Phylogeny of the damselfishes (Pomacentridae) and patterns of asymmetrical diversification in body size and feeding ecology

Charlene L. McCord [a], W. James Cooper [b], Chloe M. Nash [c, d] & Mark W. Westneat [c, d]

[a] California State University Dominguez Hills, College of Natural and Behavioral Sciences, 1000 E. Victoria Street, Carson, CA 90747
[b] Western Washington University, Department of Biology and Program in Marine and Coastal Science, 516 High Street, Bellingham, WA 98225
[c] University of Chicago, Department of Organismal Biology and Anatomy, and Committee on Evolutionary Biology, 1027 E. 57th St, Chicago IL, 60637, USA
[d] Field Museum of Natural History, Division of Fishes, 1400 S. Lake Shore Dr., Chicago, IL 60605

Corresponding author: Mark W. Westneat mwestneat@uchicago.edu

Journal: PLoS One

Keywords: Pomacentridae, phylogenetics, body size, diversification, evolution, ecotype

Abstract

The damselfishes (family Pomacentridae) inhabit near-shore communities in tropical and temperature oceans as one of the major lineages with ecological and economic importance for coral reef fish assemblages. Our understanding of their evolutionary ecology, morphology and function has often been advanced by increasingly detailed and accurate molecular phylogenies. Here we present the next stage of multi-locus, molecular phylogenetics for the group based on analysis of 12 nuclear and mitochondrial gene sequences from 330 of the 422 damselfish species. The resulting well-resolved phylogeny helps to address several important questions about higher-level damselfish relationships and the monophyly of genera, including *Chromis*, *Chrysiptera*, *Parma* and *Stegastes*. A time-calibrated phylogenetic tree scaled using fossil data and recent estimated ovalentarian clade ages, yields an older root age for the family (55.3 mya) than previously proposed, refines the age of origin for a number of diverse genera, and shows that ecological changes during the Eocene-Oligocene transition provided opportunities for damselfish diversification. We explored the idea that body size extremes have evolved repeatedly among the Pomacentridae, and demonstrate that large and small body sizes have evolved independently at least 30 times and with asymmetric rates of transition. We tested the hypothesis that transitions among three dietary ecotypes (benthic herbivory, pelagic planktivory and intermediate omnivory) are asymmetric, with higher transition rates from intermediate omnivory to either planktivory or herbivory. Using multistate hidden-state speciation and extinction models, we found that dietary ecotype is significantly associated with patterns of diversification across the damselfishes, and that the highest rates of net diversification are associated with pelagic planktivory. We also conclude that the pattern of evolutionary diversification in feeding ecology, with frequent and asymmetrical transitions between a small number of feeding ecotypes, is largely restricted to the subfamily Pomacentrinae in the Indo-West Pacific.

1. Introduction

Advances in the richness and accuracy of the Tree of Life are critically important for understanding broad evolutionary patterns and processes, as well as exploring the unique features of specific groups of organisms that capture our attention. It is a central challenge of systematics to resolve phylogenies that include large numbers of species and include previously unexamined taxa in well-supported trees, thus improving our ability to test evolutionary hypotheses and examine patterns of diversification. In recent years, we have gained an increasingly clear picture of the phylogenetic relationships among coral reef fishes through the pursuit of densely sampled species phylogenies for diverse reef fish families (1–13). By combining species-rich phylogenies with rich data layers of functional and ecological traits, significant advances have been made in our understanding of the diversification of fish lineages in coral reef ecosystems (14–22). The project presented here aims to enhance the species richness present in the phylogeny of the reef fish family Pomacentridae and examine the evolutionary history of key structural and ecological traits of the damselfishes, with a focus on the transitions in body size and dietary ecotype throughout pomacentrid diversification.

The damselfishes (Pomacentridae) are a major component of the global coral reef and temperate rocky reef ichthyofauna (23,24) and their systematic study has an unusually deep and rich history. Descriptions of damselfishes are found in publications that laid the foundations of taxonomy such as Linnæus's Systema Naturae (25), the beginnings of ichthyology with Artedi's *Ichthyologia* (26), and even the entirety of western biology with Artistotle's *History of Animals*, (27). Damselfishes were studied by several of the most prominent naturalists of the 19th century including Lacépède (28) and Cuvier (29), and the type specimen of one damselfish genus (*Stegastes*) was collected by Charles Darwin early in the voyage of *H.M.S. Beagle* (30). The past 60 years have produced a massive surge in marine research, including the publication of several hundred studies of pomacentrid fishes. Advances in molecular biology and computational ability have facilitated increasingly comprehensive molecular phylogenetic studies of damselfish relationships (1,15,31–36). These phylogenies have been used in multiple phylogenetic comparative studies that have examined the evolution of their functional morphology and ecological diversity (14,15,37–40).

The Pomacentridae range in size over an order of magnitude from the small *Chrysiptera giti* (4.5 cm) up to the giant *Microspathodon dorsalis* (45 cm), and occupy marine habitats from shallow coastal waters down to 200m depths. The circumtropical distribution of damselfishes is centered on marine coral and rocky reef habitats, with a variety of temperate species occurring at up to 50° North and South latitude. Members of the family are known for intriguing behaviors such as strong territorial aggression, and complex farming or gardening behaviors in which dense stands of filamentous algae are tended for food (41). Damselfishes have been placed in broad ecological-functional categories ("ecotypes") using combinations of traits such as diet (planktivory, herbivory, and omnivory) and primary feeding location in the water column (benthic, pelagic, intermediate) that are useful for exploring the evolution of damselfish ecomorphology (14,15). Detailed analyses of functional morphology have shown that mandible morphology and jaw protrusion ability are strong predictors of damselfish ecotype (14,15,38) and that trophic specialization is tightly tied to habitat and social behavior (37). Evolutionary-developmental studies of damselfishes have also provided insight into the ontogenetic changes associated with shifts in adult feeding ecology (42–45).

A noteworthy evolutionary pattern among damselfishes is that dietary ecotype has been shown to be unevenly distributed across the Pomacentridae, with transition rates from the

intermediate to specialized dietary ecotypes shown to be higher than in the opposite direction (37). There are frequent switches back and forth between ecotype states within some damselfish groups creating a strong pattern of convergence (14), although there is a trend in some groups to become fixed on one of the specialist ecotypes, creating asymmetrical patterns of ecotype diversification among clades. However, previous work has not found strong evidence that differences in diversification rates among feeding strategies support the evolutionary "dead-end" hypothesis, in which specialization results in reduced speciation and elevated extinction rates (46,47). Given that the asymmetry in state distribution could be the result of variation in transition rates or differences in diversification rates associated with each state, it is important to jointly estimate these two processes (48). Recent advances in the hidden-state speciation and extinction modeling framework allow us to robustly examine this variation in diversification rates and its potential association with dietary ecotype using multistate characters (49).

In this study we present advances in our understanding of damselfish phylogenetic relationships based on molecular sequence data from 330 of the 422 extant pomacentrid species (50). We use the time-calibrated phylogeny to examine evolutionary rates of transition across several important ecological traits to test hypotheses about the complex relationship between trait evolution and the tempo of damselfish diversification, and examine the origins of major clades across space and time. Our central aims were to generate a new species-rich phylogenetic framework, use that framework to provide new estimates for the divergence times of multiple damselfish clades, consider the historical timing of pomacentrid diversification, reconstruct the evolution of pomacentrid body size and feeding ecology, and explore the evolutionary patterns associated with the strikingly asymmetrical distributions of benthic, pelagic and intermediate ecotypes across the damselfish phylogeny.

2. Methods

2.1 Species sampling and genes sequenced

DNA sequence data for 335 fish species were analyzed in this study, including 330 species of damselfishes and 5 outgroup taxa from two additional families in the Ovalentaria: Embiotocidae and Cichlidae (See Table S1 for a list of all taxa and genes examined, with large data sets contributed by past authors color coded to provide clear attribution). At least one member of each damselfish genus was included in our phylogenetic analyses. Specimens sequenced for the present study (80 species including 10 not previously sequenced) were collected by the authors, purchased at local fish markets, or borrowed from outside institutions and most are associated with a voucher specimen (Table S1). We analyzed portions of twelve genes totaling 8238 bases from seven mitochondrial regions: 12s (901 bp), 16s (559 bp), ND3 (431 bp), ATP (842 bp), COI (651 bp), Cytb (1140 bp), and D-loop (439); and five nuclear loci: RAG1 (900 bp), RAG2 (800 bp), DLX2 (482 bp), Tmo4C4 (417 bp), and BMP4 (548 bp). Most nuclear sequences and three mitochondrial regions (12s, 16s and ND3) were sequenced at the Pritzker Laboratory of the Field Museum of Natural History.

Sequence data for four additional mitochondrial loci (ATP, COI, CytB and D-loop) were downloaded from GenBank and included in the concatenated supermatrix (Table S1). In some cases, sequences were excluded or replaced due to evidence from BLAST alignment or data conflict that there were identification errors or contamination. For example, prior data for *Parma oligolepis* 12S and 16S that placed it as sister to *Stegastes fasciolatus* in previous studies was determined to be contaminated or misidentified, and was replaced with new data that unites all *Parma* species as a monophyletic group. Similarly, Genbank and BOLD sequences for

Pomacentrus agassizzi all appear to be *Stegastes fasciolatus* at the time of this writing, and *Chromis dasygenys* barcode data was excluded, as it appears to be misidentified *Neopomacentrus*. These excluded sequences are listed in Table S1.

For new gene sequences, DNA sequencing protocols were similar to those described in previous phylogenetics studies from our laboratory (1,12,13). Briefly, DNA was extracted from muscle or gill tissue according to the PureGene animal tissue DNA isolation procedure (Gentra System). Double-stranded DNA products were amplified using polymerase chain reaction from aliquots of genomic DNA isolates of the 12S, 16S, ND3, tmo4c4, rag1, rag2, bmp4 and DLX2 gene fragments. Primers used for DNA amplification and sequencing are listed in Supplemental Table 2. PCR reactions were performed using an MJ Research PTC-200 Peltier Thermal Cycler (MJ Research Inc., Watertown, MA). PCR products were cycle-sequenced by creating a solution of 3.3 ul of buffer, 0.5 ul of primer, 0.7 ul of Big Dye reagent, 0.5 ul of PCR product template DNA and 5.0 ul nuclease free water to bring the total reaction volume to 10 ul. Thermal cycling protocol included one initial denaturation at 96° C followed by 25 cycles of denaturation at 96° C for 10 s, annealing at 50° C for 5 s and extension at 60° C for 4 min. All products were purified and sequences were generated on an ABI PRISM 3730 Genetic Analyzer (Applied Biosystem, Foster City, CA). All raw sequence data have been deposited in GenBank (see Table S1 for accession numbers).

Sequences for 12s and 16s were manually aligned to secondary structure models for ribosomal sequences using previously published homologous sequences for damselfishes and outgroups (1) in Mesquite version 3.51 (51). Protein coding genes were automatically aligned using the ClustalW or Muscle modules within Mesquite (52,53). Some sequences were trimmed to the size of the smallest fragment to minimize the amount of missing data in the data matrix. After excluding unalignable regions and gaps, the character count for the entire concatenated supermatrix was 8238 nucleotide characters.

2.2 Phylogenetic analysis

The best fit models of nucleotide evolution for damselfish gene sequences were selected using PartitionFinder2 (54) and maximum likelihood analyses were performed using Garli2 (55). An initial set of eight parallel (mpi) analyses (three replicates each; 24 runs) was performed using a random starting tree. The topology with the highest likelihood from this first run was used as the starting tree for a second set of 24 simultaneous maximum likelihood analyses. Termination condition settings for all analyses were (1) number of generations without topology improvement (genthreshfortopoterm) = 50000, (2) max score improvement over recent generations required for termination (scorethreshforterm) = 0.05, and required score improvement for topology to be considered better (significanttopochange) = 0.01.

