

# Phylogenetic insights into the history and diversification of fishes on reefs

S. A. Price<sup>1</sup> · T. Claverie<sup>2</sup> · T. J. Near<sup>3</sup> · P. C. Wainwright<sup>1</sup>

Received: 23 October 2014 / Accepted: 6 July 2015 / Published online: 15 July 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** Studies of the phylogenetic history of fishes on reefs and the impact of reefs on fish diversification have, to date, been limited to relatively small clades. We take advantage of a recent multi-locus, time-calibrated phylogeny of acanthomorph fishes and a broad-scale morphological dataset of body shape in reef acanthomorphs to explore the history and diversification of fish on reefs at the family level. We find that no reef family exhibits exceptional species diversity for their stem age and some, such as Aulostomidae, Zanclidae, Menidae, and Triodontidae may in fact be species poor. The inferred history of reef colonization is highly dependent on how a reef family is defined; one classification scheme raises the possibility that most modern acanthomorph families originated on reefs. We find that most reef families occupy surprisingly distinct regions of morphospace and yet, some of the most diverse reef families occupy central and highly overlapping positions within the body shape morphospace. To the extent

that proximity in morphospace reflects ecological similarity, these results imply that most reef fish families have diversified in adaptive zones away from other families. In contrast, a few of the most successful (e.g., Labridae and Pomacentridae) have achieved dominance while potentially facing stronger interactions with other lineages. Finally, we find no relationship between species diversity and body shape diversity. Assuming neither are diversity dependent, this result suggests that morphological and ecological diversification within families of reef fish may not be linked to the accumulation of species. Time-calibrated phylogenies provide the means for generating a greater understanding of the macroevolutionary processes influencing reef fish diversification, but we are currently limited by the lack of robust crown-group ages for many reef fish families.

**Keywords** Acanthomorpha · Macroevolution · Diversification · Reef fishes

Communicated by Ecology Editor Dr. Alastair Harborne

**Electronic supplementary material** The online version of this article (doi:10.1007/s00338-015-1326-7) contains supplementary material, which is available to authorized users.

✉ S. A. Price  
saprice@ucdavis.edu  
P. C. Wainwright  
pcwainwright@ucdavis.edu

<sup>1</sup> Department of Evolution and Ecology, University of California, Davis, CA 95616, USA

<sup>2</sup> CUFR de Mayotte, Route nationale 3, 97660 Dembeni, France

<sup>3</sup> Department of Ecology and Evolutionary Biology, Peabody Museum of Natural History, Yale University, New Haven, CT, USA

## Introduction

Reef fishes are iconic for their diversity and are a crucial component of coral reef ecosystems. Thus, understanding how and when the reef fish fauna evolved is an important goal that will illuminate the processes which generated a major element of biodiversity (e.g., Bellwood 1996, 2003; Bellwood and Wainwright 2002; Cowman and Bellwood 2011). However, a limited fossil record and uncertainty over the age of, as well as phylogenetic relationships among reef fish families, has restricted the scope of such investigations (Bellwood and Wainwright 2002). Recent advances in acanthomorph phylogenetics (Betancur-R et al. 2013; Near et al. 2013) therefore provide an unprecedented

opportunity to gain insights into the history and biology of reef fishes. These new phylogenies, built using multiple protein coding genes and time-calibrated with numerous fossils, resolve for the first time many inter-family relationships and ages of divergence events. Since  $\sim 92\%$  of reef fishes are acanthomorphs (FishBase: Froese and Pauly 2013), these phylogenies provide a major step forward in our ability to explore patterns of colonization, diversification, and morphospace occupation by reef fish lineages.

Recently, using the Near et al. (2013) acanthomorph phylogeny, we presented evidence that the modern reef fish fauna was formed during two waves of reef colonization on either side of the Cretaceous–Paleogene boundary  $\sim 66$  mya (Price et al. 2014). But it is unclear whether subsequent diversification leads to exceptional species diversity in reef families in comparison with other non-reef acanthomorphs. Moreover if true, can this diversification simply be explained by reef clades being older, as opposed to diversifying faster? Both intrinsic and extrinsic factors have been hypothesized to drive elevated diversification on reefs. Intrinsic factors include reefs being highly productive, as well as physically and biologically complex environments, all of which are expected to promote species coexistence (e.g., Schoener 1974), and potentially speciation (Alfaro et al. 2007) through niche differentiation and specialization. Extrinsic factors include the expansion of scleractinian coral reefs (e.g., Perrin 2002), along with changes in sea level and ocean circulation during the Paleogene and Neogene (e.g., Zachos et al. 2005; Nunes and Norris 2006). Phylogenetic studies of patterns within several important reef fish groups have concluded that reef association plays a significant role in driving cladogenesis (Tetraodontiformes: Alfaro et al. 2007; Chaetodontidae, Apogonidae, Pomacentridae, and Labridae: Cowman and Bellwood 2011). The Miocene expansion of coral reefs has been linked to an increase in lineage diversification rate within several modern reef fish families and is thought to have resulted in much of the present-day diversity of chaetodontids, labrids, and apogonids (Cowman and Bellwood 2011).

The fossil record indicates that by 50 million years ago most major functional groups of reef fishes, including herbivores, were present (Bellwood 2003). Using the Near et al. (2013) phylogeny, we inferred that the earliest lineages moving onto reefs were morphologically distinct but by the early Eocene the morphospace was saturated (Price et al. 2014). However, the overall pattern of body shape diversity (i.e., disparity) and morphospace occupation by reef fish families remains to be determined. For example, do exceptionally species-rich reef fish clades occupy relatively unique regions of body shape morphospace, as we would predict if they initially invaded reef habitats, found an empty adaptive zone, and radiated within it. Under this

scenario, we would also expect morphological disparity and species richness to be positively correlated. Indeed, across ray-finned fishes speciation rates and rates of body size evolution are correlated (Rabosky et al. 2013). Moreover, within Labridae and Haemulidae, reef-associated lineages exhibit higher rates of functional morphological diversification in traits related to prey capture and processing than non-reef lineages (Price et al. 2011, 2012, respectively). Therefore, the same intrinsic and extrinsic forces that promote higher speciation rates in reef fishes may also elevate morphological diversification rates, especially in the traits that underpin niche differentiation. We thus expect to see a positive relationship between disparity and species richness across reef fishes.

In this study we address a variety of basic questions concerning reef fish evolution. We first explore the phylogenetic history of reef invasion and ask whether there are monophyletic clusters of reef families (each cluster representing a single invasion of reef habitats), or whether every reef family invaded reefs independently. We also ask whether reef families show exceptional species richness for their age as compared to other families of acanthomorph fishes and whether reef families are consistently older than non-reef clades. Further, by using a dataset on the shape of thousands of reef associated acanthomorphs (taken from Claverie and Wainwright 2014), we explore the history of body shape diversification in reef fish. We ask if there is a relationship between species richness and morphological disparity across families. Additionally, focusing on the eight most species-rich groups of reef fishes, we ask what areas of morphospace they occupy and whether these eight groups tend to occupy unique regions of space.

