylogenetics. *Mol. Phylogenet. Evol.* **60**, 463–471 (2011).
7. C. Li, G. Ortí, G. Zhang, G. Lu, A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol. Biol.* **7**, 44 (2007).
8. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2214–2120 (2014).
9. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows – Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
10. C. Li, G. Lu, G. Ortí, Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol.* **57**, 519–539 (2008).
11. D. R. Zerbino, E. Birney, Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
12. J. M. Allen, *et al.*, Phylogenomics from whole genome sequences using aTRAM. *Syst. Biol.* **66**, 786–798 (2017).
13. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
14. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 1–11 (2005).

15. F. Abascal, R. Zardoya, M. J. Telford, TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, 7–13 (2010).
16. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
17. B. Meyer, K. Meusemann, B. Misof, MARE v0.1.2-rc (2011).
18. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
19. A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**, 758–771 (2008).
20. S. Mirarab, T. Warnow, ASTRAL-II : coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**, 44–52 (2015).
21. M. Dos Reis, Z. Yang, Approximate likelihood calculation on a phylogeny for Bayesian Estimation of Divergence Times. *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
22. A. Rambaut, S. Y. W. Ho, A. J. Drummond, B. Shapiro, Accommodating the effect of ancient DNA damage on inferences of demographic histories. *Mol. Biol. Evol.* (2009) https:/doi.org/10.1093/molbev/msn256.
23. J. F. Parham, *et al.*, Best practices for justifying fossil calibrations. *Syst. Biol.* **61**, 346–359 (2012).
24. L. Sorbini, *La collezione Baja di pesci e piante fossili di Bolca: con descrizione di nuovi generi e nuove specie*, M. civico di storia naturale di Verona, Ed. (University of California Press, 1983).
25. L. Agassiz, *Recherches sur les poissons fossiles* (Imprimerie de Petipierre, 1833).
26. A. F. Bannikov, The systematic composition of the Eocene actinopterygian fish fauna from Monte Bolca, northern Italy, as known to date. *Misc. Paleontol.* **12**, 22–34 (2014).
27. A. F. Bannikov, Fishes from the Eocene of Bolca, northern Italy, previously classified in the Sparidae, Serranidae and Haemulidae (Perciformes). *Geodiversitas* **28**, 249–275 (2006).
28. G. Carnevale, A. F. Bannikov, G. Marramà, J. C. Tyler, "The Pesciara-Monte Postale Fossil Lagerstatte: 2. Fishes and other vertebrates" in *The Bolca Fossil-Lagerstätten: A Window into the Eocene World*, Rendiconti della Società Paleontologica Italiana, Ed. (2014), pp. 37–63.

29. D. Johnson, *The limits and relationships of the Lutjanidae and associated families.* (Berkeley: University of California Press, 1980).
30. M. Friedman, G. Carnevale, The Bolca Lagerstätten: shallow marine life in the Eocene. *J. Geol. Soc. London.* **175**, 569–579 (2018).
31. T. J. Near, *et al.*, Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12738–12743 (2013).
32. R. R. Betancur, *et al.*, Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17** (2017).
33. B. Frédérich, F. Santini, Macroevolutionary analysis of the tempo of diversification in snappers and fusiliers (Percomorpha: Lutjanidae). *Belgian J. Zool.* **147**, 17–35 (2017).
34. M. E. Alfaro, Resolving the ray-finned fish tree of life. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6107–6109 (2018).
35. D. L. Rabosky, *et al.*, An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).
36. C. C. Swift, B. B. Ellwood, "Hypsocephalus atlanticus;: A new genus and species of lutjanid fish from marine Eocene limestones of Northern Florida" in *Contributions in Science*, L. A. C. Natural History Museum, Ed. (1972), pp. 15–30.
37. W. Pfeiffer, The morphology of the olfactory organ of *Hoplopagrus guentheri* Gill 1862. *Can. J. Zool.* **42**, 235–237 (1964).
38. H. S. Puri, R. O. Vernon, *Summary of the geology of Florida and a guidebook to the classic exposures*, Special Pu (1959).
39. D. S. Jordan, B. W. Evermann, *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*, B. of the U. S. N. Museum, Ed., i–xxx (1898).
40. A. G. Coates, J. A. Obando, "The geologic evolution of the Central American Isthmus" in *Evolution and Environment in Tropical America*, (University of Chicago Press, 1996), pp. 21–56.
41. C. Montes, *et al.*, Middle Miocene closure of the Central American Seaway. *Science (80-.).* (2015) https:/doi.org/10.1126/science.aaa2815.
42. A. O'Dea, *et al.*, Formation of the Isthmus of Panama. *Sci. Adv.* (2016) https:/doi.org/10.1126/sciadv.1600883.