Bayesian phylogenetic analyses of the concatenated DNA supermatrix were conducted using MrBayes (56) on an Asus GL12CM gaming PC (Intel i7-8700K, NVIDIA GTX 1080 GPU). We performed two independent runs of a Markov Chain Monte Carlo analysis with four chains (eight chains total) running in parallel for 50 million generations. Trees were saved every 1000 generations for each run and used default priors for the transition/transversion rate ratio, branch length, alpha parameter of the gamma distribution for rate heterogeneity, proportion of invariant sites, base frequencies and tree topology parameters. To help ensure that stationarity was reached, the first 10 million generations (10000 sampled trees) were discarded from each run as burn-in and the remaining 40 million generations (45000 sampled trees for each of the four runs) were used in subsequent analysis. To determine the posterior probability of clades, a

majority-rule consensus tree calculated from the post burn-in trees was constructed. The majority-rule consensus tree and maximum posterior probability phylogenetic trees were then exported and viewed using FigTree version 1.3.1 (57) and the ggtree package (58) in R version 3.3.1 (R Core Team).

2.3 Divergence time estimation

To explore the timing of damselfish divergence events we calibrated two nodes in the pomacentrid tree using fossil evidence. *Palaeopomacentrus orphae* and *Lorenzichthys olihan* from the Ypresian were used to determine the minimum age of the crown pomacentrids (59). *Chromis savornini* from the late Miocene 6.5 Ma (60) was used to date the origin of the genus *Chromis.* Due to the likely young age of *Chromis savornini*, we assigned a soft upper bound to this calibration point (48.5 Ma, age of the oldest crown pomacentrid). Outside the family Pomacentridae, we also set a prior age estimate of 96 mya for the most recent common ancestor of cichlids, embiotocids, and pomacentrids within the Ovalenteria (61). To estimate divergence times, we used a relaxed clock of log normal distributed rates in BEAST 2.6.3 (62). A standard birth-death rate model was used to estimate rates of cladogenesis. We partitioned the twelve genes used in these analyses such that each region was modeled separately under its best substitution model. Tracer v1.6 was used to assess convergence following two independent analyses of 100 million generations each. The first 25% of generations were used as burn-in and the effective sample size (ESS) for all model parameters were checked for good mixing of the final Markov chain Monte Carlo (all ESS exceeded 200 in our analyses).

2.4 Damselfish trait matrix assembly and analysis

Data for body size (body length), body depth (as % body length), dietary ecotype (benthic, pelagic, or intermediate), and presence of algal farming behavior were gathered from prior work (14,15), original species descriptions, or mined from FishBase using the rFishBase package in R (63). Continuous data for total body length were binned into discrete states, for example 3 or 5 size classes using gap coding (64) (Table S3). Size classes in the 3-state character were defined as small (S = 4.5-10.2 cm TL), medium (M = 10.5-14.1 cm TL) and large (L = 14.4-45cm TL), and the 5-state character defined by subdividing S into XS (4.6-6.5 cm) and S (6.9-10.2 cm) and subdividing L into L (14.4-20.5 cm) and XL (21-45 cm). Using the `ace` function in the R package ape (65), ancestral character states were estimated across the phylogeny using three types of transition rate models, all rates different (ARD), symmetrical transitions (SYM), and equal rates (ER). ARD was selected for all traits with the exception of the 3-size class trait, in which the SYM model was best supported. Additionally, given that the simultaneous loss and gain of a complex trait, such as dietary ecotype, without passing through an intermediate state may be rare or impossible, we constructed a set of "constrained" transition models. This assumption is supported by previous work which found no direct transitions between the benthic and pelagic states (37). These "constrained" state transition models require transitions through the intermediate state, which was modeled as Benthic→Pelagic = 0 and Pelagic→Benthic = 0.

Tip states were mapped onto the time-calibrated phylogeny, and the estimated likelihoods of each ancestral state were mapped onto internal nodes within the phylogeny, allowing visualization of independent trait origins and convergence across clades. We tested for strength of phylogenetic signal for continuous traits (body size, body depth) by computing Blomberg's K and Pagel's lambda using the `phylosig` function in the R package Phytools (65). Statistical

support was computed using 1000 randomizations. For the discrete trait ecotype, we conducted a discrete lambda test using likelihood (lambda = 0 vs estimated lambda) using the `fitDiscrete` function in the R package Geiger (66). We examined trait associations across the phylogeny, to test the association of dietary ecotype and farming behavior with body size and body slenderness (body depth/length). To perform these tests, we used a phylogenetic ANOVA with 10,000 simulations, computed using the `phylANOVA` function in Phytools.

2.5 Analysis of Ecotype Dependent Diversification using MuHiSSE

We used a multistate, hidden-state speciation and extinction (MuHiSSE) model framework (67) to evaluate the association of transitions among dietary ecotypes with rates of diversification across the damselfish phylogeny. For this analysis, each discrete ecotype was coded as the presence or absence of two binary states: benthic (01), pelagic (10) or intermediate (11). For consistency, we used the "constrained" transition model for all SSE analyses. We designed 13 models (Table 1) to test specific hypotheses about asymmetric rates of diversification and transitions associated with ecotype. These models fall into five distinct categories: 1) "trivial null" model, which assumes a single rate of diversification across phylogeny independent of character state, 2) state dependent diversification, which assumes diversification parameters are shared among lineages with the same ecotype (MuSSE), 3) state dependent diversification with hidden state(s), in which the diversification rates are shared among ecotype in combination with a hidden (non-focal) state (MuHiSSE), 4) state absorbing, in which transitions out of a particular state were constrained to zero (i.e. Benthic → Intermediate = 0) and 5) character independent diversification, which assumes diversification parameters are independent of ecotype but allowed to vary over increasing numbers of hidden states (MuCID models; Table 1). Given the high prevalence of recovering a "false positive" when using SSE models, MuCID models can act as the proper null model due to the partitioning of rate variation across hidden states that are not attributable to the focal state (67,68).

Net turnover (defined as speciation + extinction) and transitions among states were either constrained or allowed to vary depending on the model (Table 1). The extinction fraction (extinction/speciation) was held constant for all states across all models due to issues in estimating extinction in extant only phylogenies. The reparameterization of net turnover and extinction fraction alleviates problems associated with over-fitting when speciation and extinction are highly correlated, but both matter in explaining the diversity pattern (68). To account for sampling bias among character states, we used the following sampling fraction, which corresponds to the percentage of taxa in our tree with a particular state over the total number of taxa with that trait: Pelagic = 0.68, Benthic = 0.72, Intermediate = 0.76.

Following Nakov et al. (49) we used 100 random starting points to account for the initial run not recovering the MLE for the optimization of each model. Models were optimized using the MuHiSSE function in the hisse package in R (67,69) on the midway2 computing cluster through the Research Computing Center at the University of Chicago. The parameters from the optimization iteration with the highest likelihood for each model were ranked according to AICc. Diversification and transition parameters were estimated using a model averaging approach through weighted AIC (ωAIC), which reduces the subjectivity of choosing the thresholds for model choice and accounts for the uncertainty around parameter estimates across hidden states (70). This was done by averaging each estimated rate of diversification per-lineage using the weighted probability of occupying an alternative hidden state (49). We reconstructed the marginal ancestral states according to ωAIC using the MarginReconMuHiSSE function and

approximated the likelihood surface around the maximum likelihood estimates with an adaptive sampling method using the SupportRegionMuHiSSE function, in the hisse package (67) and figures were made using the utilhisse package in R (49). All R scripts included in Supplement.

Table 1. Ranked MuHiSSE results in descending order of support for tested models.

model	Description	FP	lik	aic	daic	aicw
MuHiSSE	State dependent with a hidden state	16	-1139.75	2313.54	0	0.691
MuHiSSE Relaxed	Relaxed state dependent with hidden state	17	-1139.45	2315.18	1.646	0.303
Benthic Absorbing	No transitions out of benthic state	15	-1145.74	2323.26	9.72	0.005
CID2	Character-independent 2 hidden states	9	-1161.74	2342.13	28.59	0
CID3	Character-independent 3 hidden states	14	-1159.95	2349.46	35.92	0
Pelagic Absorbing	No transitions out of pelagic state	15	-1161.20	2354.18	40.64	0
CID4	Character-independent 4 hidden states	21	-1159.10	2363.88	50.34	0
CID5	Character-independent 5 hidden states	29	-1159.00	2382.81	69.26	0
CID6	Character-independent 6 hidden states	41	-1158.94	2414.00	100.46	0
CID7	Character-independent 7 hidden states	54	-1158.88	2451.47	137.93	0
CID8	Character-independent 8 hidden states	69	-1158.69	2500.11	186.57	0
MuSSE	State dependent diversification	7	-1267.58	2551.69	238.15	0
Dull	Equal rates	6	-1271.35	2555.01	241.47	0

3. Results

Phylogenetic analyses revealed a strongly supported monophyletic Pomacentridae that originated in the Lower Eocene (55.3 mya) with the five modern subfamilies (71) arising during the mid-Eocene to upper Oligocene, from about 50-30 mya (Fig. 1 and 2). Major clades of damselfishes were recovered with high support and topologies similar to those reported in prior studies, yet with some significant revisions of relationships within several clades. A central result of this study is that deep branching events within Pomacentridae were followed by a remarkably steady rate of clade diversification throughout the past 50 million years. Body size shows weak phylogenetic signal, with multiple instances of convergence on both large and small body size across the phylogeny. Repeated patterns of evolutionary change in feeding ecology among the three main ecotypes shows that patterns of species diversification are associated with an imbalanced or asymmetric frequency of transition between states. A frequent pattern of back-and-forth transitions among ecotypes is primarily restricted to the crown group Pomacentrinae.

3.1 Pomacentrid phylogenetics

Phylogenetic analyses resolved a monophyletic Pomacentridae with five well-resolved and strongly supported damselfish subfamilies (Fig. 1, 2), supporting recent subfamily taxonomy of the family (71). The same overall damselfish topology and subfamily structure well-supported in both ML and Bayesian analyses. The sister group to the rest of the damselfishes is the subfamily Microspathodontinae, which comprises three well-supported (*pp* = 1.0) clades. The sister-group to the rest of the Microspathodontinae contains the monotypic genus *Mecaenichthys* and the now monophyletic group of four *Parma* species present in the tree (Fig. 1). Moving

upward from the base, the second major clade of the Microspathodontinae contains the giant damselfishes in the monophyletic tribe Microspathodontini, which is composed of the genera *Microspathadon* (monophyletic), *Nexilosus* (monotypic), *Hypsypops* (monotypic) and *Similiparma* (71). *Nexilosus* is represented by COI barcode sequence only, and its position is often unstable within this clade. The Microspathodontini are sister to a clade that splits into a *Plectroglyphidodon* branch and a *Stegastes* branch (*Stegastes*1 in Fig. 1). Both of these genera are polyphyletic since *Plectrophidodon lacrymatus* and the majority of the *Stegastes* species we examined are found within the *Stegastes*2 and *Stegastes*3 clades on the third major branch of the Microspathodontinae. The labeled *Stegastes*1 and *Stegastes*2 groups contain all known *Stegastes* species that are endemic to the Indo-West Pacific. The *Stegastes*3 clade is composed entirely of Atlantic and Eastern Pacific species. *Lepidozygus tapeinosoma* is found at the base of this third major branch as the sister group to *Stegastes*2 (Fig. 1). The location of *Lepidozygus tapeinosoma* nested among the Microspathodontinae renders this taxon polyphyletic. *Lepidozygus* occupies the longest monotypic branch in our trees (Fig. 1). Multiple placements for *Lepidozygus* have been recovered during independent, large-scale molecular phylogenetic analyses of the Pomacentridae and long-branch attraction make a confident placement of this species difficult.

The damselfish subfamily Chrominae contains three well-resolved major clades (Fig. 1). The sister group to the rest of Chrominae, (Fig 1: Chrominae1) is a clade composed of *Chromis flavaxilla, C. ternatensis, C. brevirostris, C. atripectoralis* and *C. viridis.* Moving up the chromine backbone to the Chrominae 2 clade (Fig. 1), the genus *Chromis* is rendered paraphyletic by *Azurina hirundo* (sister to *Chromis multilineata*) and by the genus *Dascyllus*, nested within Chrominae 2 (Fig. 1). The monophyly of *Dascyllus* is strongly supported and this genus forms the sister group to a clade that includes *C. acares, C. vanderbilti, C. delta* and a group of small *Chromis* species closely related to *C. atripes.* The largest clade of Chrominae is strongly supported as monophyletic and contains 47 of the *Chromis* species we examined (Fig. 1: Chrominae 3). Chrominae 3 divides into two branches, neither of which has strong support nor well-resolved internal relationships. Many recently described species of *Chromis*, often with COI barcodes as the only source of data currently available, are placed in Chrominae 3.