## Materials and methods

### Phylogeny

We used the recent, time-calibrated phylogeny of Acanthomorpha (Near et al. 2013), which was built using ten protein coding genes and dated using 37 fossil calibrations, to provide the framework for identifying clusters of reef families. Taxonomic sampling was sparse at the species level, but at the family-level sampling was far more complete including 228 out of a total of 322 extant acanthomorph families. We pruned the maximum clade credibility tree to a single, randomly chosen, exemplar species per family, with each species assigned to a family using Catalog of Fishes (Eschmeyer 2012). A few families were found to be non-monophyletic (Near et al. 2013); where this conclusion is supported by independent evidence (e.g., Serranidae—Smith and Craig 2007), we split the families into separate monophyletic lineages using a single

exemplar species. Tree manipulations were completed in the statistical computing framework ‘R’ (R Core Team 2014) using the packages *ape* (Paradis et al. 2004) and *geiger* (Harmon et al. 2008).

As our sampling at the species level is approximately 3 % of known acanthomorphs, the phylogeny does not provide suitable estimates of family crown ages. We searched the scientific literature for reliable crown-age estimates but found that surprisingly few acanthomorph families have well-sampled, time-calibrated phylogenies. We therefore resorted to using stem ages (the divergence between each family and its sister clade) estimated from the Near et al. (2013) phylogeny. We calculated the median stem age for every acanthomorph family using a sample of 1000 phylogenies from the Bayesian posterior distribution of trees.

### Identifying clusters of reef families and their history of reef dwelling

To answer any questions about reef fish evolution, we first needed to define what a reef fish is, which is both difficult and contentious (e.g., Bellwood 1998; Robertson 1998). We therefore used two approaches in our analyses, allowing us to identify any definition-dependent conclusions. The first is a list of families taken from Bellwood and Wainwright (2002), which represents ‘those taxa that are found on, and characteristic of, coral reefs’ henceforth the ‘B&W list’ (Bellwood and Wainwright 2002). This list includes some families that are only regionally characteristic of reefs, such as Labrisomidae, which are common on Caribbean reefs but not reefs globally. The list was updated to the most recent taxonomy (e.g., including Odacidae and Scaridae within Labridae, Caesionidae within Lutjanidae, splitting Epinephelidae from Serranidae etc.) and includes a total of 60 families within the Near et al. (2013) phylogeny. The second list is based on the percentage of extant species that are ‘reef associated’ according to FishBase (Froese and Pauly 2013), which includes fishes living on or near any shallow water, consolidated and wave-resistant structure. For each family, we calculated the percentage of extant species living on reefs using FishBase, accessed through functions in the R package *rfishbase* (Boettiger et al. 2012). These percentages were then used to calibrate a binomial distribution for every family from which we sampled 100 times. This generated 100 habitat datasets with a binary habitat variable, reef (1) and non-reef (0), sampled in proportion to the extant percentage living in each habitat type. For example, FishBase categorizes 83.3 % of extant Labridae species as living on reefs, thus the binomial distribution had a probability of success (1 = reef) of 0.833 and of failure (0 = non-reef)  $1 - 0.833 = 0.167$ . Sampling from this distribution 100

times gave us a combined dataset that listed labrids as reef dwelling 81 out of 100 times.

Both approaches used to identify reef clades generate a binary variable, with the ‘non-reef’ category encompassing families from a wide variety of marine and freshwater habitats. This means that the comparisons between reef and non-reef lineages, in terms of age and diversity, are conservative, as reefs have to exceed all other habitats combined. Despite the conservative nature of comparing reef and non-reef clades, we have successfully used this binary habitat variable to find evidence of increased rates of morphological diversification in reef lineages within several acanthomorph families (Price et al. 2011, 2012). Moreover, the composite non-reef category is unlikely to bias the inferred history of reef dwelling or the estimated evolutionary transitions onto or away from reefs. For example, the number of transitions to reef habitats and the placement of those transitions will be similar regardless of whether a simple non-reef category or a complex set of categories describing specific freshwater and marine habitats is used.

We used stochastic character mapping (Nielsen 2002) to infer the history of reef dwelling from both the B&W and the FishBase datasets. Unlike simple ancestral state reconstruction, character mapping is a Bayesian method, which allows transitions to occur along branches as well as at nodes. Stochastic mapping generates multiple possible histories in proportion to their posterior probability, so results are not conditioned on a single reconstruction. It does this through simulating character histories of habitat along the branches of the phylogeny that are consistent with the observed habitat distribution upon the phylogeny. From this set of habitat maps, we can calculate the number of transitions from non-reef to reef habitats, the posterior probability of each state (reef or non-reef) occurring at any node, as well as identify possible clusters of reef families. Clustering of reef families indicates a single colonization event and subsequent diversification upon reefs. We generated 10,000 stochastic maps in the program SIMMAP (Bollback 2006) with an uninformative prior on the symmetry of the transition rate matrix ( $\alpha = 1$  and  $\kappa = 101$ ) and a branch length prior on the rate parameter. We summarized the posterior probability of reef colonization across a sample of a thousand trees using the *densityMap* function in the *phytools* package in R (Revell 2011).

### Detecting exceptionally species-rich or species-poor reef clades

Net diversification is commonly modeled by a stochastic birth–death process, with speciation and extinction occurring at a constant rate. Under this model we would expect

to see a positive relationship between clade age and species richness. Even if the assumption of constant rates is relaxed, with rates varying between lineages, on average species richness should be positively correlated with clade age (Ricklefs 2009; Rabosky 2010). Thus, a clade that appears to be exceptionally species rich may simply be much older than the other clades. The stochastic nature of birth–death processes and the exponential increase in species richness with time complicates the identification of exceptionally species-rich or species-poor clades and thus significant departures from a time-homogenous process. Since a single birth and death rate combination can produce a very wide range of species richness's for a given clade age, exceptional clades must exceed the range of clade sizes predicted from the stochastic time-homogenous model.

We identified exceptional clades using the methods described in Magallón and Sanderson (2001). We estimated the background speciation rate from the acanthomorph phylogeny as a whole and used it to estimate a 95 % confidence interval for the expected number of species for each acanthomorph family based on their median stem age in the Near et al. (2013) phylogeny. As we currently have no good way of estimating the extinction rate across acanthomorphs, we used an extinction rate of 0 and 90 % of speciation rate to bracket the possible rates to classify any clades that fall outside these estimates as exceptionally species rich or species poor. We implemented these analyses using functions in the R package *geiger* (Harmon et al. 2008). This was compared to the extant species richness, which was calculated using the number of species listed in each family in FishBase using the R package *rfishbase* (Boettiger et al. 2012). Estimates of species richness from FishBase were checked to ensure that they were in general agreement with existing taxonomic estimates (Nelson 2006; Eschmeyer 2012). We implemented Magallón and Sanderson (2001) analyses using functions in the R package *geiger* (Harmon et al. 2008).

We also investigated whether reef lineages were older or more species rich than their non-reef counterparts. We used both non-phylogenetic (OLS) and phylogenetic (PGLS) linear regression models to analyze the relationships, with stem age and  $\log_{10}$  species richness as separate response variables. All analyses were implemented in R including functions in the *caper* (PGLS) package (Harmon et al. 2008; Orme et al. 2013). To account for the degree of phylogenetic covariation within the residuals of the PGLS, we used the maximum likelihood estimate of Pagel's  $\lambda$  (Pagel 1997).  $\lambda$  is a branch length transformation of the internal branches of the phylogeny, which varies between 0 (no phylogenetic signal) and 1 (perfect phylogenetic signal).

