ECOLOGY LETTERS

Ecology Letters, (2015) **18**: 441–450 doi: 10.1111/ele.12423

LETTER

Fossil-based comparative analyses reveal ancient marine ancestry erased by extinction in ray-finned fishes

Ricardo Betancur-R,[1,2]* Guillermo Ortí[3] and Robert Alexander Pyron[3]

Abstract

The marine-freshwater boundary is a major biodiversity gradient and few groups have colonised both systems successfully. Fishes have transitioned between habitats repeatedly, diversifying in rivers, lakes and oceans over evolutionary time. However, their history of habitat colonisation and diversification is unclear based on available fossil and phylogenetic data. We estimate ancestral habitats and diversification and transition rates using a large-scale phylogeny of extant fish taxa and one containing a massive number of extinct species. Extant-only phylogenetic analyses indicate freshwater ancestry, but inclusion of fossils reveal strong evidence of marine ancestry in lineages now restricted to freshwaters. Diversification and colonisation dynamics vary asymmetrically between habitats, as marine lineages colonise and flourish in rivers more frequently than the reverse. Our study highlights the importance of including fossils in comparative analyses, showing that freshwaters have played a role as refuges for ancient fish lineages, a signal erased by extinction in extant-only phylogenies.

Keywords

Actinopterygii, diversification, ecological transitions, marine and freshwaters, neontology, palaeontology, phylogenetic comparative methods, phylogeny, state dependent diversification.

Ecology Letters (2015) **18**: 441–450

INTRODUCTION

The boundary between marine and fresh waters constitutes a major gradient in global diversity (Lee & Bell 1999). While some species can move easily between these habitats, most lack adaptations to overcome gradients in salinity, temperature and microbiome (Logares *et al*. 2009). Among animals, ecological transitions between aquatic environments are limited to a number of taxa (e.g. annelids, mollusks, crustaceans and fishes) and have typically occurred over long timescales (Lee & Bell 1999). As a consequence, diversity of clades and patterns of community assemblage vary widely along gradients of habitat space. But, do differences in species richness and composition between marine and freshwaters imply that these habitats affect rates of diversification? To examine whether diversification patterns can be explained as a function of habitat occupancy it is necessary to disentangle the confounding effects of speciation, extinction and dispersal using comparative phylogenetic methods (Maddison 2006).

Actinopterygii (ray-finned fishes) offers a unique opportunity to test these ideas given their extraordinary diversity in marine and freshwater environments, and the exceptional fossil record of early-branching lineages. Ray-finned fishes comprise *ca*. 32 000 extant species (nearly 50% of all species of vertebrates), with similar levels of richness in both habitat types (Table 1). Studies investigating the colonisation of freshwaters by marine families (e.g. herrings, sea catfishes, silversides, grunters, sticklebacks and pufferfishes) suggest that ecological opportunity in the novel freshwater niche is often limited by high levels of incumbent freshwater fish diversity, but that marine-derived lineages undergo accelerated speciation when rivers are depauperate and thus have greater open niche spaces (Betancur-R *et al*. 2012; Davis *et al*. 2012; Bloom *et al*. 2013; Santini *et al*. 2013; Bloom & Lovejoy 2014).

A single study inferred diversification rates for freshwater and marine actinopterygians overall, failing to detect significant differences between habitat states (Vega & Wiens 2012). This study also suggested that ancestral actinopterygians inhabited freshwaters, a pattern consistent with the contemporary occurrence in rivers of early-branching lineages (e.g. bichirs, bowfins, gars and bonytongues). These conclusions may have been affected, however, by limited phylogenetic sampling (based only on RAG1 sequences for 124 actinopterygian species), limited incorporation of fossil data, and lack of comparative approaches that account for state-dependence.

The fossil record and phylogeny of ray-finned fishes suggest complex diversification dynamics, potentially explaining disparity in clade richness by differential lineage persistence in marine and freshwater habitats. For instance, most early-branching lineages are currently restricted to freshwater environments, but abundant fossil evidence indicates that many of these lineages also inhabited marine environments. Fossils of bowfins, bonytongues and other early lineages now extinct have been reported from numerous marine deposits (Grande & Bemis 1998; Hilton 2003). Thus, the persistence of ancient lineages in freshwaters but not in marine environments today

[1]*Department of Biology, University of Puerto Rico – Río Piedras, P.O. Box 23360, San Juan, Puerto Rico, 00931*
[2]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC, 20013, USA*
[3]*Department of Biological Sciences, The George Washington University, 2023 G St. NW, Washington, DC, 20052,USA*
**Correspondence: e-mail: betanri@fishphylogeny.org*

Table 1 Summary of species richness by habitat in ray-finned fishes compiled from the Catalog of Fishes (CoF; Eschmeyer & Fong 2013). Euryhaline species are species coded as occurring in both marine and freshwater environments. Brackish species are a separate category in the CoF (8.5% of the species), which we lump under alternative scenarios and gauge sensitivity of analyses to coding ambiguity. Values in parenthesis indicate sampled diversity in the molecular phylogeny (total number, percentage)

States	Richness BAM	Richness BAF	Richness BEAM	Richness BEAF
Freshwater	15 146 (278, 1.8%)	15 248 (278, 1.8%)	15 146 (278, 1.8%)	15 964 (374, 2.3%)
Marine	15 445 (1208, 7.8%)	15 343 (1208, 7.9%)	16 161 (1304, 8.1%)	15 343 (1208, 7.9%)
Euryhaline	716 (96, 13.4%)	716 (96, 13.4%)		
Total	31 307 (1582, 5.1%)	31 307 (1582, 5.1%)	31 307 (1582, 5.1%)	31 307 (1582, 5.1%)

BAM, brackish as marine (GeoSSE); BAF, brackish as freshwater (GeoSSE); BEAM: brackish and euryhaline as marine (BiSSE, SIMMAP); BEAF, brackish and euryhaline as freshwater (BiSSE, SIMMAP).

may be the result of widespread marine extinction (Parenti 2008).

Colonisation of freshwaters by marine lineages also appears more common than the reverse, another pattern long noted – though never explicitly quantified – by ichthyologists. Multiple marine-derived fish orders or families have colonised freshwaters in recent times (e.g. herrings and allies, toadfishes, gobies, flatfishes, cichlids, silversides, mullets, drums, sticklebacks and pufferfishes; Bloom & Lovejoy 2012; Davis *et al*. 2012; Bloom *et al*. 2013; Santini *et al*. 2013). In contrast, only a few groups of primary freshwater fishes have successfully diversified in the sea (e.g. sea catfishes and eeltail catfishes; McDowall 1997; Parenti 2008; Betancur-R. 2010; Betancur-R *et al*. 2012).

