JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2012.02464.x

Molecular phylogenetics reveals a pattern of biome conservatism in New World anchovies (family Engraulidae)

D. D. BLOOM*† & N. R. LOVEJOY*†

*Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada
†Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

Keywords:

anchovies;
biogeography;
freshwater;
habitat transitions;
macroevolution;
marine;
Neotropics;
phylogeny;
South America.

Abstract

Evolutionary transitions between marine and freshwater biomes are relatively rare events, yielding a widespread pattern of biome conservatism among aquatic organisms. We investigated biome transitions in anchovies (Engraulidae), a globally distributed clade of economically important fishes. Most anchovy species are near-shore marine fishes, but several exclusively freshwater species are known from tropical rivers of South America and were previously thought to be the product of six or more independent freshwater invasions. We generated a comprehensive molecular phylogeny for Engraulidae, including representatives from 15 of 17 currently recognized genera. Our data support previous hypotheses of higher-level relationships within Engraulidae, but show that most New World genera are not monophyletic and in need of revision. Ancestral character reconstruction reveals that New World freshwater anchovies are the product of a single marine to freshwater transition, supporting a pattern of biome conservatism. We argue that competition is the principal mechanism that regulates aquatic biome transitions on a continental scale.

Introduction

Understanding the frequency of major evolutionary transitions and how these events alter the trajectory of a clade is a primary interest of evolutionary biologists (Maynard-Smith & Szathmary, 1997; Anderson & Sues, 2007; Hendry *et al.*, 2010). Evolutionary transitions between biomes, such as marine and freshwater environments, are relatively rare events (Gray, 1988; Lee & Bell, 1999; Vermeij & Dudley, 2000; Vermeij & Wesselingh, 2002; Crisp *et al.*, 2009) that can have a profound impact on the history of a clade (Sumida & Martin, 1997; Gingerich *et al.*, 2001; Daeschler *et al.*, 2006; Shubin *et al.*, 2006; Niedzwiedzki *et al.*, 2010). These historical biotic interchanges can expose lineages to novel ecological opportunities, alter rates of evolution and prompt adaptive diversification (Schluter, 2000; Yoder *et al.*, 2010), as well as reshuffle community composition, alter species interactions and alter regional species diversity (Vermeij, 2005). Many clades in the tree of life have undergone macroevolutionary transitions between biomes, and understanding these events is integral to interpreting general patterns of biodiversity (Vermeij, 2006).

The integration of phylogenetics and ecology has led to the concept of phylogenetic niche conservatism (PNC) (Wiens & Donoghue, 2004; Wiens & Graham, 2005). In the broadest definition, niche conservatism predicts that closely related species will be ecologically similar, that is, they will retain (and share) a niche inherited from a common ancestor due to intrinsic (fundamental niche) or extrinsic (realized niche) constraints. From a biogeographic perspective, this translates to a tendency for lineages to track their ancestral habitat rather than exhibit transitions between different habitats (Harvey & Pagel, 1991). Alternatively, some clades show evidence for repeated transitions between habitats, demonstrating a niche lability or niche evolution model. Under the niche lability model, a trait or niche axis may evolve repeatedly within a clade, provided there is limited competition and recurring biogeographic opportunity

Correspondence: Devin D. Bloom, Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.
Tel.: +1 416 208 4823; fax: +1 416 287 7676;
e-mail: devin.bloom@gmail.com

(Wiens *et al.*, 2006). Thus, the niche evolution model posits that closely related species are not necessarily ecologically similar (Losos *et al.*, 2003).