The monophyletic *Abudefduf* is the subfamily Glyphisodontinae, with *Abudefduf declivifrons, A. concolor* and *A. taurus* sister to the rest of the clade (Fig. 2). Moving distally from the base of *Abudefduf,* the next clade is composed of *A. conformis*, *A. septemfasciatus*, *A. notatus* and *A. sordidus*. A third clade forms a nearly ladder-like topology of 14 species from the basal node of *A. whitleyi* to the crown sister-pair of *A. caudobimaculatus* + *A. vaigiensis* (Fig. 2).

The Pomacentrinae form the largest damselfish subfamily (Fig. 2). This strongly supported clade of 16 genera contains 218 species, more than half of the known damselfishes (50) and represents the most diverse lineage in the family. The Pomacentrinae contains the genera *Acanthochromis*, *Altrichthys*, *Amblygliphidodon*, *Amblypomacentrus*, *Amphiprion*, *Cheiloprion*, *Chrysiptera*, *Dischistodus*, *Hemiglyphidodon*, *Neoglyphidodon*, *Neopomacentrus*, *Pomacentrus, Pomachromis*, *Premnas*, *Pristotis* and *Teixeirichthy.* These taxa are resolved into five major clades. Lack of monophyly is found in *Chrysiptera*, *Neopomacentrus and Pristotis*, all polyphyletic (Fig. 2), while the placement of *Premnas* renders *Amphiprion* paraphyletic (Fig. 2).

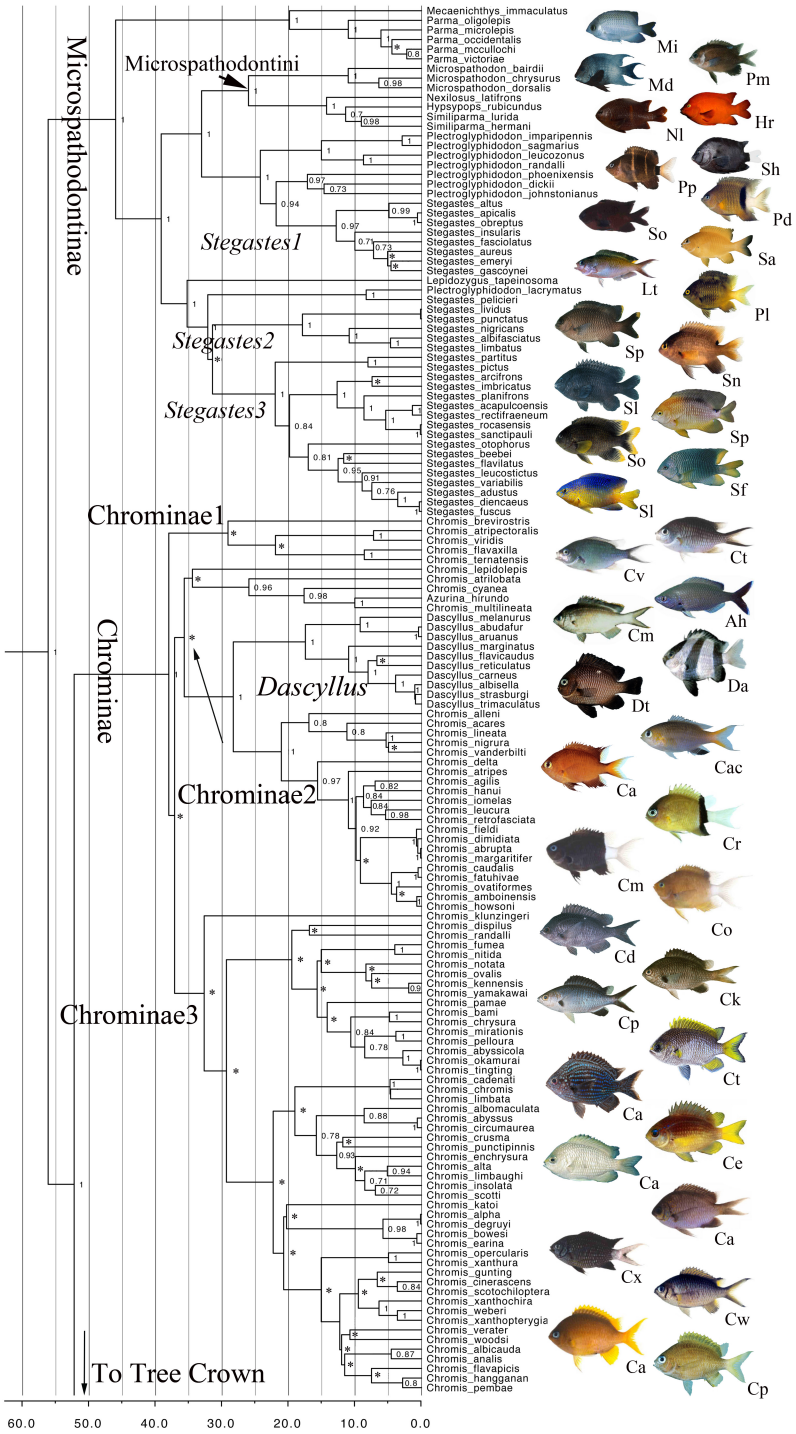

Figure 1. Time-calibrated phylogenetic tree of the Pomacentridae. Lower half of the tree including the subfamilies Microspathodontinae and Chrominae. Time axis in millions of years before present. Nodal values are Bayesian posterior support levels above 0.7, with * indicating lower posterior support. Representative photos labeled with first letter of genus and species, matching species names in the tree closest to the photo.

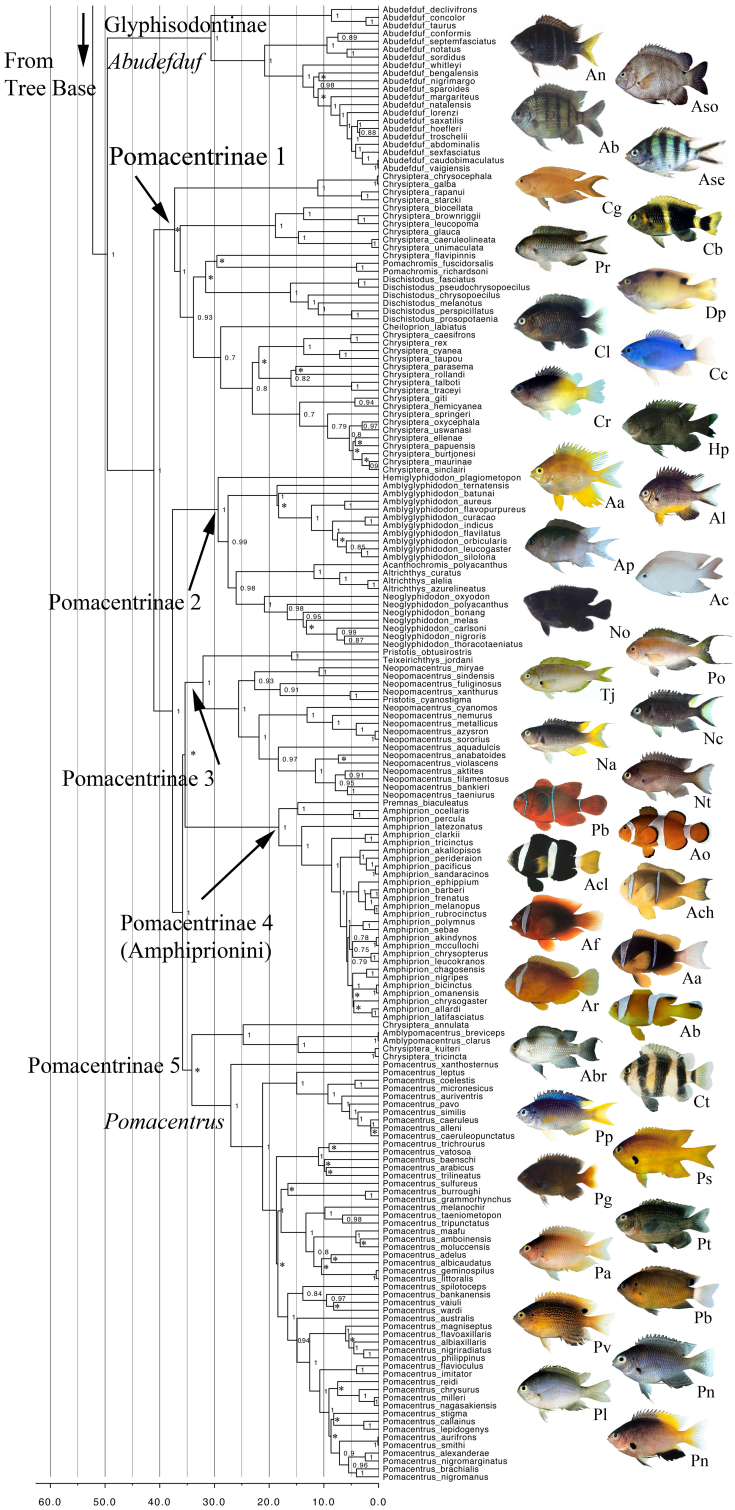

Figure 2. Time-calibrated phylogenetic tree of the Pomacentridae. Upper (crown) half of the tree including the subfamilies Glyphysodontinae and Pomacentrinae. Time axis in millions of years before present. Nodal values are Bayesian posterior support levels above 0.7, with * indicating lower posterior support. Representative photos labeled with first letter of genus and species, matching species names in the tree closest to the photo.

The sister-clade to the rest of the subfamily (Pomacentrinae 1) is an assemblage of most, but not all, *Chrysiptera* species with two lineages of species from other genera nested among them. One of these contains the genera *Pomachromis* and *Dischistodus* as two monophyletic sister groups. The other is composed of *Cheiloprion labiatus*, an unusual coral polyp specialist that is the only member of its genus. Pomacentrinae 2 is composed of five monophyletic genera. The monotypic *Hemiglyphidodon* (*H. plagiometopon*) branches from the basal node, and the next lineage is the monophyletic genus *Amblyglyphidon*, followed by *Acanthochromis* (*A. polyacanthus*) + *Altrichthys* (monophyletic), and the monophyletic *Neoglyphidodon.* These relationships have strong support with just a few nodes of low support within genera.

Pomacentrinae 3 and 4 (Fig. 2) are sister clades, with low support, with group 3 containing *Pristotis, Teixeirichthys and Neopomacentrus*, and group 4 the Amphiprionini. Relationships within Pomacentrinae 3 are strongly supported, with the sister-pair *Pristotis obtusirostris* + *Teixeirichthys jordani* forming the sister lineage to *Neopomacentrus*, which has *Pristotis cyanostigma* inserted within it, as the sister to *Neopomacentrus xanthurus*. Pomacentrinae 4 (Fig. 2) is the anemonefish tribe Amphiprionini. This tribe is composed of 29 species in the genus *Amphiprion* and *Premnas biaculeatus*. *Amphiprion* is rendered paraphyletic by *Premnas biaculeatus*, which along with the sister species *A. ocellaris* and *A. percula* forms the sister group to the rest of the tribe.

Pomacentrinae 5 (Fig. 2) contains the monophyletic genus *Pomacentrus* plus *Amblypomacentrus* and three problematic *Chrysiptera* species. The sister-group to *Pomacentrus* is *Amblypomacentrus breviceps*, *A. clarus, plus Chrysiptera annulata, C. kuiteri, and C. tricincta.* These relationships are all well-supported. The *Amblypomacentrus*/*Chrysiptera* lineage is sister to a monophyletic *Pomacentrus* clade that contains 53 species of the 77 presently known (50). With the exception of two weak nodes, the backbone of the *Pomacentrus* phylogeny inferred here is well-resolved. *Pomacentrus xanthosternus* is resolved as sister to the rest of the genus, with successive clades suggesting likely group memberships for a large number of recently described *Pomacentrus* represented by small amounts of sequence data.