We take a unified neontological and paleontological approach to estimate ancestral habitats, characterise diversification rates, and quantify habitat transitions throughout the evolutionary history of ray-finned fishes. Our analyses leverage recent advances in our understanding of the Tree of Life of fishes (Betancur-R. *et al*. 2013; Near *et al*. 2013), adding extensive analyses of the fossil record in a comparative framework. Based on the above observations, we predict that: (1) extinction has erased the signal of marine ancestry among early-branching ray-finned fish lineages and therefore ancestral habitat reconstructions excluding fossil taxa may produce misleading results; (2) diversification dynamics (net diversification and turnover rates) vary as a function of habitat occupancy, driving present-day diversity patterns; (3) marine-to-freshwater transitions are more frequent than freshwater-to-marine transitions. Our analyses support these predictions, suggesting that younger ray-finned fish clades outcompete and replace older lineages in the marine realm, while freshwater habitats have served as refuges for older lineages. Our results highlight the impact that including fossil species can have on comparative approaches (Slater *et al*. 2012, 2013), even when applying methods that attempt to account for ancestral extinction dynamics explicitly.

MATERIALS AND METHODS

Molecular phylogeny

The phylogenetic framework for our comparative analysis is based on a recent large-scale molecular study for 1407 ray-finned fish species, containing DNA sequence data from 20 nuclear and one mitochondrial markers (Betancur-R. *et al*. 2013). We augmented this sampling with nuclear sequence data for 175 additional species from another study (Near *et al*. 2013). We then analysed the concatenated dataset using Maximum Likelihood (ML) in RAxML v7 (Stamatakis 2006). We used a 24-partition scheme (i.e., a combination of codon positions and genes that best-fit the heterogeneity in the data) and 30 independent replicates under the GTRGAMMA model. We assessed branch support using 1000 replicates obtained with RAxML's rapid bootstrapping algorithm. To overcome the computational burden of analysing large phylogenetic datasets in a Bayesian framework, we time-calibrated the resulting RAxML tree under penalised likelihood in treePL (Smith & O'Meara 2012) using 126 secondary calibrations. These secondary calibrations (fixed mean nodal ages) were obtained from a Bayesian analysis of a subset of 201 taxa with 61 fossil age constraints (primary calibrations) using node-dating approaches. Further details on alignment, concatenation, ML inference, partitioning schemes, fossil calibrations, and divergence time estimates are given in the original study (Betancur-R. *et al*. 2013).

Time-tree with extant and fossil taxa

In addition to the molecular phylogeny of extant species, we compiled from the literature a large collection of trees inferred from morphological data matrices of fossil and extant species, encompassing the diversity of ray-finned fishes. We obtained credible estimates of phylogenetic placement for a total of 240 fossil taxa from 30 studies (see details in Appendix SA and Table SB1). While most studies emphasise taxonomic sampling on early-branching groups (e.g. early actinopterygians, chondrosteans, amiids and osteoglossomporphs), there is also significant coverage of crown acanthomorph clades (e.g. zeiforms, pleuronectiforms, acathuriforms and tetraodontiforms). Trees were combined, reconciled, and placed onto the molecular phylogeny using a supertree approach (Appendix SA), as implemented in the program SuperFine v1 (Swenson *et al*. 2012).

We time-scaled the resulting supertree with 1822 taxa using the R package Paleotree (Bapst 2013). We used the minimum branch length (mbl) method of the timePaleoPhy routine in Paleotree, which scales branches so they are greater than or equal to 'vartime' variable. We also imposed minimum ages from all nodes obtained in the treePL tree with 1582 extant species via the node.mins argument of timePaleoPhy. To assess sensitivity of downstream comparative analyses to

branch scaling, we generated time-trees under three different values of vartime (0.1, 0.5 and 1.0; D. Bapst pers. comm.). Larger values of vartime resulted in old root divergences and were thus excluded (e.g., a vartime of 10 yielded a root age of 740 Ma; see details on Table SB2).

Habitat coding

We compiled habitat data from all known, extant species of ray-finned fishes (31307 as of 2013) from the Catalog of Fishes (CoF; Eschmeyer & Fong 2013). We also obtained habitat information from the paleontological literature for fossils placed in the phylogeny (Table SB1). Species with broad habitat distributions that may be tolerant of wide ranges of salinity are considered euryhaline in a biogeographic sense (following Lee & Bell 1999), without discriminating among variable life histories associated with salinity preferences (e.g., diadromy, catadromy, anadromy, amphidromy; McDowall 1997; Parenti 2008). The habitat states obtained from the CoF are freshwater (F), marine (M), euryhaline (E – both F and M) and brackish (B – not coded as a separate state, but see below).

Habitat information was unavailable for some species (mostly extinct taxa), and we assessed sensitivity of analyses to coding ambiguity. For model-based approaches that allow for intermediate (euryhaline) states (e.g. GeoSSE; Goldberg *et al*. 2011), brackish species and species with unknown habitats were coded as either marine (BAM) or freshwater (BAF). Other methods require binary data (e.g. BiSSE; Maddison *et al*. 2007), in which case we coded both euryhaline and brackish taxa as either marine (BEAM) or freshwater (BEAF). Table 1 summarises species richness by habitat under different coding scenarios.

Ancestral habitat reconstructions (prediction 1)

To test the prediction that extinction has erased the signal of marine ancestry in the ray-finned fish phylogeny, we conducted ancestral state (habitat) reconstructions (ASRs) using the two trees (including and excluding fossils). We expect ASRs to differ substantially in these trees, in particular among early-branching lineages. We assume that trees containing more fossils (i.e., direct observations of past habitats) will yield more accurate results on average, and examine this assumption in the light of the results (see below). We compared the ASR results under various methods and habitat coding scenarios (BEAM and BEAF) using the R packages Phytools v0.4-21 (Revell 2012), diversitree v0.9 (FitzJohn 2012), and Ape (Paradis *et al*. 2004).

When rates of speciation, extinction, and state transition are known to differ, conducting ASR in a ML framework assuming equal rates constitutes a major violation of the models (e.g., Maddison 2006; Goldberg & Igic 2008; Pyron & Burbrink 2014). In such cases, it is necessary to infer ancestral habitats using a model that accounts for state-dependent diversification. However, this is currently not possible for the time-calibrated fossil-based tree due to algorithmic limitations, as non-ultrametric trees would demand substantial mathematical reworking of current models. We thus conducted joint ASRs using both stochastic character mapping (SIMMAP, implemented in Phytools; Bollback 2006) and maximum likelihood (implemented in Ape's function Ace). While these approaches assume fixed rates of speciation and extinction across states, we used an asymmetric model (MK2) that allows transition rates to vary. We also tested the robustness of ASRs under ML using the simpler MK1 (symmetric) model (see below), but reasons for preferring MK2 are given on Appendix SG. We generated 3000 SIMMAP replicates for each habitat state/tree and estimated the posterior probability for nodes by averaging state frequencies across replicates.