To date, most studies on phylogenetic niche conservatism and niche lability have focused on small-scale microhabitat shifts, such as switches between forest and savannah in flycatchers (Rheindt *et al.*, 2008), flowing or stagnant water in aquatic dysticid beetles (Ribera & Vogler, 2000; Ribera *et al.*, 2001), and tidal or intertidal habitats in sculpin fishes (Ramon & Knope, 2008). The prevalence of continental-scale habitat (or biome) transitions, such as those between major aquatic biomes (Crisp *et al.*, 2009), remains largely unexamined despite considerable interest (Pearse, 1927; Romer & Grove, 1935; Robertson, 1957; Parry, 1966; Halstead, 1985; Griffith, 1987; Crisp, 2006). Marine and freshwater biomes are profoundly different aquatic environments separated by stringent physiological barriers (Bloom & Lovejoy, 2011). Correspondingly, biotic interchanges between these biomes are thought to occur infrequently over geological time (Gray, 1988; Winemiller & Leslie, 1992; Lee & Bell, 1999; Mank & Avise, 2006; Logares *et al.*, 2010), and many clades have distributions in either freshwater or marine habitats, but not both. Nonetheless, transitions between marine and freshwater biomes have occurred across the tree of life, in microbes (Stahl *et al.*, 1992; Logares *et al.*, 2007, 2009, 2010), amoebae (Heger *et al.*, 2010), crabs (Daniels *et al.*, 2006), shrimp (Daniels *et al.*, 2006; Augusto *et al.*, 2009), mammals (Cassens *et al.*, 2000; Hamilton *et al.*, 2001) and fishes (Lovejoy *et al.*, 1998, 2006; Lovejoy & Collette, 2001; Yokoyama & Goto, 2005; Kawahara *et al.*, 2009; Whitehead, 2010). The fish clades for which phylogenetic data are available have shown patterns of both biome conservatism (Lovejoy *et al.*, 1998; Lovejoy & Collette, 2001; Yokoyama & Goto, 2005; Whitehead, 2010) and lability (Lovejoy & Collette, 2001; Betancur-R, 2010; Whitehead, 2010). Here, we use anchovies as a model system to investigate the evolution of transitions between marine and freshwater biomes at a continental scale. We present a new phylogenetic hypothesis for anchovies and clarify the evolutionary origins of the remarkable freshwater anchovies inhabiting the Amazon and other major rivers of South America.

The anchovy family Engraulidae is a well-defined monophyletic group (Grande & Nelson, 1985; Lavoue *et al.*, 2007, 2009) with ~140 species divided into 16 genera found in temperate and tropical regions around the world. Most anchovies are highly abundant, marine, planktivorous fishes that form large schools in near-shore habitats. However, there are some extraordinary ecological exceptions. In South America, there are 12+ anchovy species that occur in major tropical rivers, including the Amazon, Orinoco and Essequibo. Most of these species occur exclusively in freshwater, in some cases living thousands of kilometres from marine habitats. These peculiar freshwater anchovies exhibit great diversity in body size and ecology. For example, *Lycengraulis batesii* is a large-bodied piscivorous species that reaches 300 mm standard length (SL) and has canine teeth and enlarged gill raker denticles (Bornbusch, 1988; Whitehead *et al.*, 1988). At the opposite end of the spectrum is the miniaturized paedomorphic species *Amazonsprattus scintilla*, which has a maximum size of < 20 mm SL, making it the smallest known clupeomorph (Roberts, 1984; Weitzman & Vari, 1988). The currently recognized taxonomic arrangement of New World anchovies suggests that freshwater South American species are the result of multiple independent transitions from a marine environment (Nelson, 1983, 1984, 1986; Grande & Nelson, 1985) (Fig. 1). For example, Nelson (1984) suggested that the Amazonian species *Jurengraulis juruensis* is nested within marine *Cetengraulis* and *Engraulis* species (Fig. 1) and must therefore have invaded freshwater independently of other freshwater anchovy lineages. This pattern of freshwater species nested within a predominately marine group is repeated multiple times across New World anchovies. Four of the eight New World genera include both marine and freshwater species, and there are two monotypic freshwater genera (*Pterengraulis* and *Amazonsprattus*), indicating that six or more marine to freshwater transitions may have occurred in South America. This is a striking pattern given the physiological challenges of moving to a new biome (Lee & Bell, 1999; Wiens & Donoghue, 2004; Wiens & Graham, 2005; Crisp *et al.*, 2009) and suggests that New World anchovies fit a biome niche evolution model. However, a comprehensive phylogeny for anchovies has not yet been proposed, precluding analysis of the frequency of marine to freshwater transitions in this clade.