43. J. P. Bollback, SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7** (2006).
44. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
45. J. Tavera, A. Acero P., P. C. Wainwright, Multilocus phylogeny, divergence times, and a major role for the benthic-to-pelagic axis in the diversification of grunts (Haemulidae). *Mol. Phylogenet. Evol.* **121**, 212–223 (2018).
46. M. Kulbicki, *et al.*, Global Biogeography of Reef Fishes: A Hierarchical Quantitative Delineation of Regions. *PLoS One* **8** (2013).
47. M. D. Spalding, *et al.*, Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *Bioscience* **57**, 573–583 (2007).
48. N. J. Matzke, BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. *R Packag. version 0.2* (2013).
49. R. H. Ree, S. A. Smith, Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* **57**, 4–14 (2008).
50. F. Ronquist, Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Syst. Biol.* **46**, 195–203 (1997).
51. M. J. Landis, N. J. Matzke, B. R. Moore, J. P. Huelsenbeck, Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* **62**, 789–804 (2013).
52. N. J. Matzke, Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Syst. Biol.* **63**, 951–970 (2014).
53. J. Dupin, *et al.*, Bayesian estimation of the global biogeographical history of the Solanaceae. *J. Biogeogr.* **44**, 887–899 (2017).
54. F. Steininger, F. Rögl, The paratethys history. A contribution towards the Neogene geodynamics of the alpine orogene. *Ann. Geol. des Pays Hell* **3**, 1153–1165 (1979).
55. L. A. Rocha, *et al.*, Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Mol. Ecol.* **14**, 3921–3928 (2005).
56. D. R. Bellwood, P. C. Wainwright, "The History and Biogeography of Fishes on Coral Reefs" in *Coral Reef Fishes*, Peter F. Sale, Ed. (Elsevier Inc., 2002), p. 549.
57. H. A. Lessios, D. R. Robertson, Crossing the impassable: Genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proc. R. Soc. B Biol. Sci.* **273**, 2201–2208 (2006).

58. A. B. George, M. W. Westneat, Functional morphology of endurance swimming performance and gait transition strategies in balistoid fishes. *J. Exp. Biol.* **222** (2019).
59. E. G. Drucker, J. A. Walker, M. W. Westneat, "Fish Physiology 23: Mechanics of Pectoral Fin Swimming in Fishes" in *Fish Biomechanics*, R. E. S. George V Lauder, Ed. (Academic Press, 2005), pp. 369–423.
60. D. Adams, M. Collyer, A. Kaliontzopoulou, Geomorph: Software for geometric morphometric analyses. R package version 3.1.0 (2019).
61. L. J. Revell, Size-correction and principal components for interspecific comparative studies. *Evolution (N. Y).* **63**, 3258–3268 (2009).
62. D. A. Jackson, No Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecol. Ecol. Soc. Am.* **74**, 2204–2214 (1993).
63. P. R. Peres-Neto, D. A. Jackson, K. M. Somers, How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput. Stat. Data Anal.* **49**, 974–997 (2005).
64. J. Clavel, G. Escarguel, G. Merceron, mvMORPH: An R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015).
65. A. M. Davis, R. Betancur-R, Widespread ecomorphological convergence in multiple fish families spanning the marine–freshwater interface. *Proc. R. Soc. B Biol. Sci.* **284** (2017).
66. J. Felsenstein, A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. *Am. Nat.* **179**, 154–56 (2012).
67. C. T. Stayton, The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution (N. Y).* **69**, 2140–2153 (2015).
68. K. Arbuckle, C. M. Bennett, M. P. Speed, A simple measure of the strength of convergent evolution. *Methods Ecol. Evol.* **5**, 685–693 (2014).
69. M. Khabbazian, R. Kriebel, K. Rohe, C. Ané, Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* **7**, 811–824 (2016).
70. T. Ingram, D. L. Mahler, SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol. Evol.* **4**, 416–425 (2013).
71. R. G. Fitzjohn, Diversitree: Comparative phylogenetic analyses of diversification in R.