### Estimating morphospace occupation and shape disparity across reef clades

To estimate body shape, we used geometric morphometrics from 17 landmarks on lateral photos of 2875 species of Indo-Pacific fishes in 53 families of Acanthomorpha, taken from Claverie and Wainwright (2014). This dataset contains predominately reef clades: 42 of the 53 families are recognized as reef families by the B&W classification, and according to FishBase reefs support at least 10 % of the extant diversity in every one except Sciaenidae and Leiognathidae. A principal component analysis (PCA) on the Procrustes superimposed landmark coordinates from the geometric morphometric analyses was used to summarize fish shape diversity. Each family's shape disparity was calculated as the sum of the variance across all PC axes; this generated disparity estimates indistinguishable from the Procrustes Variance calculated on the Procrustes aligned coordinates implemented in the R package *Geomorph* (Adams and Otárola-Castillo 2013). To examine the impact of the highly variable number of species sampled per family, we ran a rarefaction analysis sampling 20 species per family and calculated the median disparity from 100 rarefaction samples. A few families had fewer than 20 species and were therefore excluded from the rarefaction.

Disparity, like species richness, can be influenced by clade age, as older clades have had more time to accumulate differences. Therefore, to estimate the relationship between species richness and disparity, we needed to include clade age as a covariate. Ideally we would have used the crown age of each family since disparity, unlike lineage diversity, does not begin to accumulate until the first speciation event in the tree (O'Meara et al. 2006), but this information is not currently available. We were therefore limited to using stem age, which is likely to add noise and increase type II error. We ran both OLS and PGLS models with the form: Disparity  $\sim$   $\log_{10}$  Species Richness  $\times$  Stem Age, again using Pagel's  $\lambda$  (Pagel 1997) to account for the degree of phylogenetic covariation within the residuals of the PGLS model.

For subsequent analyses of morphospace occupation, we used the first two PC axes, as they are the interpretable axes according to the broken-stick model and can be easily visualized using simple bivariate plots. PC1 is dominated by a contrast between deep-bodied and elongate shapes, and PC2 reflects the contrast between long and short dorsal and anal fins, together these two axes account for 58 % of the variance in the dataset. To examine the pattern of morphospace occupation by reef families, we calculated the minimum area convex polygons between PC1 and PC2 for each family and estimated the pairwise overlap between families using the *adehabitatHR* (Calenge 2006), *rgeos* (Bivand and Rundel 2014) and *maptools* (Bivand and

Lewin-Koh 2014) R packages. To control for the highly variable number of species per family, we calculated the median pairwise overlap of 100 rarefaction samples of 20 species per family. We then estimated the median overlap for each family to identify those families that show the greatest morphospace overlap.

Eight reef fish families have 100 or more species that are found on reefs around the world. To determine whether these eight most diverse reef fish clades occupy more distinct regions of morphospace than expected, we calculated the overlap between these eight families compared to their total morphospace occupation and compared it to the average overlap between eight randomly selected families, which we repeated 500 times. We used rarefaction with the median of 100 samples of 20 species per family to control for the variability of sampling.

The full dataset, including disparity estimates, rarefied disparity estimates, median stem age etc. is available in the Electronic Supplementary Materials, ESM, Table 1.

## Results

### Identifying clusters of reef families and the history of reef dwelling

When using the percentage of extant species living on reefs in each family to identify reef fish families, we have a strong pattern of multiple reef colonizations since the beginning of the Paleogene (Fig. 1a, c). We estimate a median of 53 non-reef to reef habitat transitions from the 1000 stochastic character maps (Fig. 1c). However, there is still evidence that reef families are more closely related than expected from random (see Price et al. 2014). From the stochastic maps, we identify a few distinct clusters of families with strong ties to reefs: syngnathiformes, blennioids, and lutjanids plus their relatives (Fig. 1a). Each of these clusters may represent a single reef colonization event, although the overall posterior probability of a reef state for most of these clusters is fairly low (ESM Table 2). Poor support for reef clusters may be partly due to the long tip branches in the tree, which are a consequence of each family being represented by a single species. Additionally, there are several pairs of families that may also represent single transitions to reefs: Acanthuridae/Zanclidae, Balistidae/Monacanthidae, Ehippididae/Drepanidae, and potentially Tetraodontidae/Diodontidae, as well as Cheilodactylidae/Cirrhitidae.

In contrast, when we used the B&W list, there is strong support for a single early invasion of reefs followed by multiple transitions away from reefs (Fig. 1b, d). From the 1000 stochastic character maps, we estimate a median of 18 non-reefs to reef habitat transitions and 88 reef to non-

reef transitions (Fig. 1d). According to the posterior probability estimated at the nodes (ESM Table 3), there is a strong probability that the colonization event took place before the node that separates Lampriformes, Gadiformes, Percopsiformes, and Polymixiiformes (node 241: posterior probability of reef = 0.961) from all other acanthomorphs *circa* 125 million years ago.

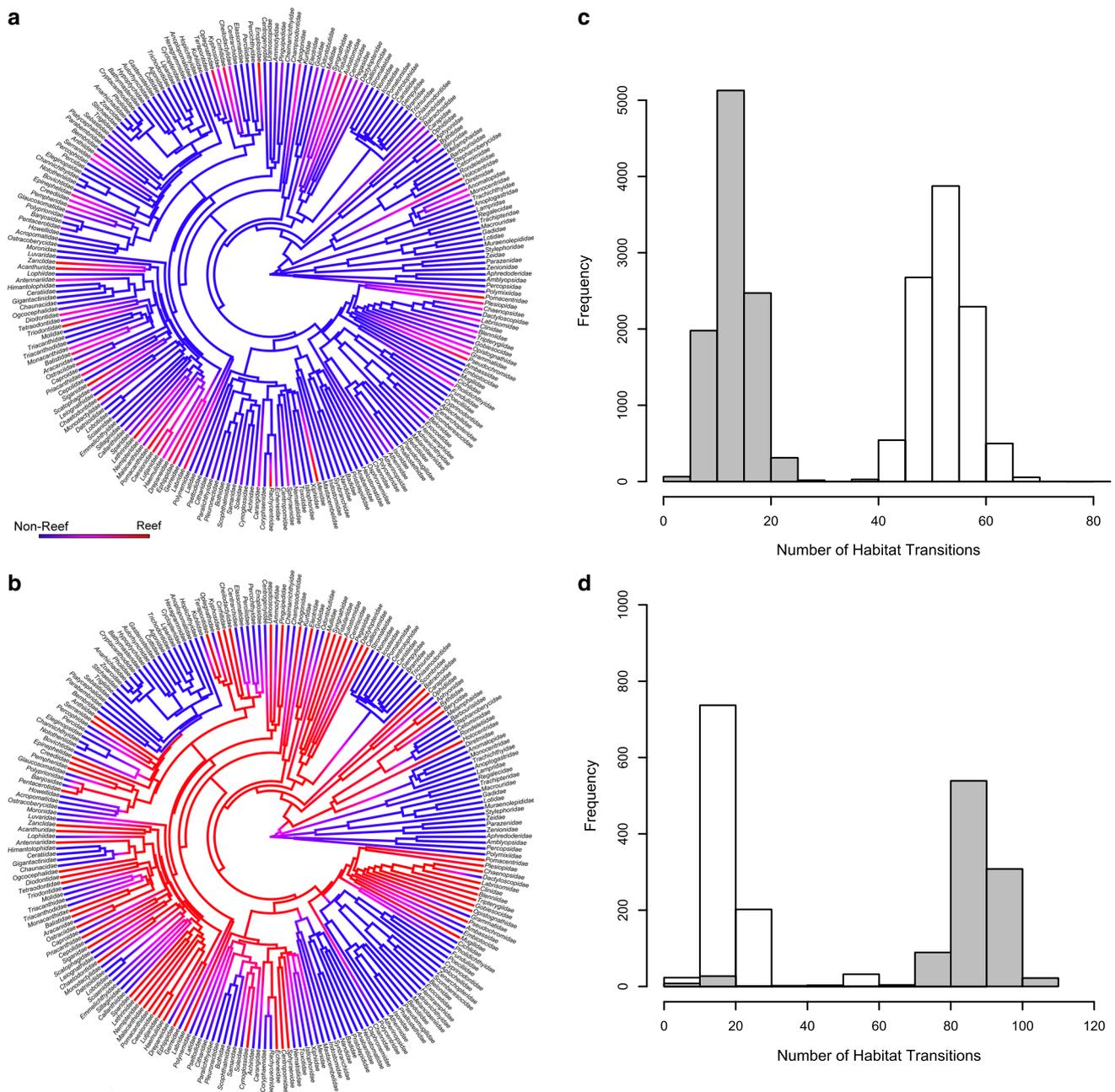
### Detecting exceptionally species-rich or species-poor reef clades