3.2 Dates of origin and divergence times for damselfishes

Time-calibrated phylogenetic analysis of the present data set (see Fig S2 for all node ages) suggests that the damselfishes split from embiotocids approximately 75 mya, with the root node of the first family divergence for the Pomacentridae at 55.3 mya. The Microspathodontinae diverged from other pomacentrids at this time and underwent important vicariance events at 46, 39, 35, 30 and 10 mya (Fig. 1). The tribe Microspathodontini (the giant damselfishes) began to diversify ~10.5 mya (Fig. 1). The Chrominae originated 52 mya and extant lineages began to diverge ~38 mya (Fig. 1). Divergence dates for Chrominae 1, 2 and 3 are approximately 29, 36, and 29 mya, respectively. *Dascyllus* arose ~29 mya and existing clades began to diversify ~17.5 mya (Fig. 1). Chrominae 3 bifurcated ~29 mya and continued to diversify at a relatively constant rate throughout the Miocene. The Glyphisodontinae (*Abudefduf*) split from the Pomacentrinae ~49 mya and the radiation of living *Abudefduf* began in the late Oligocene (~30 mya). Most of the species we examined arose during the past 10 million years (Fig. 2).

The initial divergence of the Pomacentrinae and origin of the Pomacentrinae 1 clade occurred ~41 mya (Fig. 2), with the group 1 diversifying at a relatively steady rate since its origin. Pomacentrinae 2 and 5 arose approximately 37.7 and 34.5 mya, respectively, and Pomacentrinae 3 and 4 diverged from each other ~35 mya. The *Pomachromis-Dischistodus* clade arose within Pomacentrinae 1 ~31 mya, and *Cheiloprion* arose within Pomacentrinae 1 ~28 mya.

(Fig. 2). Within Pomacentrinae 2 the major clades diverged at the following estimated times: *Hemiglyphidodon* (24.8 mya); *Amblyglyphidodon* (26 mya); *Acanthachromis* + *Altrichthys* (arose 24 mya; diverged into monophyletic genera 10 mya); and *Neoglyphidodon* (24 mya).

Pomacentrinae 3 (*Neopomacentrus*, *Pristotis* and *Teixirichthys*) split from Pomacentrinae 4 (the Amphiprionini) ~35 mya (Fig. 2). This clade diverged steadily over the past 25 my, with the Amphiprionini being relatively young, having diverged from a common ancestor ~18 mya and most species arising only 3-5 mya. Pomacentrinae 5 diverged ~35 mya, with diversification of *Pomacentrus* into 77 known species over the past ~25 million years accounting for nearly one fifth of the damselfishes and a large portion of the Indo-West Pacific coral-reef fish fauna.

3.3 Trait evolution in damselfishes: body size, dietary ecotype and farming behavior

Ancestral state estimation of body size (Fig. 3) shows that large body size (including both L and XL) has evolved repeatedly, up to 31 times, throughout the tree (usually from medium-sized relatives) and that small body size has undergone a similar number of independent origins or losses across the topology. Patterns of transition between body size states are strongly asymmetrical (Fig. 3). The giant damselfishes of the tribe Microspathodontini and the Glyphisodontinae (genus *Abudefduf*) exhibit concentrations of large species, while the various lineages of *Chromis*, *Chrysiptera* and *Pomacentrus* show a preponderance of small species. The discretized 5-state character for body length shown in Fig. 3 and body depth (as a proportion of length) both show low phylogenetic signal, significantly different from the expectation of Brownian motion across the damselfish tree. For body length, Blomberg's K was 0.29 ($p = 0.001$) and Pagel's lambda was 0.86 ($p < 0.001$) reflecting the distribution of small, medium and large taxa dispersed fairly evenly among most clades. Body depth ranged from about 27% SL in slender species such as *Azurina* to 66% in deeper bodied forms such as *Amblyglyphidodon*, with $K= 0.16$ ($p = 0.001$) and $\lambda=0.87$ ($p < 0.001$). This indicates that close relatives were significantly more variable in both body depth and length than would be expected under Brownian motion.

Analysis of dietary ecotype (Fig. 4: benthic grazer, intermediate omnivore, pelagic planktivore) reveals the frequency and asymmetry of transitions between dietary states, with intermediate generalists transitioning to specialist ecotypes with high frequency, particularly among the Pomacentrinae. The Microspathodontinae mostly share a benthic ecotype and there are several other clusters of benthic habits distributed around the tree. Most Chrominae specialize in pelagic feeding (Fig. 4) and this feeding ecotype has evolved repeatedly in damselfish history in large clades (e.g., *Amphiprion* and *Neopomacentrus*) as well as more sparsely in single species, for example in *Pomacentrus* (Fig 4). Ecotype also showed low to moderate levels of phylogenetic signal ($\lambda=0.79$, significantly lower than null expectation).

Algal farming behavior, illustrated in a mirror tree with ecotype (Fig. 5), is restricted to shallow, benthic-feeding species, and is absent from the Chrominae and from *Abudefduf*. Farming evolved early in damselfish history within the Microspathodontinae as a likely single origin with multiple losses, then again much later, either once or twice in *Chrysiptera*, 3 times in single species of Pomacentrinae 2, and up to 6 times in the genus *Pomacentrus*.

We tested pairs of traits for character correlations across the entire phylogeny, such as body size and body depth associated with ecotype, or farming, and none of these tests were significant, suggesting that body size and body slenderness evolve independently of dietary ecotype. However, we found that transitions between dietary states are prevalent at higher levels in the tree, with complex patterns of evolution and reversal, often involving the intermediate ecotype state among the Glyphisodontinae and Pomacentrinae.

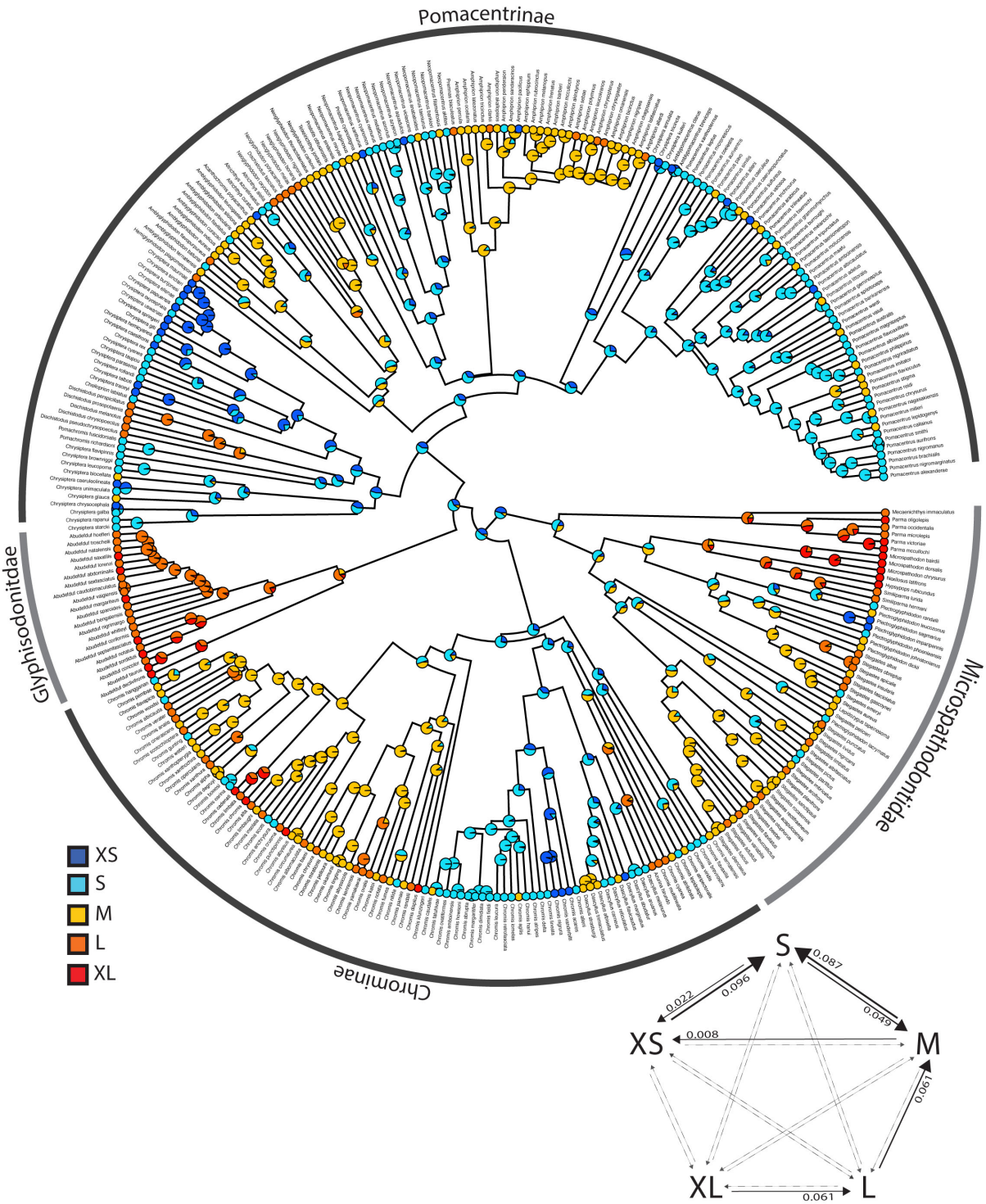

Figure 3. Ancestral state reconstruction of damselfish body size. Maximum body size (total length) gap coded into 5 states with the discretized trait mapped onto the damselfish time-calibrated phylogeny. Transition rates between character states and their rate (if non-zero) between body size states during species diversification throughout the history of the Pomacentridae is shown in lower right. XS = extra-small (4.5-6.5 cm TL), S = small (6.9-10.2 cm TL), M = medium (10.5-14.1 cm TL), L= large (14.4-20.5 cm TL), XL= extra-large (21-45 cm TL).

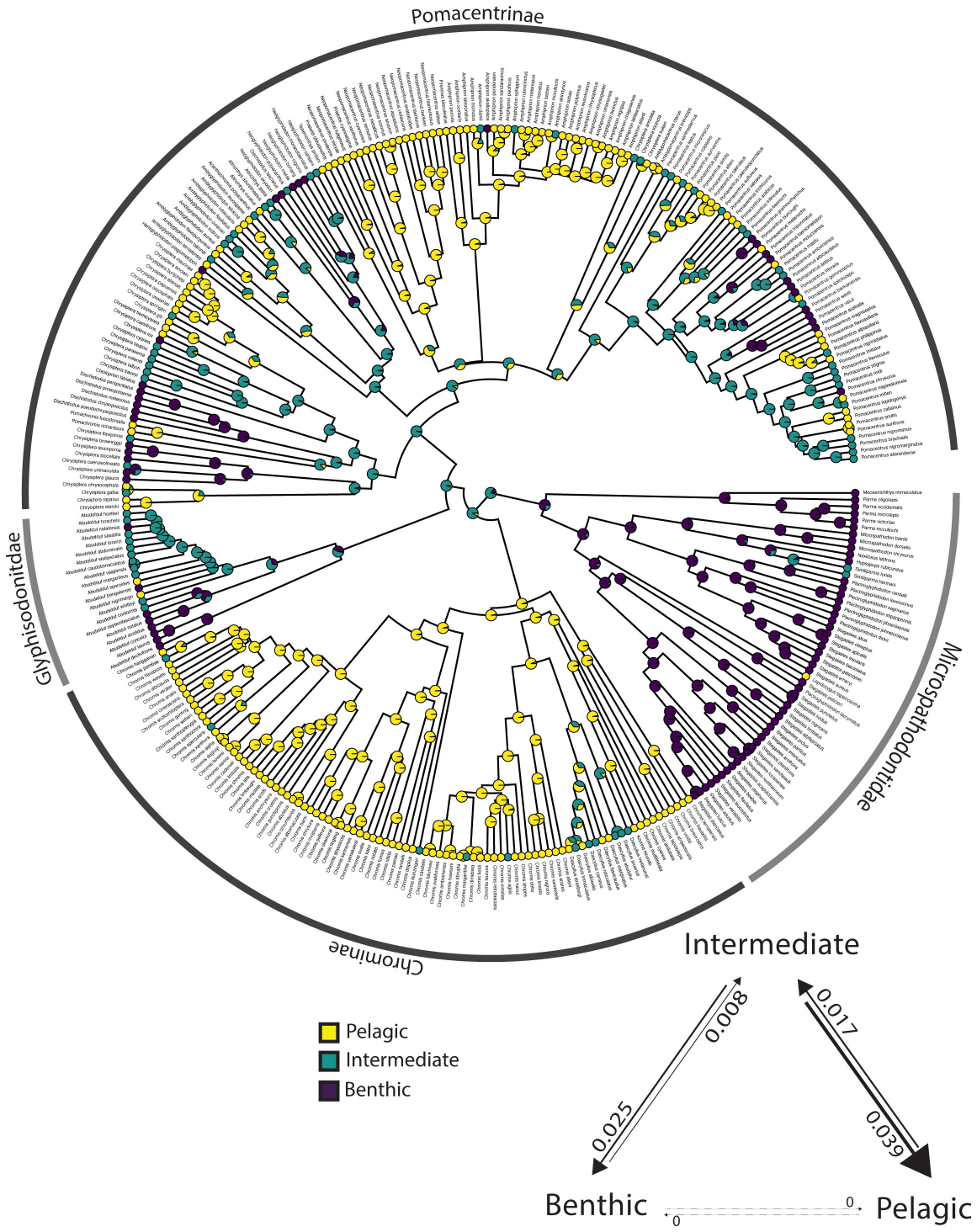

Figure 4. Ancestral state reconstruction of damselfish dietary ecotype. Dietary ecotype traits of benthic, pelagic and intermediate foraging behavior are mapped onto the damselfish time-calibrated phylogeny. Transition rates between character states and their rate (if non-zero) between ecotype states during species diversification throughout the history of the Pomacentridae is shown in lower right.