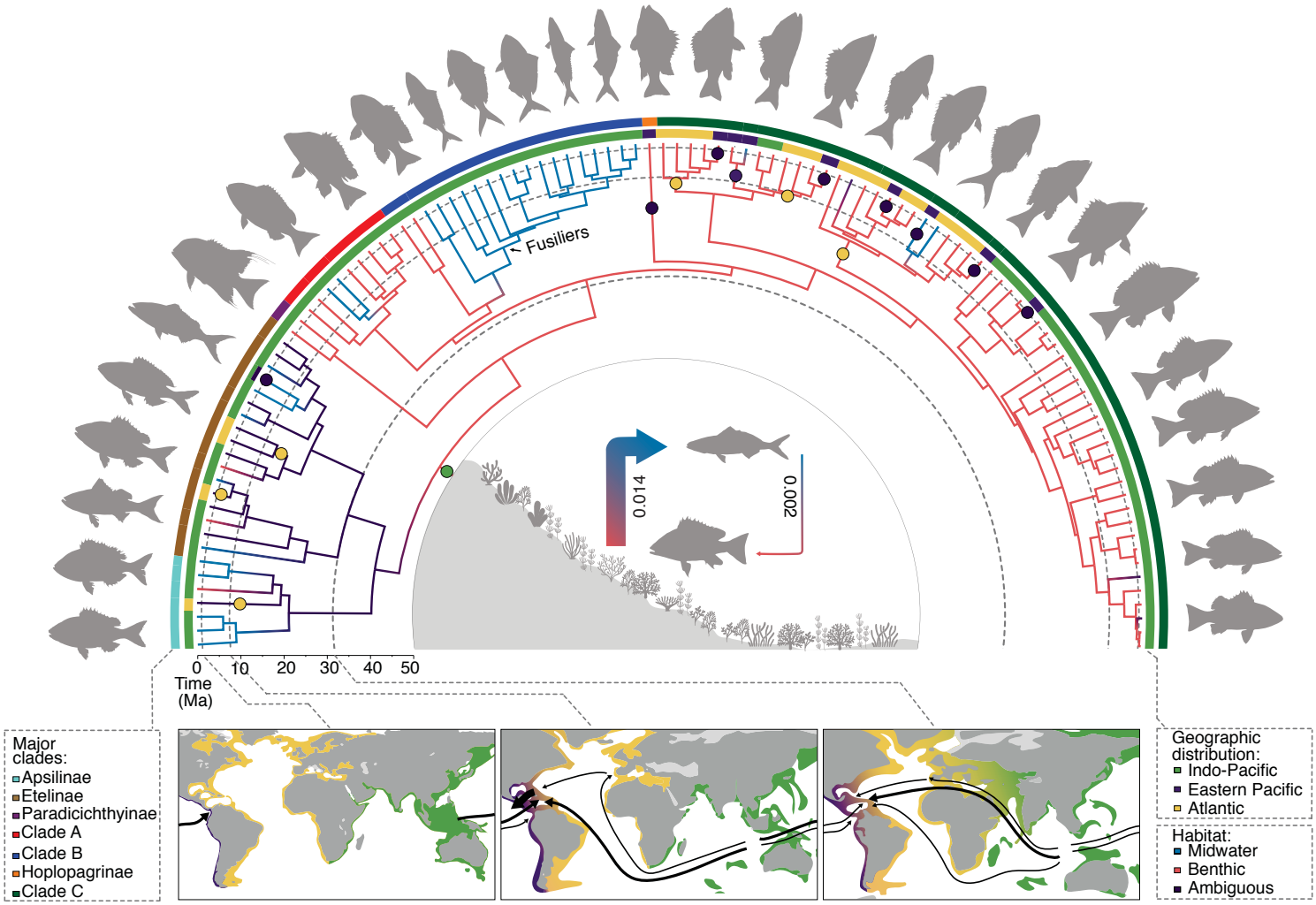
Methods Ecol. Evol. **3**, 1084–1092 (2012).