According to the methods of Magallón and Sanderson (2001), no reef family is unambiguously species rich, as most fall within the 95 % CI calculated from the overall acanthomorph speciation rate (Fig. 2). However, many reef families do fall between the high and low extinction CI, so depending on the extinction rate within acanthomorphs, some clades (including Syngnathidae, Blenniidae, Chaetodontidae, Haemulidae, Serranidae, Epinephelidae, and Labridae) may be recognized as exceptionally diverse (ESM Table 4). Gobiidae are the second most species rich family on reefs (Fig. 3) and do show exceptional richness for their age, although they are not recognized as a reef family in the B&W list as only 25 % of goby species live on reefs. Most of the taxa that show exceptional species richness are centered in the North Pacific region or are found in deep water: Cottidae, Liparidae, Stichaeidae, and Zoarcidae. However, several reef families are identified as species poor including: Zanclidae, Menidae, and Triodontidae, all of which only contain a single living species, and Aulostomidae, which contains three extant species but has a stem age older than most reef clades with origins in the Late Cretaceous.

When phylogeny is not taken into account, reef families, regardless of how they are defined, are found to be both older and more species rich than non-reef clades, but this becomes non-significant when phylogeny is taken into account (Table 1). Phylogenetic comparative methods should be implemented when the residuals from the model are phylogenetically patterned, for all four analyses this condition is met as the maximum likelihood estimate of Pagel's  $\lambda$  (Pagel 1997) is significantly greater than 0. This result suggests that reef families only appear to be older and have greater species richness because of the phylogenetic non-independence within the dataset.

### Estimating morphospace occupation and shape disparity across reef clades

Shape disparity was estimated for 53 families of acanthomorphs with close associations with reefs. The most disparate families were Syngnathidae, Synanceiidae, and



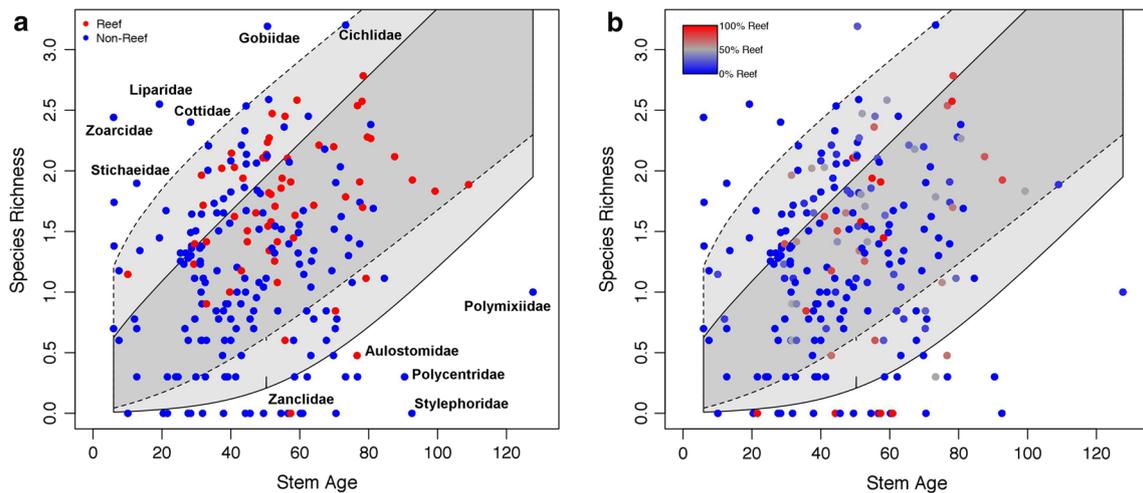
**Fig. 1** Reconstructed history of reef living according to **a** the percentage of extant species living on reefs given in FishBase **b** classification of Bellwood and Wainwright (2002). Color gradation represents the posterior probability of reef (red) or non-reef (blue) ancestry according to 1000 stochastic character maps on the maximum clade credibility phylogeny from Near et al. (2013)

Antennariidae and the least disparate were Mugilidae, Sphyrnidae, and Hemiramphidae (Fig. 4). With the caveat that we were unable to control for family crown ages, there is no strong relationship between disparity and species richness across acanthomorph families. Both the PGLS and OLS regression of disparity and  $\log_{10}$  species

converted to the family level. Histograms of the number of habitat transitions from non-reef to reef (gray) and from reef to non-reef (white) according to **c** the percentage of extant species living on reefs given in FishBase **d** classification of Bellwood and Wainwright (2002), calculated from the 1000 stochastic character maps

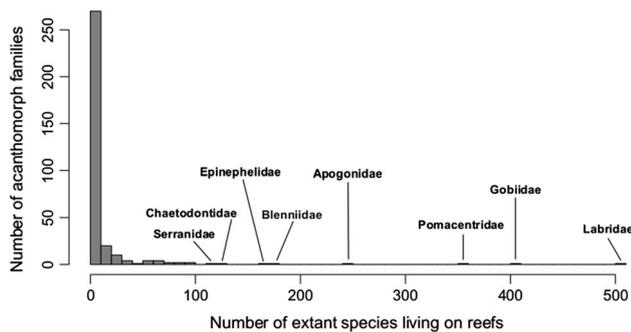
richness display a weak but non-significant positive relationship (Table 2).