3.4 Diversification patterns and dietary ecotype

The pattern of lineage diversification for the Pomacentridae is that of a highly steady lineage through time plot (Fig. S4), with non-significant difference from a constant rate model when accounting for sampling fraction and without sampling fraction (p=0.73 and p = 0.11, respectively) using the gamma statistic of Pybus and Harvey (72). The results of MuHiSSE models of trait effects on diversification show that ecotype is associated with diversification rate across the phylogeny (Table 1, Fig. 6). Transitions between states for ecotype and body size are highly asymmetrical (Figs. 3, 4) often being much more frequent in one direction than the other. We ran the ecotype transition model under both the assumption that it was unconstrained (all pathways equally likely) and that it was constrained to pass through the intermediate state, with similar results. Ecotype transitions are highly asymmetrical, with the highest estimated transition rate from intermediate-to-pelagic and the lowest was benthic-to-intermediate (Fig. 7).

Ecotype-dependent models that include hidden states were strongly favored, with the best supported model being MuHiSSE (ωAIC = 0.691), followed by MuHiSSE relaxed and Benthic Absorbing (ωAIC = 0.303, ωAIC = 0.005, respectively; Table 1). These models differ in the parameterization of transitions among hidden states and transitions out of the benthic state. All other models had ωAIC values > 1%. When averaged across all models and over hidden state reconstructions, mean net turnover and net diversification were estimated to be highest in the pelagic state, followed by intermediate and then the benthic state (Table 1; Fig. 7). These higher rates of diversification are especially prevalent in the Amphiprionini, several parts of the *Pomacentrus* tree, and in *Dascyllus* (Fig 6). The particularly elevated diversification distribution among pelagic damselfishes seen in Fig. 7 represents the anemonefishes.

4. Discussion

The synthesis of species-rich phylogenies and diverse morphological, functional, and ecological trait data yields new ways of visualizing evolutionary patterns, promotes the development of novel hypotheses, and empowers their testing using comparative methods. Marine fishes offer particularly revealing systems for these explorations due to the recent development of corresponding phylogenies and datasets (organismal and/or ecological characters) for species-rich lineages (19,21). Integrative phylogenetic approaches to reef fishes allow us to address important questions associated with local and global biogeographic patterns, the history of structural and functional evolution, and the diversification of important traits that have driven species richness for high diversity reef fish families. The central conclusions of this work are that (1) the core structure of the damselfish phylogeny is resolved with high support, but that intriguing questions remain to be answered using larger datasets, (2) the timing of several major damselfish diversification events occurs near the Eocene-Oligocene transition, (3) multistate hidden-state speciation and extinction models show that transitions between dietary ecotypes are a significant influence on damselfish diversification, (4) there is pronounced asymmetry in transition directionality between body size and ecotype that is a significant driver of diversification patterns in damselfishes, and (5) the convergent pattern of ecological diversification that has been described as a characteristic of the Pomacentridae as a whole is limited mostly to the Indo-West Pacific subfamily Pomacentrinae.

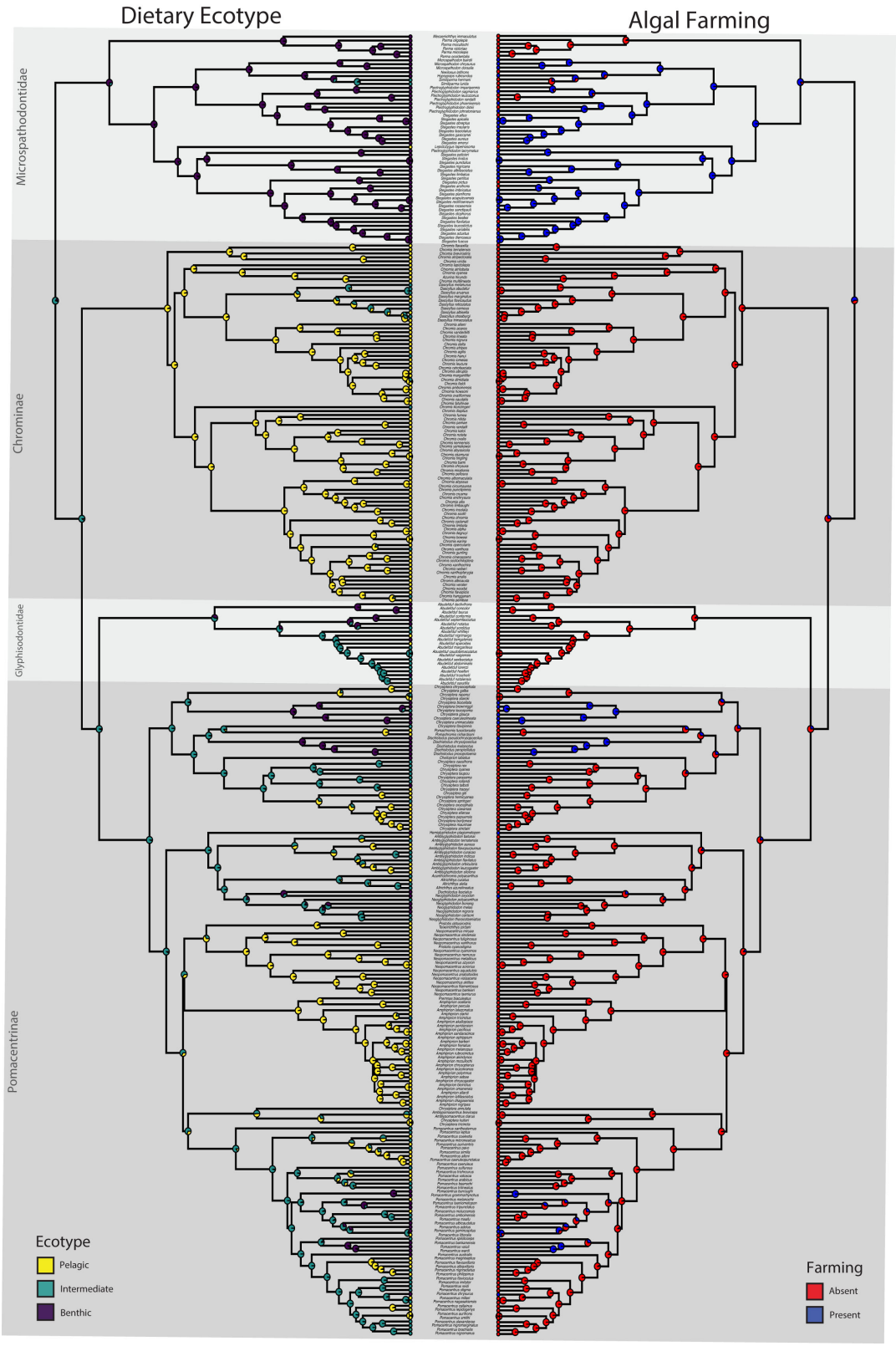

Figure 5. Mirror tree illustration of dietary ecotype and farming behavior. The patterns of dietary ecotype history (left) are illustrated with the presence and absence of the behavioral trait of algal patch "farming" (right) on the damselfish time-calibrated phylogeny.

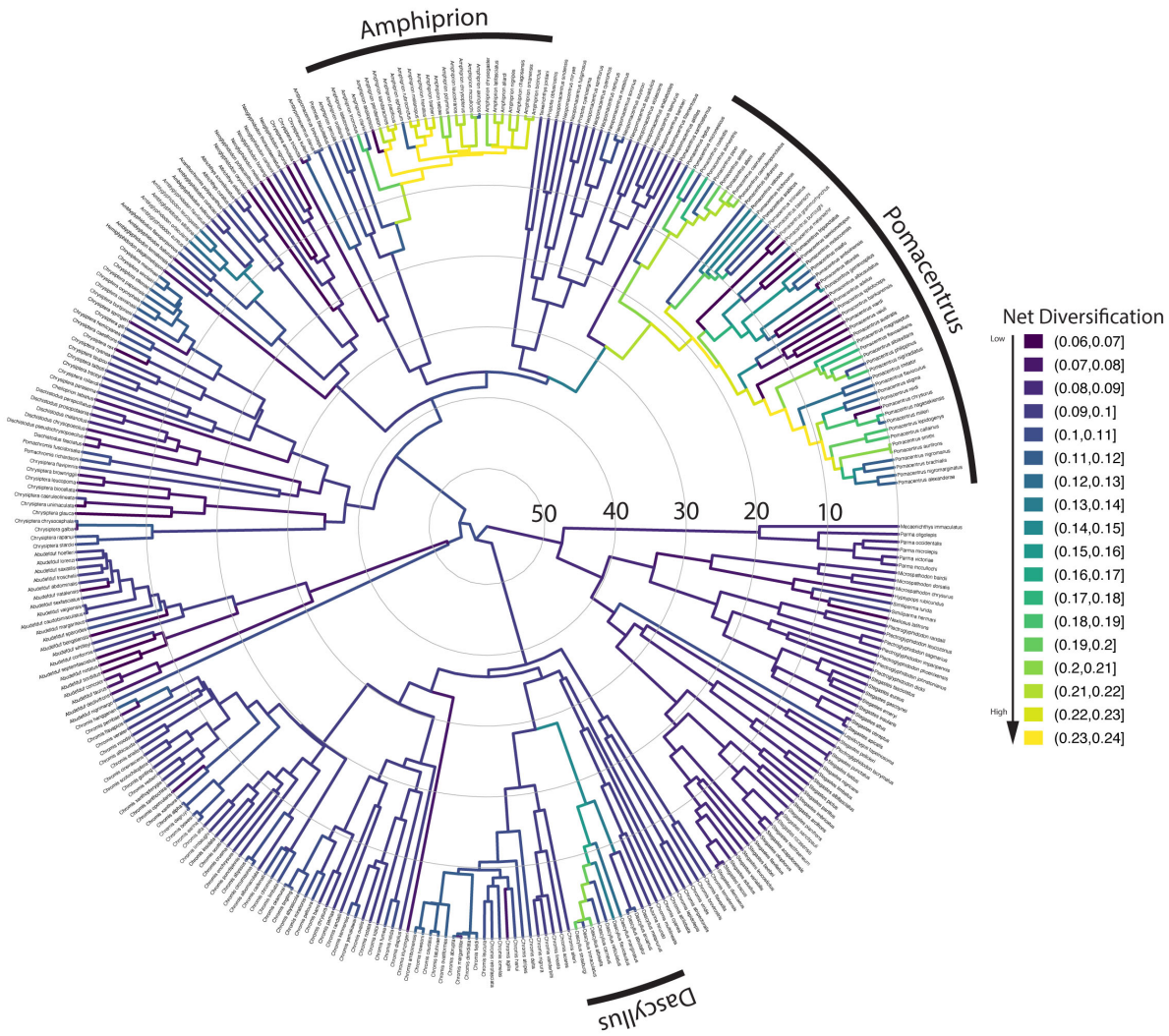

Figure 6. Net diversification rates on the time-calibrated phylogeny of the Pomacentridae. Background purple color shows the steady pattern of diversification characteristic of the family, with several independent origins of elevated diversification (*Amphiprion* is the most prominent) in lighter colors throughout the tree.