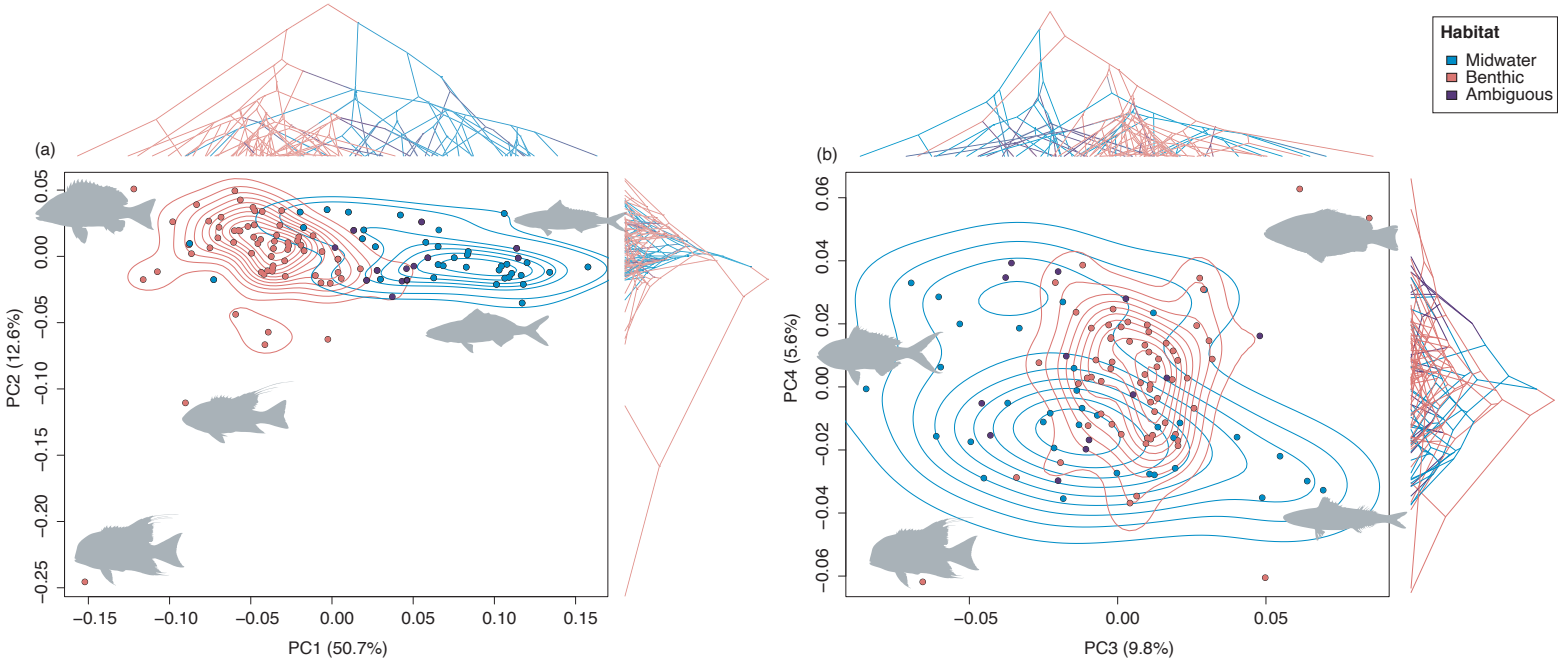
72. J. M. Beaulieu, B. C. O'Meara, Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Syst. Biol.* **65**, 583–601 (2016).
73. D. L. Rabosky, E. E. Goldberg, FiSSE: A simple nonparametric test for the effects of a binary character on lineage diversification rates. *Evolution (N. Y).* **71**, 1432–1442 (2017).
74. J. R. Gold, G. Voelker, M. A. Renshaw, Phylogenetic relationships of tropical western Atlantic snappers in subfamily Lutjaninae (Lutjanidae: Perciformes) inferred from mitochondrial DNA sequences. *Biol. J. Linn. Soc.* **102**, 915–929 (2011).
75. W. Anderson, H. T. Kami, G. D. Johnson, A new genus of Pacific Etelinae (Pisces: Lutjanidae) with redescription of the type-species. *Proc. Biol. Soc. Washingt.* **90**, 89–98 (1976).
76. K. E. Carpenter, A Phylogenetic Analysis of the Caesionidae (Perciformes: Lutjanoidea). *Copeia*, 692–717 (1990).
77. D. Johnson, Percomorph Phylogeny - Progress and Problems. *Bull. Mar. Sci.* **52**, 3–28 (1993).
78. R. A. Fritzsche, G. D. Johnson, "Limits and relationships of the Lutjanidae and associated families" (1981).
79. J. Nelson, T. Grande, M. Wilson, *Fishes of the World*, John Wiley & Sons, Ed., 5th Ed. (2016).
80. R. van der Laan, W. N. Eschmeyer, R. Fricke, Family-group names of Recent fishes. *Zootaxa* **3882**, 1–230 (2014).
81. J. Felsenstein, Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 1427–1434 (2005).
82. J. C. Uyeda, D. S. Caetano, M. W. Pennell, Comparative Analysis of Principal Components Can be Misleading. *Syst. Biol.* **64**, 677–689 (2015).
83. K. E. Carpenter, *Revision of the Indo-Pacific Fish Family Caesionidae (Lutjanoidea), with Descriptions of Five New Species* (Bishop Museum Press, 1987).
84. K. E. Carpenter, "Volume 5: Bony fishes part 3 (Menidae to Pomacentridae)" in *The Living Marine Resources of the Western Central Pacific*, FAO, (2001), pp. 2919–2941.
85. K. E. Carpenter, "Vol.8. Fusiliers fishes of the world" in *FAO Species Catalogue*, (1988).
86. K. E. Carpenter, Optimal cladistic and quantitative evolutionary classifications as