Following rarefaction to control for the different number of species within the geometric morphometric dataset, the vast majority of the 33 families overlap very little in morphospace as described by the first two PC axes of body



**Fig. 2** The relationship between stem age and species richness with 95 % confidence intervals calculated using the methods of Magallón and Sanderson (2001), assuming no extinction (*full lines*) and 90 %

extinction (*dashed lines*). Points are colored according to **a** classification of Bellwood and Wainwright (2002) **b** the percentage of extant species living on reefs given in FishBase



**Fig. 3** Histogram of the number of extant species living on reefs per acanthomorph family according to FishBase. All eight families that contain over 100 reefs species are named

shape. The median overlap between each family and every other family is less than 1 % for all families except Labridae, Serranidae, Pomacentridae, and Carangidae (Table 3, the complete pairwise estimates are in ESM Table 5).

The eight most species-rich families on reefs occupy much of the acanthomorph morphospace (Fig. 5) except the extremes of being deep-bodied (e.g., batfishes), elongate (e.g., needlefishes and mullets), and fish with very short dorsal and anal fins. By calculating the mean pairwise overlap between these eight families and comparing it to eight randomly drawn acanthomorph families, we found that the degree of overlap exhibited by the eight most species rich reef families is unusually high. The empirical estimate of the pairwise overlap, although within the 95 % CI, is at the upper extreme of that estimated for the randomly drawn families (Fig. 6), suggesting that they overlap

more than expected, even when differences in sample size are taken into account through rarefaction.

### Discussion

Despite occupying less the 1 % of the earth’s surface, reefs are home to the most species-rich fish assemblages on earth (Harmelin-Vivien 2002). Our results reveal that reef fish families are not more species rich for their age, or older than acanthomorph families that live in other ecosystems. Moreover, acanthomorph families that are more species rich do not show a corresponding elevation in body shape variability. In general, acanthomorph families associated with reefs occupy distinct areas of body shape morphospace, suggesting that these individual evolutionary radiations have influenced each other. Nevertheless, some of the most successful reef families, including Labridae and Pomacentridae, occupy central regions of body shape morphospace, which overlap with many other families.

The inferred history of reef dwelling in acanthomorphs depends on how we define a reef family. Using the B&W list, we conclude that reefs are the center of origin for all reef fishes and in fact most of Acanthomorpha. This suggests that the modern reef fish fauna evolved in situ on reefs and potentially exported diversity to other habitats. This finding is consistent with patterns inferred from the fossil record, which suggest reefs have been exporting taxonomic diversity throughout the Phanerozoic (KieSSLing et al. 2010). In contrast, using the percentage of extant species that live on reefs as a continuous estimate of the degree to which that family is an important reef fish clade,

**Table 1** Statistical table for the linear regression analyses of habitat on stem age and species richness

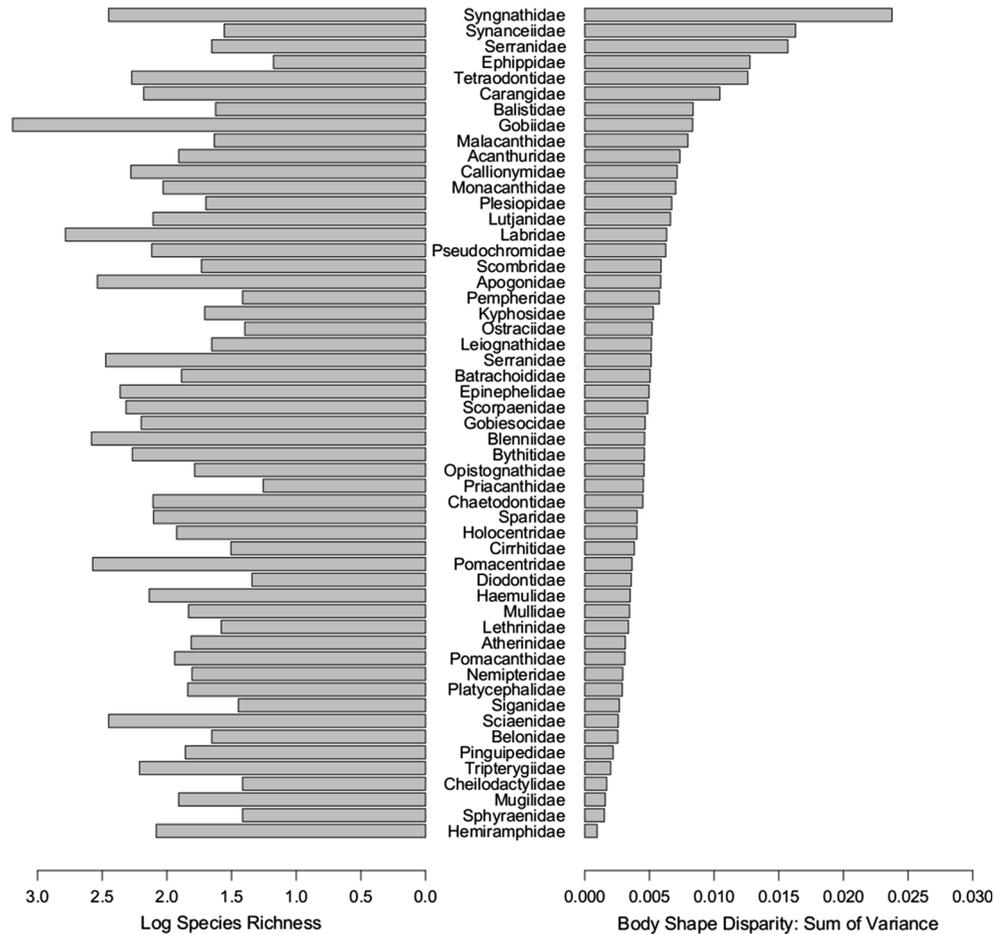
Variable	Non-phylogenetic (OLS)							
	Habitat: % extant species				Habitat: Bellwood and Wainwright (2002)			
	Coeff.	<i>p</i> value	Adj. <i>R</i> <sup>2</sup>		Coeff.	<i>p</i> value	Adj. <i>R</i> <sup>2</sup>	
Age ~ habitat								
Intercept	45.091	0.000	0.043		44.777	0.000	0.0627	
Habitat	0.152	0.001			11.851	0.000		
Log <sub>10</sub> species richness ~ habitat								
Intercept	1.045	0.000	0.025		0.972	0.000	0.1524	
Habitat	0.005	0.002			0.725	0.000		
	Phylogenetic (PGLS)							
	Habitat: % extant species				Habitat: Bellwood and Wainwright (2002)			
	Coeff.	<i>p</i> value	$\lambda$	Adj. <i>R</i> <sup>2</sup>	Coeff.	<i>p</i> value	$\lambda$	Adj. <i>R</i> <sup>2</sup>
Age ~ habitat								
Intercept	63.506	0.000	1.000	0.030	60.010	0.000	1.000	0.07957
Habitat	0.116	0.163			12.356	0.058		
Log <sub>10</sub> species richness ~ habitat								
Intercept	2.246	0.000	0.593	-0.012	2.261	0.000	0.545	-0.01161
Habitat	-0.002	0.443			-0.143	0.436		