For comparison, we also obtained ASRs using the time-tree with extant taxa under the binary state speciation and extinction model (BiSSE), which accounts for state-dependent diversification. An important advantage of the BiSSE model implemented in diversitree is that it can take complete yet terminally unresolved phylogenies by specifying sampling fractions (FitzJohn *et al*. 2009). Thus, our ASR BiSSE estimates are based on habitat states for all 31307 extant ray-finned fish species, not just the 1582 tips in the extant-only tree (see sampling fractions in Table 1). The BiSSE reconstructions were conducted in a Bayesian framework (see details below).

Finally, we assessed sensitivity of ASRs to other type of analyses (e.g. ML-MK1 vs. ML-MK2), alternative scaling of fossil branches in Paleotree (vartime values 0.0, 0.5, and 1.0), and alternative tree topologies (six additional trees obtained with RAxML search replicates). Analyses of ASR sensitivity are reported in Appendix SC.

State-dependent diversification (predictions 2 and 3)

We implemented models that account for state dependent diversification (SSE) to assess whether net diversification (speciation – extinction) and turnover (extinction/speciation) rates differ among marine and freshwater clades (prediction 2) and whether marine-to-freshwater transitions (qMF) are more frequent than freshwater-to-marine transitions (qFM; prediction 3).

We used GeoSSE (Goldberg *et al*. 2011), an extension of the BiSSE model that allows lineages to occur simultaneously in two habitats (i.e. euryhaline species). Because the implementation of BiSSE in diversitree includes features that are unavailable for GeoSSE (e.g., uneven sampling fractions), we also complemented and compared these analyses using BiSSE, gauging sensitivity of both analyses to coding ambiguity (BAM, BAF, BEAM and BEAF; Table 1).

As mentioned above, SSE analyses are only possible with the molecular time-tree including 1582 terminals. This tree comprises ca. 5% of all extant ray-finned fish species and has an overrepresentation of marine taxa (Table 1). We thus defined sampling fractions for both GeoSSE and BiSSE, assuming evenly distributed sampling across clades (Table 1). We also ran BiSSE under uneven sampling by defining sampling fractions across 50 clades (Table SD1), to account for sampling biases.

We first fit a series of models accounting for state dependent diversification using both GeoSSE and BiSSE. The GeoSSE model implements parameters for speciation in

freshwater, marine, and euryhaline taxa, extinction in freshwater and marine taxa, and marine-to-freshwater dispersal and vice versa (seven parameters). The BiSSE model is similar, except that it lacks the ability to account for speciation in euryhaline taxa (six parameters). We conducted model testing in a ML framework and selected the best-fit models using the ΔAIC and Akaike weights (AW) statistics. Details for ten GeoSSE and nine BiSSE models tested are given in Table SD2.

For the best-fit models, we integrated over parameter uncertainty using 10 000 generations sampled from an exponential prior (1/2 * the state-independent diversification rate *r*), with 10% of the initial samples discarded as burn-in. We summarised the posterior distributions of samples to assess variation in net diversification, turnover, and transition (qMF, qFM) rates across habitats. As predicted, inclusion or exclusion of fossil taxa may affect the ASRs, which may further bias other parameter estimates (e.g. speciation and extinction rates). We thus repeated the SSE analyses on a nested clade (Clupeocephala; ca. 130 Ma younger than Actinopterygii) that retains a significant proportion of the extant diversity of ray-finned fishes (ca. 96% of the species) but includes fewer ambiguous nodes (see Results). Finally, we compared the results obtained for transition rates (prediction 3) using SSE with estimates derived from the ASRs using SIMMAP (see above).

Clade-based analyses (additional tests of prediction 2)

We also tested for individual differences in freshwater and marine clades, to assess further whether diversification dynamics varied across habitats. We estimated stem and crown ages and calculated species richness for 67 non-nested clades, based on the time-tree of extant taxa. Most of these clades represent family-, ordinal-, or supraordinal-level groups in the recent bony fish classification (Betancur-R. *et al*. 2013). Other clades are not formally classified but have repeatedly been obtained in recent large-scale multi-locus analyses (Betancur-R. *et al*. 2013; Near *et al*. 2013) (Table SF1).

We first extracted the state probabilities from the ASRs, averaging across results from alternative models and coding scenarios, and discretized habitat states for clades with low ambiguity in nodal reconstructions. This resulted in 20 freshwater (freshwater probability > 0.9) and 35 marine clades (freshwater probability < 0.1); 12 remaining clades had ambiguous state probabilities (> 0.1 and < 0.9) and were excluded (Table SF1).

We then tested for a significant relationship between clade age (both crown and stem) and diversity. A positive age–diversity relationship is expected under a constant birth–death process (McPeek & Brown 2007; Pyron & Burbrink 2012; Pyron & Wiens 2013), but this has proved rare for stem-group ages in comparative datasets (Rabosky *et al*. 2012; Stadler *et al*. 2014). If diversity shows a positive relationship with crown ages, but an insignificant or negative relationship with stem ages, it may be an indication that younger clades outcompete older clades as extinction prunes lineages from the stem, driving the disconnect between stem ages and richness through time (Pyron & Burbrink 2012; Pyron & Wiens 2013).

We also calculated net diversification rates for the 55 target clades using both stem and crown equations of the method-of-moments (Magallon & Sanderson 2001; implemented in the R package Geiger, Harmon *et al*. 2008). We used the relative extinction fractions (e) estimated with BiSSE/GeoSSE for freshwater (e = 0.53) and marine (e = 0.30) lineages, and assessed confidence limits under arbitrarily high (e = 0.90) and low (e = 0.0) extinction fractions (Magallon & Sanderson 2001). We then estimated mean diversification rates for each habitat and extinction fraction, and assessed statistical significance among treatments using the non-parametric Mann–Whitney *U*-test. Under a constant birth–death process, both crown and stem equations are expected to yield similar values of net diversification. However, higher extinction rates in stem lineages may result in lower stem-based estimates, as more lineages are pruned from the stem branch.

A recent study showed that decoupling of clade age and species richness may arise from how higher taxa are defined, rather than clade-specific patterns resulting from ecological limits on diversity (Stadler *et al*. 2014). For instance, young taxa are often identified as exceptionally rich due to the 'pull of the present' effect (Stadler 2011). To standardise our comparisons, we thus repeated all clade-based analyses for a subset of the clades whose crown age is 80–50 Ma (9 freshwater and 19 marine clades; Table SF1), with the caveat that this may reduce power.