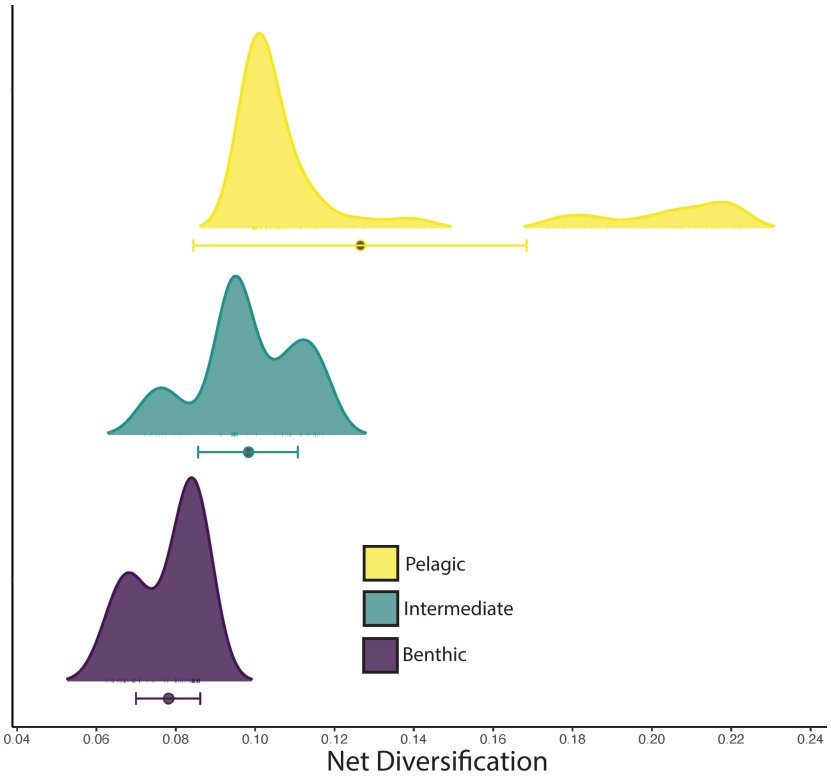

Figure 7. The distribution of net diversification across damselfish ecotypes. Model averaged mean and standard deviation of diversification for each ecotype is indicated by a colored circle and bar, respectively. The height of the peak indicates the number of tips for each ecotype with that particular estimated net diversification rate. Mean estimated net diversification was highest in the pelagic state (with *Amphiprion* showing the highest level to the right), followed by the intermediate state, with the benthic state showing slowest rate of net diversification.

4.1 Damselfish phylogeny: insights from increased resolution and remaining puzzles

Our understanding of phylogenetic relationships among the Pomacentridae has improved tremendously in completeness, resolution and support as species sampling and character matrix density has increased. From early work by Tang (34), Jang-Liaw (31) and Quenouille et al. (33) to Cooper et al. (1), Frédérich et al. (15), and the present study, the molecular phylogeny of damselfishes has grown from 23 species to now include 330 of the 422 known species. All regions of the tree are densely sampled, and although many species are represented by just one or a few gene sequences, the overall topology is well-resolved and generally well-supported under both likelihood and Bayesian approaches.

The phylogeny of the subfamily Microspathodontinae (Fig. 1) has been relatively stable over the past decade. Several taxonomic issues have been resolved (71), but other puzzles remain. *Parma* species are resolved as the monophyletic sister group to *Mecaenichthys*, as we have eliminated the older sequence data for *Parma oligolepis* that appears to be a misidentification or contamination linking it to *Stegastes*. *Nexilosus* is now resolved as sister to the *Similiparma-Hypsypops* clade, using recent CO1 data only, and eliminating the older sequences for this species that were generated using ancient DNA approaches on a preserved specimen (1). However, the resolution of *Nexilosus* in this position remains uncertain and its relationship with other damselfishes merits further investigation.

The genus *Stegastes* is not monophyletic, with *Plectroglyphidodon* sister to one *Stegastes* clade and *Lepidozygus* sister to another that also contains *Plectroglyphidodon lacrymatus* (Fig. 1). Although the Bayesian posterior support for *Lepidozygus* in this position is high, we note that the long branch for this species has made it hop around in recent analyses (1,15,33) so view its position here with caution. We consistently get *Plectroglyphidodon lacrymatus* as sister to *Stegastes pelicieri* in *Stegastes* 2 (Fig. 1), and this is supported by multiple genes from multiple specimens. However, a similar finding in this part of the tree, that *Pomacentrus agassizi* is sister to *Stegastes fasciolatus*, was rejected and *P. agassizii* deleted from our matrix due to evidence of identification errors for specimens in GenBank and BOLD (specimens labelled *P. agassizii* appear to be *S. fasciolatus* in the photos). Once these issues are resolved with additional data from specimens with certain identification, substantive name changes will be required in this region of the tree to yield monophyletic genera.

Within *Stegastes*, *Stegastes* 1 is the sister lineage to the remainder of *Plectroglyphidodon* (Fig. 1). This clade includes the type species for the genus (*P. johnstonianus*; (73). The members of this clade should therefore be retained in *Plectroglyphidodon*. *Stegastes* 1 should be described as a new genus or its species should be redescribed as members of *Plectroglyphidodon*. *Stegastes* 3 (Fig. 1) includes *Stegastes imbricatus,* the type species for the genus (30) collected by Charles Darwin in the Cape Verde archipelago, during the first landfall of the *H.M.S. Beagle* after leaving England. *Stegastes* 3 and its sister lineage, *Stegastes* 2 (including *Plectroglyphidodon lacrymatus*), should therefore be retained in *Stegastes* (or re-described as a member of this genus in the case of *P. lacrymatus*).

The topology of the pomacentrid subfamily Chrominae (Fig. 1) has also remained stable across recent phylogenetic hypotheses, with three solidly monophyletic groups of taxa that render *Chromis* itself solidly polyphyletic. The monotypic *Azurina* should receive formal taxonomic revision, and the position of *Dascyllus* nested deep within Chrominae group 2 (Fig. 1) is consistent and intriguing. Taxonomic work in this group will be based on the history that both *Chromis* and *Dascyllus* were described by Cuvier, with *Chromis* Cuvier 1814 having priority over *Dascyllus* Cuvier 1829 by 15 years (50). However, we also show that the finer, detailed species relationships within the Chrominae are uncertain, with many clades poorly supported (Fig. 1), so we suggest that major taxonomic name changes, as well as detailed evolutionary studies focused on within-chromine patterns, may require additional research.

Recent phylogenetic work (1,15) and analyses focused on the genus *Abudefduf* (74,75) have provided both Sanger and UCE sequence data for all 21 known species, rendering *Abudefduf* as one of the most clearly resolved, monophyletic pomacentrid groups (Glyphysodontinae; Fig. 2). The time calibrated reconstruction here of about 30 mya for the genus is older than the 25 mya recovered by Campbell et al. (75), due to new time-calibrated outgroup nodes in the present tree (61). However, the 25 mya age remains within the 95% highest posterior density error bars of our estimate. The species-level topology of *Abudefduf* shown here is also similar but not identical to that resolved under ML for the ultra-conserved element data (75), with the positions of *A. concolor* and *A. lorenzi* being the primary differences, and the topology based on the larger UCE dataset likely to be preferred.

The subfamily Pomacentrinae is full of phylogenetic intrigue, taxonomic issues, and evolutionary fascination (Fig. 2). Pomacentrinae 1 contains the two main clades of *Chrysiptera* with a clade of *Pomachromis* + *Dischistodus* (sister genera, each resolved as strongly monophyletic) nested between them with high Bayesian posterior probability, and support in other analyses as well (Fig. S1, S2). Pomacentrinae 1 is therefore also likely to require

significant taxonomic revision. Pomacentrinae 3 (Fig. 2) also raises taxonomic questions with the clade containing *Pristotis*, *Teixeirichthys*, and *Neopomacentrus* rendered non-monophyletic by intercalation of *Pristotis*. In contrast, Pomacentrinae 2 is a diverse group of damselfishes whose relationships are congruent with current taxonomy, with *Hemiglyphidodon*, *Amblyglyphidodon*, *Acanthochromis*, *Altrichthys*, and *Neoglyphidodon* all resolved as monophyletic.

The remarkable clade of anemonefishes, or clownfishes (Tribe Amphiprionini; Pomacentrinae 4; Fig. 2) is resolved with high support as a monophyletic clade that diversified almost 15 million years ago and underwent a burst of diversification during the past 5 million years. These findings are consistent with other recent estimates of 5-13 mya for the origin and radiation of the Amphiprionini (15,76,77). Litsios et al. (76) supported the long-held hypothesis that mutualism with sea anemones promoted the adaptive diversification of the anemonefishes. Marcionetti et al. (78) identified 17 genes that underwent positive selection in the early stages of the anemonefish radiation, two of which are potentially associated with chemical components of the toxin discharge mechanisms of sea anemones. Although the monotypic genus *Premnas* is distinguished from *Amphiprion* by the possession of a large suborbital spine that is not seen in other members of the tribe (23), the synonymy of *Premnas* with *Amphiprion* seems clear and necessary so as to render anemonefish taxonomy monophyletic.

Our results show strong support for the monophyly of *Pomacentrus* (Pomacentrine 5; Fig. 2), which is the second largest genus of damselfishes, and good support for most of the relationships within this clade (Fig. 2). However, it should be noted that most of the resolution within the *Pomacentrus* tree is supported by barcode or other small loci such as d-loop, provided by recent new species descriptions (79,80). This phylogeny for *Pomacentrus*, which contains 52 of the 80 recognized species, is a another step forward in the species tree of this genus, as prior phylogenetic work has included up to 34 *Pomacentrus* species (15).

4.2 The timing of diversification across ocean basins

New time-calibrated phylogenies yield fresh insight into the timing of key lineage splitting and diversification events among the damselfishes, including the timing of damselfish diversification in the Atlantic/Eastern Pacific (AEP) and Indo-West Pacific (IWP) regions. Most damselfishes live in the IWP, including the entire subfamily Pomacentrinae, yet seven damselfish sub-clades inhabit the broader AEP region. For example, although the Microspathodontinae are mostly fishes of the IWP, two microspathodontine lineages inhabit the AEP: the Microspathodontini (the giant damselfishes) and a small clade of *Stegastes*. In a similar pattern, most Chrominae are IWP residents, but at least two lineages have diversified in the AEP. The genus *Abudefduf* diversified in the Atlantic/Eastern Pacific twice: once early in their diversification (*A. declivifrons*, *A. concolor*, and *A. taurus*) and a second time more recently (*A. hoefleri*, *A. saxatilis*, and *A. troschelii*).

In contrast to the other major clades, the Pomacentrinae are restricted to the Indo-West Pacific and the majority of them inhabit coral reefs (1,71). Although no extant pomacentrine damselfishes are native to either the Atlantic or East Pacific, their late Eocene divergence date (40 mya) considerably precedes the closing of the circumtropical Tethys seaway between 12 and 18 mya (Early Miocene; Fig. 2) when warm-water connections between the Atlantic and Indo-West Pacific were severed (81). Their diversification across the tropical Indo-West Pacific and their absence on coral reefs of the AEP is therefore somewhat of a paradox.

Most major branching events in pomacentrine evolution roughly coincide with the Eocene-Oligocene Transition about 33.5 mya (Fig 2), a significant global biotic reorganization

with major changes to ocean ecosystems (82). Many pomacentrine branching events occurred around the EOT, suggesting that these fishes may have benefitted from this period of ecological upset. Yet the steady radiation of pomacentrines throughout the tropical IWP is in contrast to their absence from the AEP, which may be the result of a coral extinction event in the late Oligocene and Early Miocene. Between 24 and 16 mya, a large portion of the reef-building coral genera that inhabited what is now the Caribbean went locally extinct, but persisted in the Indo-West Pacific (83,84). This loss of Caribbean coral reef habitat extended into the period during which the Tethys Seaway closed 12-18 mya (85). Coral reef habitat loss in the Caribbean may have caused pomacentrine extinctions, and the Tethyan closure, the East Pacific Barrier, and colder ocean waters around Southern Africa prevented their recolonization of the Atlantic and Eastern Pacific. The damselfishes are a rich system for more detailed exploration of biogeographic patterns and reef fish community assembly across the marine realm.