We used two approaches to identify reef fishes. The first is based on the percentage of extant species that are ‘reef associated’ according to FishBase (Froese and Pauly 2013) and the second is a list of families taken from Bellwood and Wainwright (2002).  $\lambda$  is the maximum likelihood estimate of the branch length transformation of the internal branches of the phylogeny, which varies between 0 (indicating no phylogenetic signal) and 1 (indicating perfect phylogenetic signal)

we infer a non-reef ancestral acanthomorph (Price et al. 2014). Consistent with our previous findings, which used a similar definition of reef fish, we find that there have been multiple colonizations of reefs throughout the history of acanthomorphs (Price et al. 2014). Our conflicting findings mirror the uncertainty in the literature about the origins of reef fishes (as reviewed by Bellwood and Wainwright 2002) and a lack of consensus upon what is, or is not, a reef fish (e.g., Bellwood 1996,1998; Robertson 1998). Early attempts to infer where various reef fishes originated using phylogenies concluded that most clades probably evolved in non-reef environments (Bellwood 1996; Bellwood and Wainwright 2002). The fossil record provides little additional help; many reef families first appear in the Eocene Monte Bolca Lagerstätte. There is, however, no direct evidence of the Monte Bolca deposit being a preserved coral reef, though it was certainly a shallow tropical coastline in the Tethys region in close proximity to coral reefs (Bellwood 1998). Overall the weight of opinion appears to support a non-reef origin and multiple invasions of reef habitats by fishes (Bellwood and Wainwright 2002; Price et al. 2014). Increasingly comprehensive phylogenies with new fossil analyses will continue to elucidate this question.

It is not useful to identify clusters of reef fishes in the context of the results from the B&W list because reefs were recovered as the most likely ancestral habitat. However, looking at the stochastic character maps from the FishBase reef dataset, we identify a few small, monophyletic clusters of reef families that potentially represent single origins of reef living. These clusters include syngnathiformes, lutjanids, and their relatives, blennies and their relatives, as well as several pairs of families, notably: balistids/monacanthids, acanthurids/zanclids, and ephippids/drepanids. The lutjanid cluster (including lutjanids, haemulids, lethrinids, sparids, pomacanthids, and malacanthids), syngnathiformes and the blennioids (including blenniids, grammistids, opistagnathids, gobiesocids, and trypterygids) each appear to represent particularly old transitions to reefs that date back before the Cretaceous-Paleogene boundary to about 80-90 mya. However, the posterior probability of each of these groups having a single reef origin is fairly low (see ESM Table 2). Support for these reef clusters are low, in part because we used a single exemplar species for each family thereby generating artificially long branches to the tips of the phylogeny. Consequently, a major benefit of a comprehensive phylogeny with full crown-group diversity for each family

**Fig. 4** Species richness and body shape disparity as measured by the sum of variance



**Table 2** Statistical table for the linear regression analyses of disparity versus species richness for both OLS and PGLS

	Non-phylogenetic (OLS)		Phylogenetic (PGLS)		
	Coeff.	<i>p</i> value	Coeff.	<i>p</i> value	$\lambda$
Intercept	0.006	0.637	0.005	0.709	1.000
Log <sub>10</sub> species richness	0.001	0.876	0.002	0.689	
Median stem age	0.000	0.780	0.000	0.642	
Log <sub>10</sub> species richness × median stem age	0.000	0.899	0.000	0.829	

$\lambda$  is the maximum likelihood estimate of the branch length transformation of the internal branches of the phylogeny, which varies between 0 (indicating no phylogenetic signal) and 1 (indicating perfect phylogenetic signal)

(~17,000 species) is that it would provide a more reliable estimate of the history of reef living across acanthomorphs, as it would take into account the distribution of reef-dwelling lineages within each family as well.

We find little evidence that any reef fish families are exceptionally species rich for their age compared to other acanthomorphs (Fig. 2). As we currently have no good way of estimating the extinction rate across acanthomorphs, we estimated the expected species richness at the extremes: no extinction and extinction at 90 % of the speciation rate. The species richness of 14 reef fish families identified by

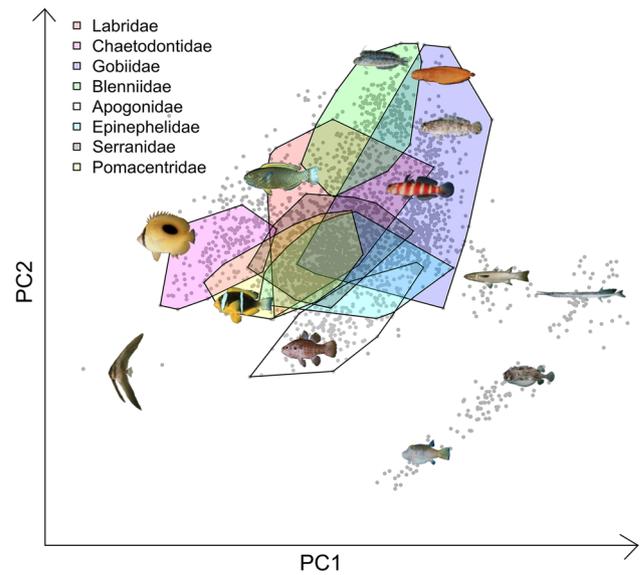
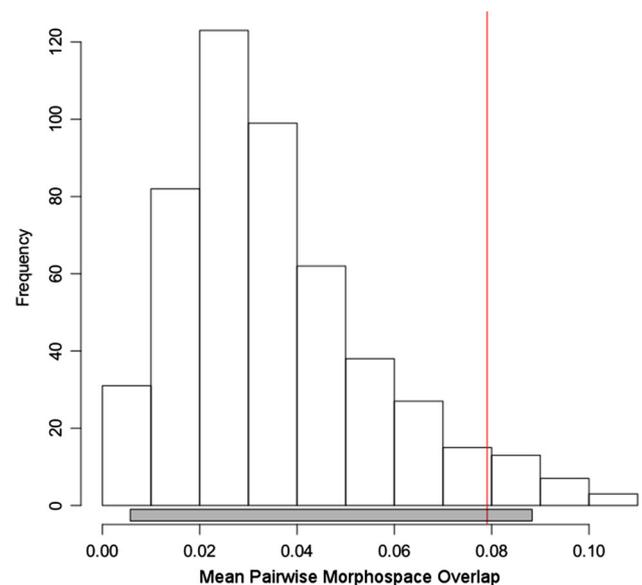
B&W fall between the estimates for the extinction extremes: Labridae, Blenniidae, Serranidae, Syngnathidae, Epinephelidae, Tetraodontidae, Haemulidae, Chaetodontidae, Monacanthidae, Pentacerotidae, Eleotridae, Clinidae, Lutjanidae, and Labrisomidae. These groups deserve closer inspection in future studies as they may potentially exhibit exceptional diversity for their age.

Gobies do exhibit exceptional richness, and although not commonly recognized as a ‘reef fish family,’ they are one of the most species rich families on reefs with 401 species found on reefs worldwide according to FishBase (Fig. 3).