RESULTS

The new molecular phylogenetic matrix includes data from 1582 extant species and 21 genes (available from FigShare http://dx.doi.org/10.6084/m9.figshare.1297403). Matrix completeness (number of sequences per taxon) across the concatenated alignment is 43%, and all species include sequences from at least three genes, with a mean of nine. The RAxML trees estimated with the new dataset (including optimal and suboptimal replicates) are all highly congruent with the previous study (Betancur-R. *et al*. 2013). The placement of 240 fossils in the time-calibrated supertree (1822 taxa) also provides a comprehensive summary of the current state of ray-finned fish paleontological systematics (see Fig. S1). Below we report results obtained to test each of the predictions.

Prediction 1 – Extinction has erased the signal of marine ancestry for early branching lineages

As predicted, ancestral states along the tree backbone inferred with the fossil-based tree (SIMMAP and ML reconstructions only) are substantially different from those using the time-tree with extant terminals (Figs. 1, SC1–SC4). In the extant-only tree, most early nodes have very high probabilities of freshwater occupancy. Conversely, analyses using the fossil-based tree reconstructed those nodes as marine with high probability. Most differences are in Triassic and Paleozoic (380–200 Ma) nodes. The more nested Clupeocephala clade, selected for assessing the robustness of downstream SSE analyses, however, includes fewer ambiguities in ancestral states (Fig. 1).

These results are generally robust to method (SIMMAP and ML) and model (MK1 and MK2) choice, scaling of fossil branches in Paleotree (Fig. SC2–SC4), and alternative tree topologies (six additional RAxML trees; Fig. SC5). While ASRs based on the BEAM coding resulted in nodes having

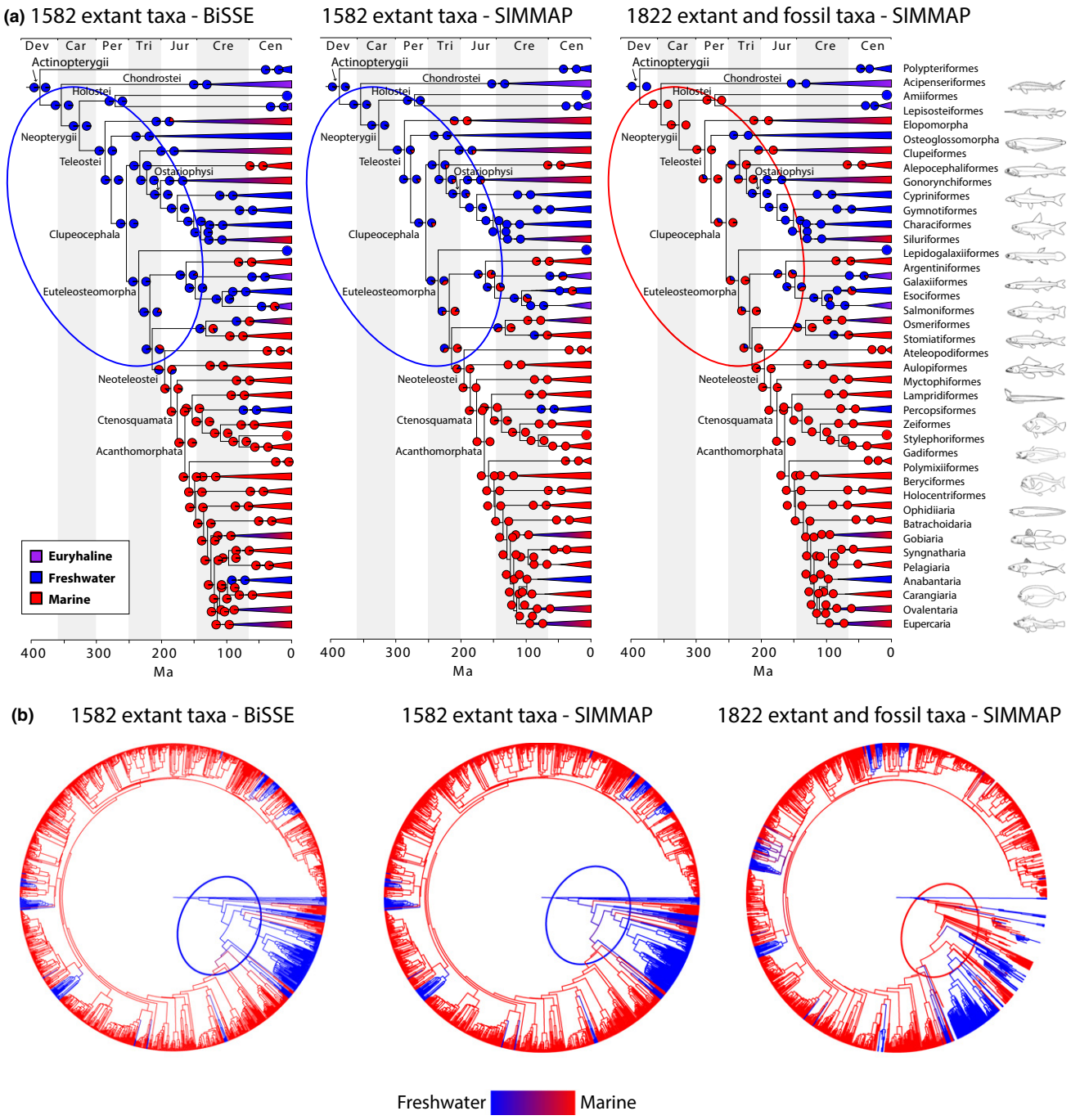

Figure 1 Ancestral state (habitat) reconstructions (ASR) including and excluding fossils. (a) Collapsed trees using alternative ASR models (BiSSE and SIMMAP) and coding scenarios (left pies BEAF – brackish and euryhaline as marine; right pies BEAM – brackish and euryhaline as freshwater). (b) Expanded trees under BEAM coding only (see also Fig. SC1), with colour gradients denoting state probability. Euryhaline species are not coded as intermediate states, but terminal clades including euryhaline taxa are depicted in (a). Ovals highlight major differences in ancestral habitat reconstructions among early-branching linages.

slightly higher probability of marine ancestry than those using BEAF (left and right pies in Fig. 1a), the reconstructions are largely robust to coding ambiguity. Ancestral states based on the tree with extant taxa reconstructed under BiSSE are similar to those using SIMMAP or ML. However, BiSSE nodes have lower probability of uncertainty and show fewer differences between coding scenarios. Taken together, these results support the prediction that extinction has confoundingly erased the real signal of early marine ancestry in Actinopterygii.

Prediction 2 – Diversification dynamics vary as a function of habitat occupancy

Despite similar levels of species richness in marine and freshwaters (ca. 16 000 in each habitat), a model in which speciation

and extinction rates are constant across habitats is strongly rejected by the SSE analyses in all six comparisons (ΔAIC = 26–68, AW = 9.9×10^{-7}–2.3×10^{-16}; Table 2). This prediction is tested on the basis of diversification analyses obtained with BiSSE (4 analyses: BEAM even, BEAF even, BEAM uneven, and BEAF uneven) and GeoSSE (2 analyses: BAM even and BAF even) as well as clade-based comparisons, using the time-tree with extant taxa.