4.3 Ecotype transitions and asymmetric damselfish diversification patterns

The damselfishes are a curious example of a lineage that has undergone a rapid, steady, yet somewhat constrained adaptive radiation (14,15). Damselfish diversification to become one of the largest and most important clades on coral reefs is characterized, and perhaps driven by, complex patterns of divergence and convergence in size, shape and feeding ecology, within a curiously bounded set of ecomorphological end points. The regular, clock-like nature of damselfish speciation through time (Fig. S4) rejects early-burst or late-burst evolutionary hypotheses (86) for the generation of high species richness. The metronomic pace of damselfish radiation (excepting *Amphiprion*) is in contrast to patterns seen in some fish groups such as cichlids (87). These steady diversification rates are significantly associated with ecotype transitions in these fishes, in combination with a hidden (non-focal) trait according to our MuHiSSE tests of trait effects on diversification across the phylogeny (Fig. 6). Perhaps the most intriguing finding of this study is that transitions among size classes and among dietary ecotypes in the damselfishes are strongly asymmetrical (Figs. 6, 7), with transitions through intermediate states the rule, frequent back-and-forth transitions in some groups, but with some clades often getting somewhat evolutionarily "stuck" at the extremes. This is similar to prior results (37) showing high rates of transitions out of the intermediate dietary ecotype into the pelagic and benthic ecotypes. However, in contrast to this prior work (37), that found that the intermediate ecotype has the highest diversification rate as opposed to the "specialists", we found higher diversification rates associated with the pelagic ecotype, followed by the intermediate and the benthic ecotype. The highest diversification rate in damselfishes is associated with a pelagic feeding ecotype, and the clownfishes of the genus *Amphiprion* lead this trend.

It is intriguing that the Pomacentridae exhibit just three major feeding niches: planktivory, omnivory and herbivory (with the exception of the corallivorous *Cheiloprion*). Diversification of the damselfishes has not involved expansion into an increasing number of ecological niches, but rather repeated convergence on a limited set of similar feeding morphologies and ecologies. This pattern has been described as driving damselfish diversification due to the repeated pattern of convergence on these three primary ecotypes (14), and has been considered to be characteristic of the damselfishes as a whole. However, the present analysis of dietary ecotype using a more comprehensive phylogeny shows that it is primarily the Pomacentrinae that have undergone this ecotype diversification (Fig. 4, 6). Although all three ecotypes are distributed across the phylogeny, most major clades specialize on a single ecotype, or exhibit a small number of ecotype transitions (Fig. 4, 5). Examples of this

include the typically large body size and herbivorous ecotype of the microspathodontine clade. At the other extreme, almost all Chrominae are relatively small and planktivorous, with the *Dascyllus* lineage departing from this pattern as a rare reversion to the intermediate state within the clade (Fig. 4, 5). Although most *Dascyllus* species feed on plankton, many also consume considerable amounts of algae (88–90). Their trophic morphology is characteristic of omnivorous damselfishes and the feeding kinematics of *Dascyllus aruanus* are similar to those of other pomacentrid omnivores (38). The Glyphisodontinae (*Abudefduf*) are a relatively species-poor lineage but they have evolved all three ecotypes, but with only a small amount of the back-and-forth pattern in dietary and body size transitions seen among the Pomacentrinae (Fig. 4).

The Pomacentrinae (and in particular *Pomacentrus*) show a much stronger pattern of ecotypic convergent evolution, with repeated transitions between all three ecotypes (Fig. 4-6). Although the Pomacentrinae exhibit the highest degree of geographic and habitat limitations of any of the subfamilies (only found in the Indo-West Pacific; almost entirely restricted to coral reefs) and are one of the two most recently evolved major clades, they have produced roughly twice as many species as the Chrominae, almost three times as many species as the Microspathodontinae, and are approximately ten times more species-rich than their sister lineage Glyphisodontinae. Members of all five damselfish subfamilies can be found on Indo-West Pacific coral reefs, but only the Pomacentrinae have demonstrated a pronounced ability to frequently transition between three ecotypes that lie along a bentho-pelagic axis within these ecosystems. The enhanced evolvability of the Pomacentrinae relative to other damselfishes may have contributed to the tremendous success of this lineage.

Algal farming or gardening behavior is a complex behavioral trait that has evolved multiple times in three main damselfish lineages, occurring in fishes of diverse body sizes but restricted to the shallow benthic habitats where desirable algae grow (Fig. 5). Cultivating and defending an algal patch has long been recognized as an unusual level of behavioral complexity in fish feeding behavior (41,91,92), the intricacy of which recently increased as it has been shown to include the domestication of mysid shrimps by damselfishes to help fertilize the garden (93). The independent origin of this complex behavior up to a dozen times across the tree, associated with intense aggressive territorial defense behavior, and associated each time with the benthic ecotype that shows the lowest diversification rate across the tree, makes this a promising system for study of the factors that accelerate or slow diversification in damselfishes.

In summary, an expanded, time-calibrated damselfish phylogeny allows us to make important additions to our understanding of damselfish ecological radiation. First, there is asymmetry in the direction of ecotype transition, with shifts away from the intermediate omnivorous ecotype to either the pelagic or benthic ecotype. Second, it is primarily the crown group in the family, the Pomacentrinae, showing the strongest pattern of convergence in ecotype. All three ecotypes have evolved repeatedly across the damselfish phylogeny, but before the appearance of the Pomacentrinae, transitions in ecomorphology were more likely to be associated with major branching events rather than at finer levels within clades. Increased phylogenetic resolution, feeding mechanics studies, dietary ecology and developmental biology of the damselfishes will continue to reveal surprising evolutionary trends in this spectacular group of fishes.

Acknowledgments
Thanks to Lydia Smith and Jillian Henss for assistance with DNA sequencing, and to the Division of Fishes and the Pritzker Molecular Laboratory at the Field Museum of Natural

History for collections and data acquisition support. We also thank the fish collections of the Australian Museum, Scripps Institution of Oceanography, National Museum of Natural History in the Smithsonian Institution, Institute of Zoology in the Academia Sinica, and the Gil Rosenthal lab of Texas A & M University for supplying pomacentrid tissues. Supported by NSF grant DEB 1541547.

Literature Cited

1. Cooper WJ, Smith LL, Westneat MW. Exploring the radiation of a diverse reef fish family: Phylogenetics of the damselfishes (Pomacentridae), with new classifications based on molecular analyses of all genera. Mol Phylogenet Evol. 2009 Jul;52(1):1–16.

2. Westneat MW, Alfaro ME. Phylogenetic relationships and evolutionary history of the reef fish family Labridae. Mol Phylogenet Evol. 2005 Aug;36(2):370–90.

3. Thacker CE. Phylogeny of Gobioidei and Placement within Acanthomorpha, with a New Classification and Investigation of Diversification and Character Evolution. Copeia. 2009 Feb 23;(1):93–104.

4. Tornabene L, Ahmadia GN, Berumen ML, Smith DJ, Jompa J, Pezold F. Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: Eviota). Mol Phylogenet Evol. 2013 Jan;66(1):391–400.

5. Santini F, Carnevale G, Sorenson L. First multi-locus timetree of seabreams and porgies (Percomorpha: Sparidae). Ital J Zool. 2014 Jan 2;81(1):55–71.

6. Craig MT, Hastings PA. A molecular phylogeny of the groupers of the subfamily Epinephelinae (Serranidae) with a revised classification of the Epinephelini. Ichthyol Res. 2007 Feb;54(1):1–17.

7. Sorenson L, Santini F, Carnevale G, Alfaro ME. A multi-locus timetree of surgeonfishes (Acanthuridae, Percomorpha), with revised family taxonomy. Mol Phylogenet Evol. 2013 Jul;68(1):150–60.

8. Longo G, Bernardi G. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. Mol Phylogenet Evol. 2015 Jul;88:55–63.

9. Thacker CE, Roje DM. Phylogeny of Gobiidae and identification of gobiid lineages. Syst Biodivers. 2011;9(4):329–47.

10. Thacker CE, Roje DM. Phylogeny of cardinalfishes (Teleostei: Gobiiformes: Apogonidae) and the evolution of visceral bioluminescence. Mol Phylogenet Evol. 2009 Sep;52(3):735–45.

11. Fessler JL, Westneat MW. Molecular phylogenetics of the butterflyfishes (Chaetodontidae): Taxonomy and biogeography of a global coral reef fish family. Mol Phylogenet Evol. 2007 Oct;45(1):50–68.

12. McCord CL, Westneat MW. Phylogenetic relationships and the evolution of BMP4 in triggerfishes and filefishes (Balistoidea). Mol Phylogenet Evol. 2016 Jan;94:397–409.

13. Smith LL, Fessler JL, Alfaro ME, Streelman JT, Westneat MW. Phylogenetic relationships and the evolution of regulatory gene sequences in the parrotfishes. Mol Phylogenet Evol. 2008 Oct;49(1):136–52.

14. Cooper WJ, Westneat MW. Form and function of damselfish skulls: rapid and repeated evolution into a limited number of trophic niches. BMC Evol Biol. 2009;9(1):1–17.

15. Frédérich B, Sorenson L, Santini F, Slater GJ, Alfaro ME. Iterative Ecological Radiation and Convergence during the Evolutionary History of Damselfishes (Pomacentridae. Am Nat. 2013;181(1):94–113.

16. Westneat MW, Alfaro ME, Wainwright PC, Bellwood DR, Grubichl JR, Fessler JL, et al. Local phylogenetic divergence and global evolutionary convergence of skull function in reef fishes of the family Labridae. Proc R Soc B-Biol Sci. 2005 May 22;272(1567):993–1000.

17. Wainwright PC, Smith WL, Price SA, Tang KL, Sparks JS, Ferry LA, et al. The Evolution of Pharyngognathy: A Phylogenetic and Functional Appraisal of the Pharyngeal Jaw Key Innovation in Labroid Fishes and Beyond. Syst Biol. 2012 Dec;61(6):1001–27.

18. Collar DC, Wainwright PC, Alfaro ME. Integrated diversification of locomotion and feeding in labrid fishes. Biol Lett. 2008 Feb 23;4(1):84–6.

19. Floeter SR, Bender MG, Siqueira AC, Cowman PF. Phylogenetic perspectives on reef fish functional traits. Biol Rev. 2018;93(1):131–51.

20. Bellwood D, Klanten S, Cowman P, Pratchett M, Konow N, Van Herwerden L. Evolutionary history of the butterflyfishes (f: Chaetodontidae) and the rise of coral feeding fishes. J Evol Biol. 2010;23(2):335–49.

21. Siqueira AC, Morais RA, Bellwood DR, Cowman PF. Trophic innovations fuel reef fish diversification. Nat Commun. 2020;11(1):1–11.

22. Konow N, Bellwood DR, Wainwright PC, Kerr AM. Evolution of novel jaw joints promote trophic diversity in coral reef fishes. Biol J Linn Soc. 2008;93(3):545–55.

23. Allen GR. Damselfishes of the World. Mentor, Ohio: Aquarium Systems; 1991.

24. Frédérich B, Parmentier E. Biology of damselfishes. CRC Press; 2016.

25. Linnaeus C. Systema naturae. Vol. 1. Stockholm Laurentii Salvii; 1758.

26. Artedi P, Linnaeus C. Petri Artedi Ichthyologia sive opera omnia de piscibus. apud Conradum Wishoff; 1738.

27. Aristotle. Historia Animalium, translated by Arthur L. Peck. II Camb MA Lond. 1965;

28. Lacépède BGÉ de. Histoire naturelle des poissons. Plasson Paris. 1802;Vol. 4: i-xliv:1–728, pls. 1-16.

29. Cuvier GS, Valenciennes A. Histoire naturelle des poissons. Hist Nat Poissons. 1846;19:3.

30. Jenyns L, Darwin C. The Zoology of the Voyage of HMS Beagle, Under the Command of Captain Fitzroy, During the Years 1832 to 1836. Smith, Elder & Company; 1842.