**Table 3** Median rarefied morphospace overlap between families calculated on PC1 and PC2

Family	Median rarefied overlap
Labridae	0.041
Chaetodontidae	0.000
Gobiidae	0.004
Blenniidae	0.000
Apogonidae	0.000
Epinephelinae	0.025
Serranidae	0.026
Pomacentridae	0.007
Acanthuridae	0.000
Balistidae	0.002
Lutjanidae	0.002
Callionymidae	0.000
Carangidae	0.042
Cirrhitidae	0.000
Haemulidae	0.000
Hemiramphidae	0.000
Holocentridae	0.000
Lethrinidae	0.000
Monacanthidae	0.000
Mullidae	0.000
Nemipteridae	0.000
Ostraciidae	0.000
Pempheridae	0.000
Pinguipedidae	0.000
Platycephalidae	0.000
Pomacanthidae	0.000
Pseudochromidae	0.000
Scorpaenidae	0.001
Siganidae	0.000
Sparidae	0.000
Synanceiidae	0.000
Tetraodontidae	0.000
Tripterygiidae	0.000

Contrasting sources suggest the number may be much higher, with up to one half of all gobies ~1000 species found on reefs (Herler et al. 2011). Our findings are consistent with an earlier analysis that identified a diversification rate shift on the branch leading to Gobiidae (Near et al. 2013). Gobies are small, generally less than 10 cm in length, often cryptic, and found in almost every reef microhabitat, as well as many other marine and some freshwater environments (Herler et al. 2011). Work is needed on their evolutionary history but one can speculate that their small size and the associated short generation times may play a role in the high diversification rates. To

**Fig. 5** Body shape morphospace as described by PC1 and PC2 for each acanthomorph family (gray dots) with the eight most species rich reef families delineated by their minimum area convex polygons. The morphospace is illustrated with photos of fish that represent body shapes typical in each region of morphospace**Fig. 6** Histogram of rarefied mean pairwise morphospace overlap on PC1 and PC2 between eight randomly sampled acanthomorph families with the 95 % confidence interval indicated by the gray bar. The eight most species rich reef families were excluded from the sampling; their rarefied mean pairwise overlap is indicated by the vertical red line

determine whether reef environments were important drivers of elevated diversification rates in gobies requires a comprehensive time-calibrated phylogeny, which is currently not available.

Interestingly, several reef families appear to be unusually species poor, including the monotypic families: Zanclidae, Menidae, and Triodontidae, as well as Aulostomidae, which includes three extant species but is old, diverging from their sister-taxon in the Cretaceous. With the exception of zanclids, these families have a fairly rich fossil record. Triodontidae is a tetraodontiform and the extant genus *Triodon* has a fossil record stretching back to the Eocene where it has greater species diversity (Santini and Tyler 2003). Menidae have an extensive fossil record starting in the Paleocene and were once widespread and more diverse, with at least 16 named species (Friedman and Johnson 2005). Because of the hard lower boundary on diversity (there must be at least one species), most of these families only require one or two additional species to no longer be considered species poor; therefore, a more detailed understanding of the fossil record of these and other acanthomorph groups may help to confirm the identification of these lineages as depauperate. Finally, while it may be tempting to suggest that because the majority of acanthomorph families fall within the expected species richness, there is little evidence of diversification rate heterogeneity across Acanthomorpha. However, with unknown extinction rates and not including the possibility of time or diversity dependence in rates, it is too early to conclude this with confidence.

There are many reasons why we may expect to see a positive relationship between body shape disparity and lineage diversity (as reviewed by Rabosky et al. 2012). The same conditions that promote speciation: high productivity, biological and physical complexity, and a high degree of biological interactions may also promote morphological diversification, especially within traits that promote niche differentiation. Moreover, morphological versatility may actually promote speciation (Vermeij 1973a, b) by allowing the clade to more rapidly exploit ecological opportunities. However, we found no evidence of a significant relationship between species richness and body shape disparity across reef acanthomorphs. Taken at face value, this result suggests that speciation and body shape evolution have not been tightly linked during reef fish diversification. Our result differs from the recent finding that speciation rates and rates of body size evolution are correlated across ray-finned fishes (Rabosky et al. 2013). While our study focuses on body shape evolution rather than body size, there are also substantial differences between these studies in taxonomic coverage and the methods used. We employed models that assume a simple, time-homogenous birth–death process, whereby diversity and disparity increase through time, while Rabosky et al. (2013) used models that allow rates to change over time. Our result should also be treated with caution, as we were unable to correct for crown age, thus we resorted to using stem age.

Under certain circumstances, using stem age can be misleading when looking at the relationship between species diversity and age (Stadler et al. 2014) and disparity does not begin to accumulate until the first speciation event. Therefore, the relationship between speciation and body shape evolution among reef acanthomorphs will continue to be a problem that benefits from the development of a much more comprehensive, time-calibrated phylogeny of acanthomorphs.

We identify eight clades of acanthomorphs that are particularly diverse on reefs, all of which have more than 100 reef-dwelling species worldwide: Labridae, Gobiidae, Chaetodontidae, Pomacentridae, Epinephelidae, Serranidae, Blenniidae, and Apogonidae. Together these clades occupy much of the reef associated acanthomorph morphospace, except for the extremes that have only been reached by a few families: deep and elongate body shape or exceptionally small dorsal and anal fins (Fig. 5). The rarefaction analyses of 33 major reef fish families reveal that the median morphospace overlap between each family and every other family is zero. In other words, the vast majority of reef families occupy fairly distinct regions of morphospace. To the extent that morphological similarity reflects ecological similarity, this pattern is consistent with a history of strong ecological interactions between families during the invasion of reefs and subsequent diversification, leading to families evolving within relatively unoccupied regions of morphospace. With this general pattern of unique morphospace occupation by reef families as the predominant backdrop, it is especially interesting to find that the eight most species-rich reef families overlap in body shape morphospace to an unusually high degree even after rarefaction. The most species-rich reef clades occupy more similar regions of morphospace than do reef families generally (Fig. 6). Average body size does differ considerably among these groups, which can have a huge influence on ecology but is excluded from our geometric morphometric shape data. Most of the overlap is driven by three families: Labridae, Pomacentridae, and Serranidae (in our data set, serranids are represented entirely by members of Anthiinae), which, along with carangids, are the only groups to exhibit greater than 1 % median morphospace overlap with every other family in the morphological dataset (see Table 3). What is special about these groups? These three reef families occupy a central position within the acanthomorph shape-space described by PC1 and PC2 (Fig. 5). The other diverse reef families occupy more peripheral positions within the morphospace: blennies and gobies are more elongate, whereas apogonids and chaetodontids are more deep bodied. Thus, while most reef fish families appear to have found success by dominating relatively private regions of morphospace, some of the most successful groups show the opposite pattern. Labrids,

pomacentrids, and serranids did not find success by invading novel regions of morphospace, they radiated in the most densely packed regions of morphospace, where ecological interactions with other species might be expected to be strongest.

Although the new robust phylogenies of acanthomorphs and teleost fishes have allowed us to gain unique insights into the history and diversification of reef fish, there are still many unanswered questions. In particular, the lack of well-supported crown ages for many reef fish families is currently hindering our exploration of the macroevolutionary patterns and processes that have influenced the evolution of the modern reef fish fauna.