This is contrary to previous results using clade-based approaches on a much less densely sampled phylogeny (Vega & Wiens 2012). The MCMC plots obtained from the SSE analyses using either the complete ray-finned fish tree (Fig. 2a) or the clupeocephalan subtree (Fig. SD1a) indicate substantially higher rates of lineage diversification for marine vs. freshwater lineages, respectively. Likewise, clade-based estimates on 20 freshwater and 35 marine clades show higher net diversification rates for marine than freshwater lineages (Table SF2). We obtained similar results for the 9 freshwater and 19 marine clades that have a crown age of 80–50 Ma, but these comparisons yielded marginal or no statistical significance presumably due to limited power (Table SF3).

Scatterplots of species richness and crown age resulted in a positive correlation for marine clades (P = 0.006–0.0001), suggesting constant species accumulation over time in marine environments (Figs. 3, SF1). Contrary to the expectations of a constant birth-death model, there is no significant relationship between crown age and diversity for freshwater clades (P = 0.43–0.76), and there is a negative (P = 0.01–0.03; Fig. 3) or non-existent (P = 0.12–0.55; Fig. SF1) relationship between stem age and richness for both freshwater and marine clades, respectively. These results suggest that the stem vs. crown disconnect is driven by greater extinction in marine than in freshwater clades, but see caveats below for conflicting estimates of clade turnover obtained with SSE and clade-based analyses.

Prediction 3 – Marine-to-freshwater transitions are more frequent than the reverse

As predicted, a model in which transitions rates are symmetric is rejected by the SSE analyses (ΔAIC = 4–74, AW = 5.3×10^{-2}–5.3×10^{-17}; Table 2), except for BEAM uneven

Table 2 Tests of SSE models where (1) speciation and extinction are fixed as equal across habitats (λM~λF; μM~μF) (see Vega & Wiens 2012) while transition rates are free to vary (qMF≠qFM), and (2) transition rates are fixed as equal (qMF~qFM) while speciation and extinction are free to vary (λM≠λF; μM≠μF). For both tests, ΔAIC and Akaike Weights (AW) are computed in comparison to a model in which constrained parameters are free to vary

Method	Sampling	Coding scenarios	λM~λF -μM~μF (ΔAIC)	λM~λF -μM~μF (AW)	qMF~qFM (ΔAIC)	qMF~qFM (AW)
BiSSE	Even	BEAF	58***	3.13E-13***	62***	2.83E-14***
BiSSE	Even	BEAM	59***	1.38E-13***	65***	8.39E-15***
GeoSSE	Even	BAF	69***	7.87E-16***	74***	9.27E-17***
GeoSSE	Even	BAM	68***	2.32E-16***	70***	5.31E-17***
BiSSE	Uneven (50 clades)	BEAF	26***	9.87E-07***	4*	0.05343871*
BiSSE	Uneven (50 clades)	BEAM	22***	8.39E-15***	2	1.71E-01

BAM, brackish as marine (GeoSSE); BAF, brackish as freshwater (GeoSSE); BEAM, brackish and euryhaline as marine (BiSSE); BEAF, brackish and euryhaline as freshwater (BiSSE); AW, Akaike weights.
*Marginally significant.
***Strongly significant.

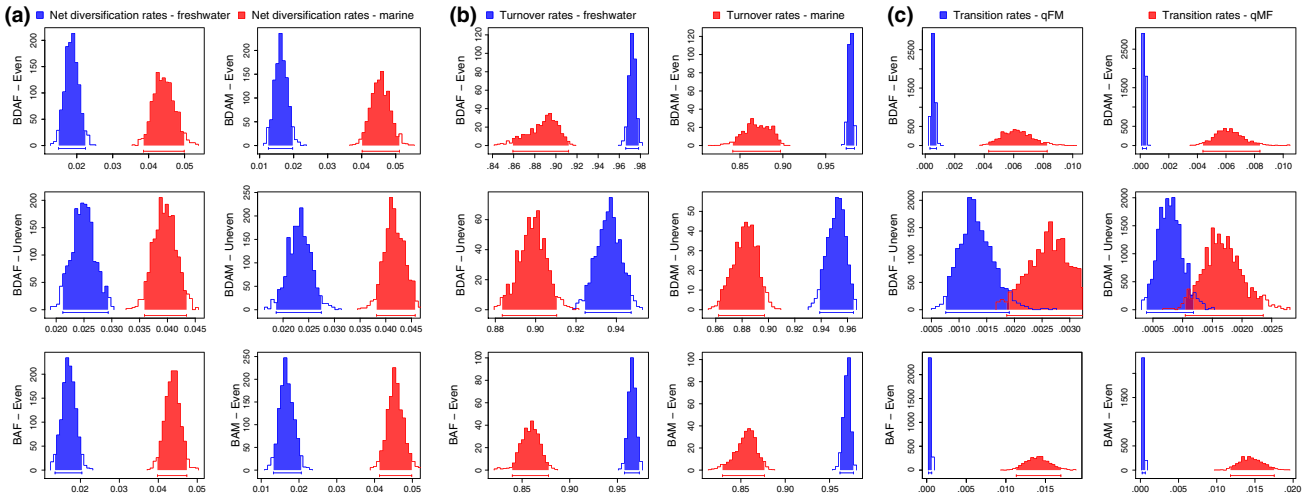

Figure 2 MCMC plots of (a) net diversification (speciation – extinction), (b) turnover (extinction/speciation), and (c) transition rates (qFM: freshwater-to-marine; qMF: marine-to-freshwater) across habitats in Actinopterygii. Sensitivity of analyses to model implementation (BiSSE and GeoSSE), habitat-coding scenarios, and sampling fraction definitions (even or uneven for 50 clades) is assessed. BAM: brackish as marine (GeoSSE); BAF: brackish as freshwater (GeoSSE); BEAM: brackish and euryhaline as marine (BiSSE); BEAF: brackish and euryhaline as freshwater (BiSSE).