31. Jang-Liaw N-H, Tang KL, Hui C-F, Shao K-T. Molecular phylogeny of 48 species of damselfishes (Perciformes: Pomacentridae) using 12S mtDNA sequences. Mol Phylogenet Evol. 2002 Dec 1;25(3):445–54.

32. McCafferty S, Bermingham E, Quenouille B, Planes S, Hoelzer G, Asoh K. Historical biogeography and molecular systematics of the Indo-Pacific genus Dascyllus (Teleostei: Pomacentridae. Mol Ecol. 2002;11(8):1377–92.

33. Quenouille B, Bermingham E, Planes S. Molecular Systematics of the Damselfishes (Teleostei: Pomacentridae): Bayesian Phylogenetic Analyses of Mitochondrial and Nuclear DNA Sequences. Mol Phylogenet Evol. 2004;31(1):66–88.

34. Tang KL. Phylogenetic relationships among damselfishes (Teleostei: Pomacentridae) as determined by mitochondrial DNA data. Copeia. 2001;2001(3):591–601.

35. Santini S, Polacco G. Finding Nemo: molecular phylogeny and evolution of the unusual life style of anemonefish. Gene. 2006;385:19–27.

36. Bernardi G. Monophyletic origin of brood care in damselfishes. Mol Phylogenet Evol. 2011;59(1):245–8.

37. Gajdzik L, Aguilar-Medrano R, Frédérich B. Diversification and functional evolution of reef fish feeding guilds. Ecol Lett. 2019;22(4):572–82.

38. Cooper WJ, Carter CB, Conith AJ, Rice AN, Westneat MW. The evolution of jaw protrusion mechanics is tightly coupled to bentho-pelagic divergence in damselfishes (Pomacentridae). J Exp Biol. 2017;220(4):652–66.

39. Olivier D, Parmentier E, Frédérich B. Insight into biting diversity to capture benthic prey in damselfishes (Pomacentridae). Zool Anz- J Comp Zool. 2016;264:47–55.

40. Frédérich B, Olivier D, Litsios G, Alfaro ME, Parmentier E. Trait decoupling promotes evolutionary diversification of the trophic and acoustic system of damselfishes. Proc R Soc B Biol Sci. 2014;281(1789):20141047.

41. Hata H, Kato M. A novel obligate cultivation mutualism between damselfish and Polysiphonia algae. Biol Lett. 2006;2(4):593–6.

42. Cooper WJ, VanHall R, Sweet E, Milewski H, DeLeon Z, Verderber A, et al. Functional morphogenesis from embryos to adults: Late development shapes trophic niche in coral reef damselfishes. Evol Dev. 2020;22(3):221–40.

43. Frédérich B, Colleye O, Vandewalle P. Allometric growth in the damselfishes of the genus Dascyllus (Pomacentridae. J Morphol. 2007;268(12):1074–1074.

44. Frédérich B, Sheets HD. Evolution of ontogenetic allometry shaping giant species: a case study from the damselfish genus Dascyllus (Pomacentridae. Biol J Linn Soc. 2010;99(1):99–117.

45. Frederich B, Vandewalle P. Bipartite life cycle of coral reef fishes promotes increasing shape disparity of the head skeleton during ontogeny: an example from damselfishes (Pomacentridae. BMC Evol Biol. 2011;11:1–21.

46. Simpson GG. Tempo and mode in evolution. Columbia University Press; 1944.

47. Colles A, Liow LH, Prinzing A. Are specialists at risk under environmental change? Neoecological, paleoecological and phylogenetic approaches. Ecol Lett. 2009;12(8):849–63.

48. Maddison WP. Confounding asymmetries in evolutionary diversification and character change. Evolution. 2006;60(8):1743–6.

49. Nakov T, Beaulieu JM, Alverson AJ. Diatoms diversify and turn over faster in freshwater than marine environments. Evolution. 2019;73(12):2497–511.

50. Fricke R, Eschmeyer W. Eschmeyer's catalog of fishes: genera, species, references. Laan R, editor. 2020.

51. Maddison WP, Maddison DR. Mesquite: a modular system for evolutionary analysis. Version [Internet]. 2018;3(51). Available from: http://www.mesquiteproject.org

52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.

53. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 Nov 1;23(21):2947–8.

54. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol. 2016;34(3):772–3.

55. Zwickl D. GARLI: genetic algorithm for rapid likelihood inference. 2006. Ph. D. dissertation, The University of Texas at Austin. http://garli …; 2014.

56. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012 May 1;61(3):539–42.

57. Rambaut A. FigTree v1. 4. 2012.

58. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017 Jan 1;8(1):28–36.

59. Bellwood DR. Fossil pharyngognath fishes from Monte Bolca, Italy, with a description of a new pomacentrid genus and species. Studi E Ric Sui Giacimenti Terziari Bolca. 1999;8:207–17.

60. Arambourg C. Les poissons fossiles d'Oran. Vol. 6. J. Carbonel; 1927.

61. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci. 2018;115(24):6249–54.

62. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLOS Comput Biol. 2014 Apr 10;10(4):e1003537.

63. Boettiger C, Lang DT, Wainwright PC. rfishbase: exploring, manipulating and visualizing FishBase data from R. J Fish Biol. 2012 Nov 1;81(6):2030–9.

64. James W. Archie. Methods for Coding Variable Morphological Features for Numerical Taxonomic Analysis. Syst Zool. 1985;(3):326.

65. Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3(2):217–23.

66. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. GEIGER: investigating evolutionary radiations. Bioinformatics. 2008;24(1):129–31.

67. Beaulieu JM, O'Meara BC. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol. 2016;65(4):583–601.

68. Caetano DS, O'Meara BC, Beaulieu JM. Hidden state models improve state-dependent diversification approaches, including biogeographical models. Evolution. 2018;72(11):2308–24.

69. Beaulieu J, O'Meara B, Caetano D, Boyko J. Package 'hisse.' 2020.

70. Conway M, Olsen BJ. Contrasting drivers of diversification rates on islands and continents across three passerine families. Proc R Soc B. 2019;286(1915):20191757.

71. Cooper WJ, Santini F. A revised damselfish taxonomy with a description of the tribe Microspathodontini (Giant Damselfishes. In: Biology of the Damselfishes. Editors: Eric Parmentier; 2016.

72. Pybus OG, Harvey PH. Testing macro–evolutionary models using incomplete molecular phylogenies. Proc R Soc Lond B Biol Sci. 2000;267(1459):2267–72.

73. Fowler HW, Ball SC. Descriptions of new fishes obtained by the Tanager Expedition of 1923 in the Pacific islands west of Hawaii. Proc Acad Nat Sci Phila. 1924;76:269–74.

74. Aguilar-Medrano R, Barber PH. Ecomorphological diversification in reef fish of the genus Abudefduf (Percifomes, Pomacentridae. Zoomorphology. 2016;135(1):103–14.

75. Campbell MA, Robertson DR, Vargas MI, Allen GR, McMillan WO. Multilocus molecular systematics of the circumtropical reef-fish genus Abudefduf (Pomacentridae): history, geography and ecology of speciation. PeerJ. 2018;6:5357.

76. Litsios G, Sims CA, Wuest RO, Pearman PB, Zimmermann NE, Salamin N. Mutualism with sea anemones triggered the adaptive radiation of clownfishes. BMC Evol Biol. 2012;12:15.

77. Elliott JK, Lougheed SC, Bateman B, McPhee LK, Boag PT. Molecular phylogenetic evidence for the evolution of specialization in anemonefishes. In: Proceedings of the Royal Society Of London Series B-Biological Sciences. 1999. p. 677–85.

78. Marcionetti A, Rossier V, Roux N, Salis P, Laudet V, Salamin N. Insights into the Genomics of Clownfish Adaptive Radiation: Genetic Basis of the Mutualism with Sea Anemones. Genome Biol Evol. 2019;11(3):869–82.

79. Allen GR, Drew JA. A new species of Damselfish (Pomacentrus: Pomacentridae) from Fiji and Tonga. Aqua Int J Ichthyol. 2012;18(3):171–81.

80. Allen G, Erdmann M, Purtiwi P. Descriptions of four new species of damselfishes (Pomacentridae) in the Pomacentrus philippinus complex from the tropical western Pacific Ocean. J Ocean Sci Found. 2017;25:47–76.

81. Harzhauser M, Kroh A, Mandic O, Piller WE, Gohlich U, Reuter M, et al. Biogeographic responses to geodynamics: A key study all around the Oligo-Miocene Tethyan Seaway. Zool Anz. 2007;246(4):241–56.

82. Hooker JJ, Collinson ME, Sille NP. Eocene-Oligocene mammalian faunal turnover in the Hampshire Basin, UK: calibration to the global time scale and the major cooling event. J Geol Soc. 2004;161:161–72.

83. Edinger EN, Risk MJ. OLIGOCENE MIOCENE EXTINCTION AND GEOGRAPHIC RESTRICTION OF CARIBBEAN CORALS - ROLES OF TURBIDITY, TEMPERATURE, AND NUTRIENTS. Palaios. 1994;9(6):576–98.

84. Zachos J, Pagani M, Sloan L, Thomas E, Billups K. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science. 2001;292(5517):686–93.

85. Harzhauser M, Piller WE. Benchmark data of a changing sea - Palaeogeography, palaeobiogeography and events in the Central Paratethys during the Miocene. Palaeogeogr Palaeoclimatol Palaeoecol. 2007;253(1–2):8–31.

86. Harmon LJ, Losos JB, Jonathan Davies T, Gillespie RG, Gittleman JL, Bryan Jennings W, et al. Early bursts of body size and shape evolution are rare in comparative data. Evol Int J Org Evol. 2010;64(8):2385–96.

87. Arbour JH, López-Fernández H. Ecological variation in South American geophagine cichlids arose during an early burst of adaptive morphological and functional evolution. Proc R Soc B Biol Sci. 2013;280(1763):20130849.

88. Gerber RP, Marshall N. Ingestion of detritus by the laboon pelagic communicty at Eniwetok Atoll. Limnol Oceanogr. 1974;19(5):815–24.

89. Sano M, Shimizu M, Nose Y. Food Habits of Teleostean Reef Fishes in Okinawa Island, Southern Japan. Bull Univ Mus Univ Tokyo. 1984;25:1–128.

90. Kuo S-R, Shao K-T. Feeding Habits of Damselfishes (Pomacentridae) from the Southern Part of Taiwan. J Fish Soc Taiwan. 1991;18(3):165–76.

91. Lobel PS. Herbivory by damselfishes and their role in coral reef community ecology. Bull Mar Sci. 1980;30(1):273–89.

92. Lassuy DR. Effects of "farming" behavior by Eupomacentrus lividus and Hemiglyphidodon plagiometopon on algal community structure. Bull Mar Sci. 1980;30(1):304–12.

93. Brooker RM, Casey JM, Cowan Z-L, Sih TL, Dixson DL, Manica A, et al. Domestication via the commensal pathway in a fish-invertebrate mutualism. Nat Commun. 2020;11(1):1–9.

Supporting information

S1 Fig. Best ML tree. Maximum likelihood topology for the damselfishes.
S2 Fig. Time tree posterior support. Time-calibrated tree for the damselfishes showing posterior probability support for nodes of the tree.
S3 Fig. Time tree node ages. Time-calibrated tree for the damselfishes showing reconstructed ages of nodes of the tree.
S4 Fig. Lineage through time plot. The lineage through time plot for the damselfishes.
S1 Table. Genbank accession numbers. Genbank accession numbers for the 12 genes and 335 taxa used in the damselfish phylogenetic analysis, with color coding by major data contributor.
S2 Table. Primers. A list of primers used in gene sequencing.
S3 Table. Trait data for damselfishes. Trait data on ecotype, farming behavior, body size, and body depth for the Pomacentridae.

S1 File. TreeFile1 BestML.tre. Treefile for the maximum likelihood topology of the damselfishes.
S2 File. TreeFile2 TimeTree335.tre. Treefile for the time-calibrated topology of the damselfishes from BEAST analysis, with 5 outgroups.
S3 File. TreeFile3 DamselsOnlyTimeTree300.phy. Treefile for the time-calibrated topology of the damselfishes from BEAST analysis, damselfish only, in .phy format for upload in R.