Based on morphological criteria, Grande & Nelson (1985) divided the 16 genera of the anchovies (family Engraulidae) into two subfamilies, Coilinae and Engraulinae. Coilinae is an entirely Old World group found in the Indo-Pacific and includes the genera *Coilia*, *Lycothryssa*, *Papuengraulis*, *Setipinna* and *Thryssa*. Although often referred to as a New World clade, the Engraulinae includes the Indo-Pacific genera *Stolephorus* and *Encrasicholina*, the cosmopolitan genus *Engraulis*, and the New World genera *Anchoa*, *Anchovia*, *Anchoviella*, *Amazonsprattus*, *Lycengraulis*, *Cetengraulis*, *Jurengraulis* and *Pterengraulis*. These New World genera and the widespread genus *Engraulis* together are thought to form the clade Engraulini (Nelson, 1970; Grande & Nelson, 1985). Lavoue *et al.* (2009) was the only previous molecular phylogenetic study to investigate higher-level relationships within Engraulidae. Using mitogenomics, they supported the relationships proposed by Grande & Nelson (1985) and confirmed that *A. scintilla* is a member of Engraulini (Nelson, 1984), but they lacked the necessary taxon sampling to evaluate relationships between genera and species. Despite being recognized as one of the most important ecological and economical groups of fishes at all times (Whitehead, 1985; Whitehead *et al.*, 1988), very

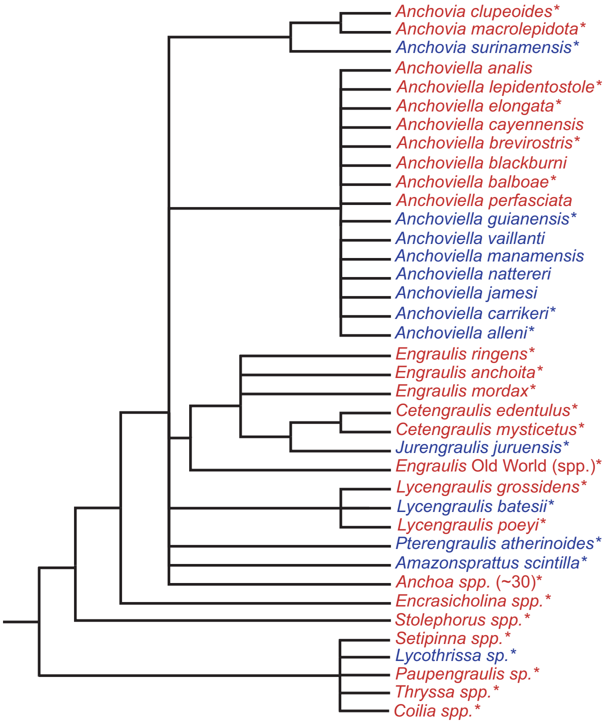

Fig. 1 Summary of previous hypotheses of anchovy relationships from Grande & Nelson, 1985; Nelson, 1984; and Whitehead *et al.*, 1988. Marine species are shown in red and freshwater species in blue.

little is known about the evolutionary relationships of anchovies.

In this study, we propose the first comprehensive molecular phylogeny for the anchovy family Engraulidae, including representatives from 15 of the 17 recognized genera, based on both mitochondrial and nuclear genes. Using this tree, we reconstruct transitions between marine and freshwater habitats in the New World Anchovies (Engraulini) and thereby test whether anchovies fit the niche conservatism or niche evolution model. We also evaluate previous hypotheses of higher-level anchovy relationships and clarify the origins of the remarkable South American freshwater anchovies.