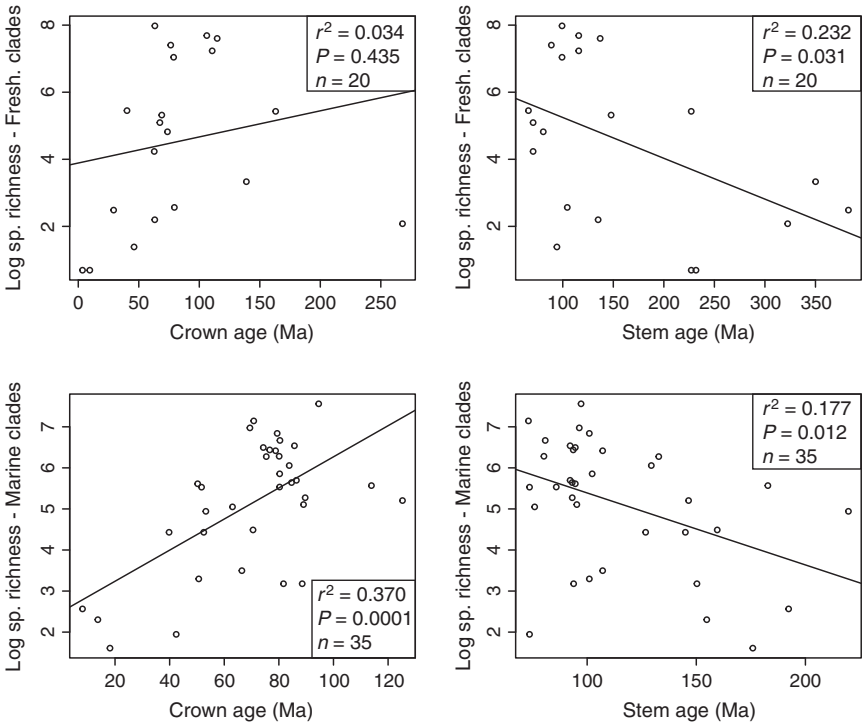

Figure 3 Correlations of (log) species richness against crown and stem ages, for 35 marine and 20 freshwater clades (see also Fig. SF1).

(ΔAIC = 2, AW = 0.17; Table 2). All analyses derived from the tree with extant taxa indicate that qMF (marine to freshwater transitions) are more frequent than qFM (freshwater to marine transitions; Table SE1; Fig. 2c). The magnitudes of transition ratios vary substantially (qMF/qFM = 2.0−31.6), depending on the model used (BiSSE, GeoSSE), sampling fraction schemes (even or uneven), and coding scenarios (BEAM, BEAF, BAM, BAF), but the directionality is consistent. The SIMMAP reconstructions using the tree with extant and fossil taxa (not possible with SSE) also indicate higher marine-to-freshwater than freshwater-to-marine transition ratios, but the differences are smaller (qMF/qFM = 1.2−1.4; Table SE1). Comparisons of the fit of the datasets (extant only and fossil-based) to MK1 and MK2 models assuming fixed diversification parameters revealed lack of statistical power and are reported on Appendix SG.

Caveats – Conflicting estimates of lineage turnover and SSE model inadequacy

While both BiSSE/GeoSSE and the clade-based estimates consistently indicate that net diversification rates are higher for marine than freshwater clades (see above), these two approaches resulted in apparently contradictory estimates of clade turnover (Fig. 2b; Tables 3, SF4). The clade-based results conform to the expectation that turnover is faster for marine than freshwater clades. Marine clades feature substantially higher crown- vs. stem-based estimates of net diversification relative to freshwater clades (Tables 3, SF4) and stronger disparity in scatterplots of clade richness against crown vs. stem ages (Figs. 3, SF1). The SSE analyses, however, indicate that turnover rates are faster in freshwater lineages (turnover rate ratios marine/freshwater = 0.88–0.96; Fig. 2b, Table SD2). The SSE results are largely robust to clade selection

Table 3 Estimates of net diversification rates for the 55 target clades (35 marine and 20 freshwater) using both crown and stem equations implemented in the method-of-moments (Magallon & Sanderson 2001). Rates were calculated based on the mean values of ε (extinction) estimated with BiSSE/GeoSSE for freshwater (ε = 0.53) and marine (ε = 0.30) habitats and confidence limits were assessed under arbitrarily high (ε = 0.90) and low (ε = 0.0) extinction fractions. Rate estimates based on crown equations are significantly higher than those based on stem equations for marine (MA) clades, whereas all crown vs. stem comparisons for freshwater (FW) clades are non-significant. These results show strong departures from the null, constant birth-death model in marine but not in freshwater clades – a signature of young-to-old lineage out competition in marine habitats

Extinction (ε)	Mean net div. rates FW - Crown	Mean net div. rates FW - Stem	*U* test *P* value	Mean net div. rates MA - Crown	Mean net div. rates MA - Stem	*U* test *P* value
0.0	0.0504	0.0419	0.6359	0.0705	0.0503	0.006**
FW = 0.53; MA = 0.30	0.0468	0.0360	0.3507	0.0689	0.0470	0.002**
0.9	0.0330	0.0249	0.3793	0.0433	0.0296	0.002**

**$P < 0.01$.

(the Actinopterygii tree or the Clupeocephala subtree; Figs. 2b, SD2), although the Clupeocepahala subtree shows substantial overlap for turnover rates in marine and freshwater lineages with uneven sampling (Fig. SD1). While some authors have questioned the feasibility of estimating extinction rates from molecular phylogenies (Rabosky 2010), simulation work has shown that SSE-based approaches have power to detect extinction given sufficient data (e.g., > 300 terminals; Davis *et al*. 2013).

Another concern recently raised is that SSE methods can suffer from model inadequacy producing large Type I error rates. These problems can be the result of unexamined characters (Rabosky & Goldberg In press) or phylogenetic pseudoreplication (Maddison & FitzJohn 2015). We do not think phylogenetic pseudoreplication is an issue for our analyses given the numerous transitions between states (i.e. patterns found are not an artifact of small sample sizes). Differences in speciation and turnover rates detected, however, can potentially be the result of characters not included in the model and could explain the conflicting estimates obtained with clade-based and SSE approaches. Even if SSE model inadequacy is compromising some of the other results reported here, the Type I error would not significantly impact the main conclusions of the study about extinction erasing the marine ancestry.

DISCUSSION

Our results highlight the importance of integrating paleontological and neontological phylogenetic approaches for large-scale comparative analyses (Albert *et al*. 2009; Slater *et al*. 2013). We show that ASRs differ substantially among ancient ray-finned fish lineages when fossils are excluded or included in the phylogeny (Fig. 1). In particular, extinction has erased the signal of marine ancestry from early-branching lineages, misleading ancestral-state estimates using the extant-only tree. This is in agreement with the observation that many of these lineages had marine representatives in the past (e.g., bowfins, bonytongues, and other extinct lineages) in spite of their contemporary freshwater confinement (Parenti 2008).

The disparity in ASRs between the trees examined here is probably the result of two extreme situations. On the one hand, ancient lineages are most likely to persist in fresh than marine waters (see below), which explains the prevalence of ancient nodes reconstructed as freshwater in the tree including modern taxa only. On the other hand, the probability of fossilisation is higher for marine than for freshwater organisms (e.g., Shipman 1993), suggesting that our fossil-based tree may have an overrepresentation of marine extinct tips that is biasing the ASRs towards high probability of marine ancestry.