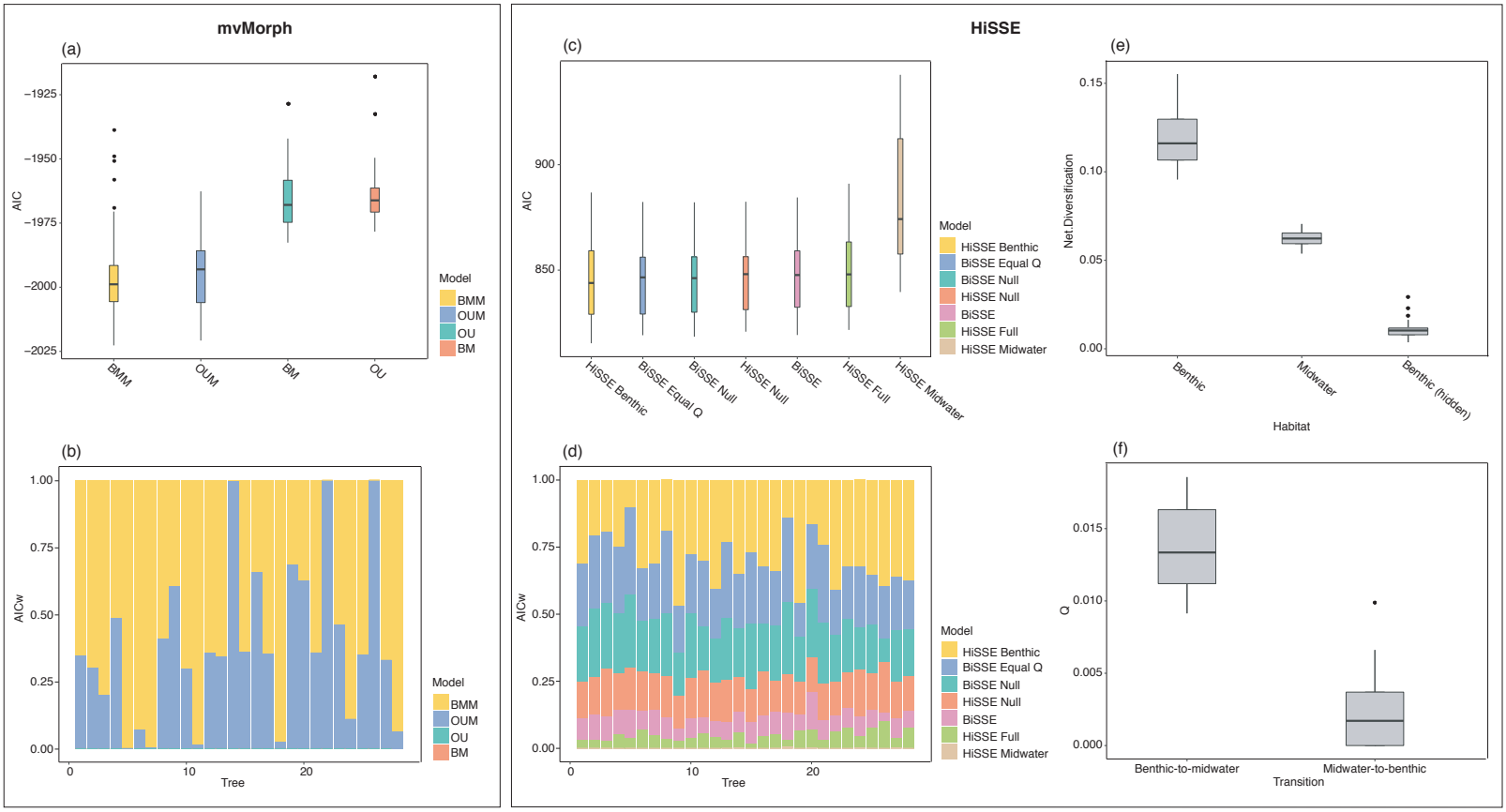
illustrated by fusilier fishes (Teleostei: Caesionidae). *Syst Biol* **42**, 142–154 (1993).