**Acknowledgments** We thank Jack Randall for making available his collection of photographs of Indo-Pacific fishes. This research was supported by NSF Grants IOS-0924489, DEB-1061981.

## References

- Adams DC, Otarola-Castillo E (2013) geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol Evol* 4:393–399
- Alfaro M, Santini F, Brock CD (2007) Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). *Evolution* 61:2104–2126
- Bellwood DR (1996) The Eocene fishes of Monte Bolca: the earliest coral reef fish assemblages. *Coral Reefs* 15:11–19
- Bellwood DR (1998) What are reef fishes? - Comment on the report by D. R. Robertson: Do coral-reef fish faunas have a distinctive taxonomic structure? *Coral Reefs* 17:187–189
- Bellwood DR (2003) Origins and escalation of herbivory in fishes: a functional perspective. *Paleobiology* 29:71–83
- Bellwood DR, Wainwright PC (2002) The history and biogeography of fishes on coral reefs. In: Sale PF (ed) *Coral reef fishes: diversity and dynamics in a complex ecosystem*. Academic Press, Boston, pp 5–32
- Betancur RR, Broughton RE, Wiley EO, Carpenter K, Lopez JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton JCI, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC, Hough DJ, Lu G, Grande T, Arratia G, Orti G (2013) The tree of life and a new classification of bony fishes. *PLoS Currents* 5: ecurrents.tol.53ba26640df0cca ee75bb165c8c26288
- Bivand R, Lewin-Koh N (2014) maptools: Tools for reading and handling spatial objects. R package version 0.8-30. <http://CRAN.R-project.org/package=maptools>
- Bivand R, Rundel C (2014) rgeos: Interface to Geometry Engine - Open Source (GEOS). R package version 0.3-6. <http://CRAN.R-project.org/package=rgeos>
- Boettiger C, Lang DT, Wainwright PC (2012) rfishbase: exploring, manipulating and visualizing FishBase data from R. *Fish Biol* 81:2030–2039
- Bollback JP (2006) SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88
- Calenge C (2006) The package adehabitat for the R software: a tool for the analysis of space and habitat use by animals. *Ecol Model* 197:516–519
- Claverie T, Wainwright PC (2014) A morphospace for coral reef fishes: elongation is the dominant axis of body shape evolution. *PLoS One* 9:e112732
- Cowman PF, Bellwood DR (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *J Evol Biol* 24:2543–2562
- Eschmeyer WN (2012) *Catalog of fishes*. California Academy of Sciences (<http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>)
- Friedman M, Johnson G (2005) A new species of Mene (Perciformes: Menidae) from the Paleocene of South America, with notes on paleoenvironment and a brief review of menid fishes. *J Vert Paleontol* 25:770–783
- Froese R, Pauly D (2013) FishBase. World Wide Web electronic publication. <http://www.fishbase.org>
- Harmelin-Vivien ML (2002) Energetics and fish diversity on coral reefs. In: Sale PF (ed) *Coral reef fishes. Dynamics and diversity in a complex ecosystem*. Academic Press, Boston, pp 265–274
- Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* 24:129–131
- Herler J, Munday PL, Hernaman V (2011) Gobies on coral reefs. In: Patzner RA, Van Tassell JL, Kovacic M, Kapoor BG (eds) *The biology of gobies*. Science Publishers, New Hampshire, pp 493–539
- Kiessling W, Simpson C, Foote M (2010) Reefs as cradles of evolution and sources of biodiversity in the Phanerozoic. *Science* 327:196–198
- Magallón S, Sanderson MJ (2001) Absolute diversification rates in angiosperm clades. *Evolution* 55:1762–1780
- Near TJ, Dornburg A, Eytan RI, Keck BP, Smith WL, Kuhn KL, Moore JA, Price SA, Burbrink FT, Friedman M, Wainwright PC (2013) Tempo of diversification in the superradiation of spiny-rayed fishes. *Proc Natl Acad Sci* 110:12738–12743
- Nelson JS (2006) *Fishes of the World*. John Wiley and Sons, Hoboken, New Jersey
- Nielsen R (2002) Mapping mutations on phylogenies. *Syst Biol* 51:729–739
- Nunes F, Norris RD (2006) Abrupt reversal in ocean overturning during the Palaeocene/Eocene warm period. *Nature* 439:60–63
- O’Meara BC, Ane C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60:922–933
- Orme D, Freckleton R, Thomas G, Petzoldt T, Fritz S, Isaac N, Pearse W (2013) caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. <http://CRAN.R-project.org/package=caper>
- Pagel M (1997) Inferring evolutionary processes from phylogenies. *Zool Scripta* 26:331–348
- Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20:289–290
- Perrin C (2002) Tertiary: the emergence of modern reef ecosystems. In: Kiessling W, Flügel E, Golonka J (eds) *Phanerozoic reef patterns*. SEPM (Society for Sedimentary Geology), Tulsa, pp 587–621
- Price SA, Holzman RA, Near TJ, Wainwright PC (2011) Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecol Lett* 14:462–469
- Price SA, Tavera J, Near TJ, Wainwright PC (2012) Elevated rates of morphological and functional diversification in reef-dwelling haemulid fishes. *Evolution* 67:417–428
- Price SA, Schmitz L, Oufiero C, Eytan RI, Dornburg A, Smith WL, Friedman M, Near TJ, Wainwright PC (2014) Two waves of

- colonization straddling the K–Pg boundary formed the modern reef fish fauna. *Proc R Soc Lond B Biol Sci* 281:20140321
- R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>
- Rabosky D (2010) Primary controls on species richness in higher taxa. *Syst Biol* 59:634–645
- Rabosky D, Slater G, Alfaro M (2012) Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biol* 10:e1001381
- Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME (2013) Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat Commun* 4:1958
- Revell LJ (2011) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3:217–223
- Ricklefs RE (2009) Speciation, extinction and diversity. In: Butlin RK, Bridle JR, Schluter D (eds) *Speciation and patterns of diversity*. Cambridge University Press, Cambridge, pp 257–277
- Robertson DR (1998) Do coral-reef fish faunas have a distinctive taxonomic structure? *Coral Reefs* 17:179–186
- Santini F, Tyler JC (2003) A phylogeny of the families of fossil and extant tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), upper Cretaceous to recent. *Zool J Linn Soc* 139:565–617
- Schoener TW (1974) Resource partitioning in ecological communities. *Science* 185
- Smith WL, Craig MT (2007) Casting the percomorph net widely: the importance of broad taxonomic sampling in the search for the placement of serranid and percoid fishes. *Copeia* 1:35–55
- Stadler T, Rabosky DL, Ricklefs RE, Bokma F (2014) On age and species richness of higher taxa. *Am Nat* 184:447–455
- Vermeij GJ (1973a) Biological versatility and earth history. *Proc Natl Acad Sci* 70:1936–1938
- Vermeij GJ (1973b) Adaptation, versatility, and evolution. *Syst Zool* 22:466–477
- Zachos JC, Rohl U, Schellenberg SA, Sluijs A, Hodell DA, Kelly DC, Thomas E, Nicolo M, Raffi I, Lourens LJ, McCarren H, Kroon D (2005) Rapid acidification of the ocean during the Paleocene-Eocene thermal maximum. *Science* 308:1611–1615