The true history of habitat evolution in extant and extinct ray-finned fishes most likely lies in between these two extremes, but a signal of marine ancestry is clearly obscured by the extinction of early-branching marine lineages and their absence from the time-tree of extant taxa. Not accounting for state-dependent diversification dynamics is problematic for reconstructing ancestral states on molecular time-trees of extant taxa (Maddison 2006). However, this problem is alleviated as more extinct branches are added, particularly towards the root of the tree, adding direct empirical observations of past character states. Thus, the early signal of marine ancestry in our fossil-based tree should be more robust than the ancestral inference from extant taxa alone in the molecular time-tree, even when accounting for state dependence.

Our analyses also indicate that younger clades may be more likely to out-compete older clades in marine than in freshwaters, explaining the prevalence of ancient lineages in freshwaters but not in marine habitats today. While there appears to be unequivocal evidence that net diversification rates are faster for marine than for freshwater ray-finned fish lineages, state-diversification and clade-based approaches offer conflicting insights into clade turnover. The results obtained using method-of-moments statistics indicate high marine turnover, in agreement with the prediction that freshwater environments are refugia for ancient lineages that have been out-competed in marine environments. The SSE analyses, however, suggest that turnover rates are faster in freshwater habitats. One possible explanation for this result is that higher levels of endemism in riverine habitats may increase stochastic extinction probability. For instance, introduction of alien species has resulted in extinctions of native fauna in rivers but not in marine systems (Lockwood 2004).

There are other examples of ancient fish lineages persisting in freshwater refuges outside Actinopterygii. For example, modern lungfishes (dipnoans) are confined to rivers but multiple marine lungfish fossil forms exist (e.g. Clement & Long 2010). This is not to say that all freshwater habitats are less competitive than their marine counterparts, but that living remnants have greater probability of persistence in rivers that are not ecologically saturated. The extent of ecological saturation of riverine habitats also affects diversification dynamics in modern taxa. For instance, we have shown that both clades and community assemblages of marine-derived freshwater arids are species-rich in rivers with low levels of pre-existing competition (e.g. in the Australia-New Guinea region), whereas the opposite is true for freshwater environments with high incumbent diversity (e.g. the Amazon and Mekong rivers; Betancur-R *et al*. 2012).

Marine ecosystems also exhibit great disparity in species richness patterns of fish assemblages, with tropical shallow reefs being considerably more diverse than deep-sea or temperate environments. In fact, one hypothesis for the origin of deep-sea ichthyofauna is that the deep ocean is a refuge for living fossils of ancient lineages that were outcompeted by contemporary forms in shallow habitats (Wilson & Hessler 1987). Factors affecting diversification dynamics in fishes in relation to depth or latitude remain to be addressed using explicit phylogenetic approaches.

Finally, our results support the notion that marine-to-freshwater transitions in fishes are more frequent than the reverse. Transitions from marine to freshwater also occurred repeatedly outside fishes: at least 16 (out of ca. 30) marine-derived phyla have colonised freshwaters, whereas only seven have made the transition to land (Lee & Bell 1999). Even though all analyses consistently indicate that transition rates are

asymmetric, there is great disparity in rate estimates obtained from alternative analyses (Table SE1). In the light of the abundant instances of marine-to-freshwater transitions in contemporary fish clades (e.g. herrings and allies, sea catfishes, toadfishes, gobies, flatfishes, cichlids, atherinomorphs, mullets, drums and pufferfishes), we feel that some of these results may be underestimating the extent of transition asymmetry in ray-finned fishes, possibly due to taxonomic sampling effects.

Even at shallow evolutionary scales we find support for asymmetric transition rates. For instance, ariid catfishes had an initial shift from fresh to marine waters ca. 70−40 Ma, followed by at least 13 independent events of freshwater recolonization in more recent times (< 20 Ma; Betancur-R. 2010; Betancur-R *et al*. 2012). Apart from ariids, plotosid catfishes constitute the only other fish family of primary freshwater origin that has successfully diversified in the sea. What key innovations promoted fresh-to-marine water transitions and subsequent diversification in these two groups but not in others? We hypothesise that the catfish body plan, with rigid dorsal and pectoral spines, may have provided ariids and plotosids defence in an otherwise overly competitive environment, nowadays dominated by spiny-finned fish (acanthomorph) forms. Physiological constraints could also explain this asymmetry in transition rates (i.e. movements from hypertonic to hypotonic environments may be less stressful than the reverse), but this possibility needs to be explored.

In summary, we show that modern representatives of Paleozoic lineages of ray-finned fishes persist in rivers, but have been driven to extinction in marine environments. This extinction apparently misleads comparative analyses conducted on extant-only trees. Marine lineages are also more likely to colonise and diversify in freshwaters with open ecological niches than are freshwater lineages to flourish in marine habitats. Our study adds to the growing body of evidence that excluding fossils from phylogenetic comparative analyses can mislead inferences in macroevolutionary research. Further integration of paleontological data into phylogenetic comparative approaches will likely be crucial to accurately infer historical dynamics of habitat and diversification.

ACKNOWLEDGEMENTS

We are grateful to A. Acero for a decade-old discussion that led to the conception of this study. We thank L. Revell, D. Bapst, R. Fitzjohn and E. Goldberg for their assistance in R code. G. Slater, three anonymous reviewers, L. Revell, L. Hughes and the GWU Systematics Discussion Group made helpful comments. W.N. Eschmeyer and J. Fong provided access to their Catalog of Fishes database. This research was funded in part by the Peter Buck fellowship to RBR and US NSF grants DBI-0905765 and DEB-1441719 to RAP and DEB-1019308 to GO.

AUTHOR CONTRIBUTIONS

RBR, GO and RAP conceived the study. RBR collected the data, performed the analyses and wrote the manuscript. All authors contributed to the writing and approved the final version of the paper.

REFERENCES

Albert, J.S., Johnson, D.M. & Knouft, J.H. (2009). Fossils provide better estimates of ancestral body size than do extant taxa in fishes. *Acta Zool*., 90, 357–384.

Bapst, D.W. (2013). A stochastic rate-calibrated method for time-scaling phylogenies of fossil taxa. *Methods Ecol. Evol*., 4, 724–733.

Betancur-R., R. (2010). Molecular phylogenetics supports multiple evolutionary transitions from marine to freshwater habitats in ariid catfishes. *Mol. Phylogenet. Evol*., 55, 249–258.

Betancur-R, R., Ortí, G., Stein, A.M., Marceniuk, A.P. & Alexander Pyron, R. (2012). Apparent signal of competition limiting diversification after ecological transitions from marine to freshwater habitats. *Ecol. Lett*., 15, 822–830.

Betancur-R., R., Broughton, R.E., Wiley, E.O., Carpenter, K., Lopez, J.A., Li, C. *et al*. (2013). The tree of life and a new classification of bony fishes. *PLoS Currents Tree of Life*, 2013, 1, April 18, 2013.