Methods

Taxon sampling

Our data set comprises 60 species (117 individuals) representing all nine New World genera and 15 of the 17 currently recognized genera in the anchovy family (Roberts, 1984; Peng & Zhao, 1988; Whitehead *et al.*, 1988). We focused sampling on New World taxa, in order to provide the most robust test of habitat transitions in South American lineages (Appendix S1).

Currently, there are 12 described freshwater species from six genera found in South American freshwaters; however, a number of additional species await formal taxonomic description and species limits are poorly known (Whitehead, 1973). A number of freshwater specimens included in our study could not be unequivocally assigned to a particular described species, and our molecular data indicated they might represent undescribed taxa. The taxonomic status of these individuals was beyond the scope of this study; however, we included any freshwater individual that could potentially represent a distinct species. Our data set included eight currently recognized and three putative species from continental freshwaters of South America, with representatives of all six genera found in Neotropical freshwaters. We also comprehensively sampled marine species, including 30 New World species. We specifically targeted lineages that were previously proposed as sister to freshwater taxa from the Pacific, Atlantic and Caribbean Oceans. For outgroups, we included 10 species representing the major lineages of Clupeiformes. Trees were rooted with *Denticeps clupeoides*, a basal clupeoid (Lavoue *et al.*, 2007, 2009; Li & Orti, 2007). When possible, multiple individuals of each species were sequenced for all genes. Specimens were collected using seine nets, dip-nets and cast-nets, or purchased from fish markets. Muscle or fin tissue was stored in either 95% ethanol or a salt solution consisting of 20% DMSO and 0.25 M EDTA saturated with NaCl.

DNA extraction, PCR and sequence acquisition

Whole-genomic DNA was extracted using the DNEasy spin column tissue kit (Qiagen Inc., Valencia, CA, USA). We collected DNA sequence data from fragments of two mitochondrial genes 16s and cytochrome *b* (Cyt*b*). The 16s fragment was PCR-amplified using primers 16S135F and 16S1072R from Li & Orti, (2007). Cytochrome *b* was amplified using newly designed primers CytbAnchF (5′ TGACTTGAAAAACCACCGTTGTTATTCAAC 3′) and CytbAnchR (5′ CTAGCTTTGGGAGYTAGDGGTGGRAGTT 3′). Additionally, we sequenced fragments of the nuclear recombination activating genes 1 and 2 (RAG1 and RAG2). Primers for PCR amplification of RAG1 were RAG12510F from Li & Orti, (2007) and RAG14078R from Lopez *et al.* (2004). The primers RAG2AnchF (5′ TTCAAGCTTCGCCCYATCTCTTTCTCCAA 3′) and RAG2AnchR (5′ CTCCATGCACTGGGCGTGGACCCA 3′) were newly designed for this study. PCR for 16s and cytb were performed in 25-μL reactions, which included 2.5 μL 10× PCR buffer, 2 μL $MgCl_2$, 2 μL dNTPs (10 mM), 2 μL of each primer (10 mM), 0.5 μL Taq polymerase, 1–4 μL genomic DNA and the remaining volume of H_20. PCR thermocycling conditions were 95 °C for 2 min, followed by 30–40 cycles of 95 °C for 30 s, 53 °C for 1 min, 72 °C for 90 s and a final extension of 72 °C for 5 min. PCR for RAG1 and RAG2 were conducted in 50-μL reactions containing 5 μL 10× PCR buffer, 4 μL $MgCl_2$, 2 μL dNTPs (10 mm), 2 μL each primer (10 mm), 1 μL Taq polymerase, 1–5 μL genomic DNA and the remaining volume of H_20. Thermocycling conditions for the RAG genes were 95 °C for 4 min, 35–40 cycles of 95 °C for 1 min, 50–55 °C for 1 min, 72 °C for 90 s, and a final extension of 72 °C for 5 min. The PCR products for all four genes were purified using Qiagen spin column PCR purification kit. Both 16s and Cytb were sequenced using the PCR amplification primers. Internal sequencing primers used for RAG1 were 3222F from Li & Orti (2007) and the newly designed RAG1SEQF (5′ TACCACAAGATGTACCGCAC 3′). Internal sequencing primers for RAG2 were RAG2-526F and RAG2-1096F from Li & Orti (2007), as well as newly designed RAG2SEQR (5′ CAGCTTAGGGCTGCCCAA-CAGAAGCTCGAC 3′). Samples were sequenced at the SickKids Centre for Applied Genomics, Toronto, Canada.