87. R. Fricke, W. N. Eschmeyer, R. van der Laan, Eschmeyer's Catalog of Fishes: Genera, Species, References (2020) (September 20, 2019).

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


Fusiliers
0.014
0.002
0
10
20
30
40
50
Time
(Ma)
Major clades:
Apsilinae
Etelinae
Paradicichthyinae
Clade A
Clade B
Hoplopagrinae
Clade C
Geographic distribution:
Indo-Pacific
Eastern Pacific
Atlantic
Habitat:
Midwater
Benthic
Ambiguous


(a)
(b)
Habitat
Midwater
Benthic
Ambiguous
PC1 (50.7%)
PC2 (12.6%)
PC3 (9.8%)
PC4 (5.6%)
−0.15
−0.10
−0.05
0.00
0.05
0.10
0.15
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06


mvMorph
(a)
AIC
−1925
−1950
−1975
−2000
−2025
BMM
OUM
BM
OU
Model
BMM
OUM
OU
BM
(b)
AICw
1.00
0.75
0.50
0.25
0.00
0
10
20
Tree
Model
BMM
OUM
OU
BM
HiSSE
(c)
AIC
900
850
HiSSE Benthic
BiSSE Equal Q
BiSSE Null
HiSSE Null
BiSSE
HiSSE Full
HiSSE Midwater
Model
HiSSE Benthic
BiSSE Equal Q
BiSSE Null
HiSSE Null
BiSSE
HiSSE Full
HiSSE Midwater
(d)
AICw
1.00
0.75
0.50
0.25
0.00
0
10
20
Tree
Model
HiSSE Benthic
BiSSE Equal Q
BiSSE Null
HiSSE Null
BiSSE
HiSSE Full
HiSSE Midwater
(e)
Net.Diversification
0.15
0.10
0.05
0.00
Benthic
Midwater
Benthic (hidden)
Habitat
(f)
Q
0.015
0.010
0.005
0.000
Benthic-to-midwater
Midwater-to-benthic
Transition