Bloom, D.D. & Lovejoy, N.R. (2012). Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae). *J. Evol. Biol*., 25, 701–715.

Bloom, D.D. & Lovejoy, N.R. (2014). The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). *Proc. Biol. Sci*., 281, 20132081.

Bloom, D.D., Weir, J.T., Piller, K.R. & Lovejoy, N.R. (2013). Do Freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (Atherinopsidae). *Evolution*, 28, 2689–2690. n/a-n/a.

Bollback, J.P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, 7, 88.

Clement, A.M. & Long, J.A. (2010). Air-breathing adaptation in a marine Devonian lungfish. *Biol. Lett*., 6, 509–512.

Davis, A.M., Unmack, P.J., Pusey, B.J., Johnson, J.B. & Pearson, R.G. (2012). Marine-freshwater transitions are associated with the evolution of dietary diversification in terapontid grunters (Teleostei: Terapontidae). *J. Evol. Biol*., 25, 1163–1179.

Davis, M.P., Midford, P.E. & Maddison, W. (2013). Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol*., 13, 38.

Eschmeyer, W.N. & Fong, J.D. (2013). Species of Fishes by family/subfamily. Available at: http://research.calacademy.org/ichthyology/catalog/family Last accessed 15 Jan 2013.

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

FitzJohn, R.G., Maddison, W.P. & Otto, S.P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol*., 58, 595–611.

Goldberg, E.E. & Igic, B. (2008). On phylogenetic tests of irreversible evolution. *Evolution*, 62, 2727–2741.

Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011). Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol*., 60, 451–465.

Grande, L. & Bemis, W.E. (1998). A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy. An empirical search for interconnected patterns of natural history. *J. Vertebr. Paleontol*. (Memoir 4, supplement) 18, 690.

Harmon, L., Weir, J., Brock, C., Glor, R. & Challenger, W. (2008). GEIGER: investigating evolutionary radiations. *Bioinformatics* (Oxford, England), 24, 129–131.

Hilton, E.J. (2003). Comparative osteology and phylogenetic systematics of fossil and living bony-tongue fishes (Actinopterygii, Teleostei, Osteoglossomorpha). *Zool. J. Linn. Soc*., 137, 1–100.

Lee, C.E. & Bell, M.A. (1999). Causes and consequences of recent freshwater invasions by saltwater animals. *Trends Ecol. Evol*., 14, 284–288.

Lockwood, J.L. (2004). How do biological invasions alter diversity patterns? A biogeographic perspective. In *Frontiers of biogeography*. (eds Lomolino, M.V., Heaney, L.R.). Sinauer Associates Sunderland, Massachusetts, pp. 271–310.

Logares, R., Brate, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K. & Rengefors, K. (2009). Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol*., 17, 414–422.

Maddison, W.P. (2006). Confounding asymmetries in evolutionary diversification and character change. *Evolution*, 60, 1743–1746.

Maddison, W.P. & FitzJohn, R.G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol*., 64, 127–136.

Maddison, W., Midford, P. & Otto, S. (2007). Estimating a binary character's effect on speciation and extinction. *Syst. Biol*., 56, 701–710.

Magallon, S. & Sanderson, M.J. (2001). Absolute diversification rates in angiosperm clades. *Evolution*, 55, 1762–1780.

McDowall, R.M. (1997). The evolution of diadromy in fishes (revisited) and its place in phylogenetic analysis. *Rev. Fish Biol. Fisheries*, 7, 443–462.

McPeek, M.A. & Brown, J.M. (2007). Clade age and not diversification rate explains species richness among animal taxa. *Am. Nat*., 169, E97–E106.

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L. *et al*. (2013). Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. *Proc. Natl Acad. Sci*., 110, 12738–12743.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* (Oxford, England), 20, 289–290.

Parenti, L.R. (2008). Life History Patterns and Biogeography: an Interpretation of Diadromy in Fishes1. *Ann. Mo. Bot. Gard*., 95, 232–247.

Pyron, R.A. & Burbrink, F.T. (2012). Extinction, ecological opportunity, and the origins of global snake diversity. *Evolution*, 66, 163–178.

Pyron, R.A. & Burbrink, F.T. (2014). Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecol. Lett*., 17, 13–21.

Pyron, R.A. & Wiens, J.J. (2013). Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc. Royal Society B: Biol. Sci*., 280, 20131622.

Rabosky, D.L. (2010). Extinction Rates Should Not Be Estimated from Molecular Phylogenies. *Evolution*, 64, 1816–1824.

Rabosky, D.L. & Goldberg, E. (In press). Model Inadequacy and Mistaken Inferences of Trait-Dependent Speciation. *Syst. Biol*., in press.

Rabosky, D.L., Slater, G.J. & Alfaro, M.E. (2012). Clade age and species richness are decoupled across the eukaryotic tree of life. *PLoS Biol*., 10, e1001381.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol*., 3, 217–223.

Santini, F., Nguyen, M.T., Sorenson, L., Waltzek, T.B., Lynch Alfaro, J.W., Eastman, J.M. *et al*. (2013). Do habitat shifts drive diversification in teleost fishes? An example from the pufferfishes (Tetraodontidae). *J. Evol. Biol*., 26, 1003–1018.

Shipman, P. (1993). *Life History of a Fossil: An Introduction to Taphonomy and Paleoecology*. Harvard University Press, Cambridge, Massachusetts.

Slater, G.J., Harmon, L.J. & Alfaro, M. (2012). Integrating fossils with molecular phylogenies improves inference of trait evolution, 66, 1–14.

Slater, G.J., Harmon, L.J. & Freckleton, R. (2013). Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods Ecol. Evol*., 4, 699–702.

Smith, S.A. & O'Meara, B.C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* (Oxford, England), 28, 2689–2690.

Stadler, T. (2011). Mammalian phylogeny reveals recent diversification rate shifts. *Proc. Natl Acad. Sci. USA*, 108, 6187–6192.

Stadler, T., Rabosky, D.L., Ricklefs, R.E. & Bokma, F. (2014). On age and species richness of higher taxa. *Am. Nat*., 184, 447–455.

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

Swenson, M.S., Suri, R., Linder, C.R. & Warnow, T. (2012). SuperFine: fast and accurate supertree estimation. *Syst. Biol*., 61, 214–227.

Vega, G.C. & Wiens, J.J. (2012). Why are there so few fish in the sea? *Proc. R. Soc. B*, 282, 1–7.

Wilson, G.D.F. & Hessler, R.R. (1987). Speciation in the Deep Sea. *Annu. Rev. Ecol. Syst*., 18, 185–207.

SUPPORTING INFORMATION

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Helene Morlon
Manuscript received 11 October 2014
First decision made 20 November 2014
Second decision made 14 January 2015
Manuscript accepted 29 January 2015