Alignment

Forward and reverse sequences were edited, used to build consensus sequences and then exported for analysis using SEQUENCHER 4.6. (Genecodes). Multiple alignments for each gene were conducted using CLUSTAL X (Thompson *et al.*, 1997). Default settings were used for cytb, RAG1 and RAG2. The resulting alignments were evaluated in MacClade (Maddison & Maddison, 2001) to ensure no stop codons were present. The 16s data were subjected to gap opening and extension parameters (10/10, 10/5, 20/5, 25/5, 30/5, 35/5), the resulting alignments were compared qualitatively, and it was determined that 16s alignment was stable across this range of alignment parameters.

Data analysis

Aligned sequences were used to produce four data sets: (i) the two mitochondrial genes combined, (ii) RAG1, (iii) RAG2 and (iv) all genes (16s, cytb, RAG1, RAG2) concatenated into a single total evidence matrix. The total evidence matrix was partitioned by gene for maximum likelihood (ML) and Bayesian (BI) analysis. Congruence among partitions was assessed using an incongruence length difference (ILD) test implemented in PAUP* (Swofford, 2002). We tested for selection and recombination in our nuclear data set to confirm the appropriateness of these genes for reconstructing the evolutionary history of anchovies from different environments and selective regimes. In order to test for positive selection, codon-based likelihood methods were used to estimate d_N/d_S ratios in the RAG1/RAG2 data set. Random sites (Nielsen & Yang, 1998; Yang *et al.*, 2000) models were implemented using the codeml program of the PAML software package (Yang, 2007). We used PhiPack (Bruen *et al.*, 2006) to test for possible recombination in the RAG genes.

Maximum parsimony (MP) analysis was applied to the four-gene data set using PAUP* (Swofford, 2002). For MP tree searches, we used the heuristic search algorithm with 1000 random addition replicates, and TBR branch swapping. All characters were equally weighted and gaps were treated as missing data. Bootstrap support values were calculated using 1000 replicates with 10 random sequence additions per replicate.

A best-fit model of sequence evolution and parameter estimation for each gene was determined under the Akaike information criterion using ModelTest (Nylander *et al.*, 2004). Partitioned ML tree searches were performed with GTR+G models for each partition using the program RAXML (Stamatakis, 2006). ML bootstrap estimates were based on 100 replicates using the rapid bootstrapping algorithm in RAXML.

Bayesian inferences are known to improve when heterogeneity is accommodating using mixed-model partitioned approaches (Brandley *et al.*, 2005; Brown & Lemmon, 2007); therefore, we partitioned our data by gene, using the best-fit model of evolution chosen by ModelTest. We conducted a BI analysis using MRBAYES v3.1.2 software (Ronquist & Huelsenbeck, 2003). Four independent runs were conducted, and each search consisted of four chains sampling every 100 generations for 20 million generations. All parameters were unlinked and default priors were used. Adequate mixing of Metropolis-coupled chains was checked to ensure acceptance rates fell between 10% and 70%. Convergence was assessed in several ways. First likelihood vs. generation plots were evaluated using the sump command in MrBayes. Second, average standard deviation of split frequencies was checked to ensure they remained below 0.01 and potential scale reduction factors were 1.0. Finally, cumulative posterior probability plots were constructed using the *compare* command in AWTY (Nylander *et al.*, 2008). Based on these measures, we conservatively determined that convergence had been reached within 4 million generations, and these were discarded as burn-in. The remaining 16 000 trees from each run were combined, and the frequency of clade occurrence represented posterior probabilities of clades.

Habitat reconstruction

The evolutionary history of habitat transitions was inferred using ancestral character reconstruction. We classified species as either marine or freshwater using literature sources, museum sources and personal observations. Estuarine species were categorized as marine, because these species are rarely or never found in entirely freshwater habitats, and likely do not reproduce in freshwater habitats.

Habitat type (marine or freshwater) was coded as a discrete, unordered, binary character. MP and ML character reconstruction were implemented using MESQUITE VERSION 2.6 (Maddison & Maddison, 2009). The MP criterion minimizes the number of character state changes needed to explain the current state at the tips of the tree, while allowing a single character state per branch. MP reconstruction is agnostic to branch length information, and any character state may change to any other state. ML determines the likelihood of a character state at each internal node using the Mk model (Pagel, 1999), thus providing a measure of uncertainty for character states, while taking into account branch length information (Schluter *et al.*, 1997). Ideally, reconstructions would be conducted on a chronogram to determine the explicit timing of biome transitions. However, the scarcity of fossil anchovies precluded our ability to generate a time-calibrated phylogeny (see Discussion). Thus, character reconstructions were optimized on the BI tree from the concatenated four-gene data set, utilizing branch length information for ML reconstructions.

Results

Molecular data

The 16s data set resulted in 804 aligned characters including gaps, 398 of which were parsimony-informative. We also removed all gaps and constructed MP trees and found that removing gaps had no effect on the topology of the 16s tree; thus, further discussion will only focus on the alignment that included gaps. The cytochrome b data set yielded 1131 base pairs, including 490 informative sites. The RAG1 and RAG2 data resulted in 1493 and 1219 base pairs, of which 550 and 519 where parsimony-informative. An intron spanning 390 base pairs was detected in *Anchoa cubana*, this intron was previously reported by Li & Orti (2007); however, the specimen was erroneously identified as *Anchoa lyolepis*. Chi-square tests of homogeneity indicated that none of the data sets consisted of biased base pair composition (data not shown). We found no evidence for positive selection in our RAG1/RAG2 data set, as a model incorporating selection was not found to be a better fit to the data than one without selection (M1a-M2a LRT, d.f. = 2, $P = 1.00$) in our analysis using random sites models (Yang, 2007). We also failed to detect evidence for recombination in either RAG1 ($P = 0.489$) or RAG2 ($P = 0.480$). Uncorrected sequence variation ranged from 0.12% to 22.0% for 16s, 0.17–22.9% in cytochrome b and ~0.5–18% in both RAG1 and RAG2 genes. The combined data set consisted of 4647 characters, of which 1957 were parsimony-informative (Appendix S2).

Phylogenetic relationships

The equally weighted MP analysis of the four-gene data set produced three equally parsimonious trees of 13121 steps (Fig. 2). The ML analysis produced a well-resolved tree with a score of –63198.938072 (not shown, but discussed below). The BI analysis was run four times with identical results recovered from each run; the resulting tree is shown in Fig. 3.

All analyses and partitions strongly supported the monophyly of the anchovy family Engraulidae. We also recovered the separation of Engraulidae into two major clades corresponding to subfamilies Coilinae and Engraulinae. The subfamily Coilinae consisted of the Indo-Pacific genera *Setipinna*, *Lycothrissa*, *Coilia* and *Thryssa*, whereas Engraulinae included Indo-Pacific genera *Stolephorus* and *Encrasicholina* as sister to the clade Engraulini including all New World anchovies and the genus *Engraulis*. Genera within Coilinae were all monophyletic, albeit with limited taxon sampling for those groups.

Within Engraulini, only two (*Lycengraulis* and *Cetengraulis*) of the six polytypic genera were monophyletic. The Eastern Pacific *Engraulis anchoita* and Western Atlantic *E. ringens* formed a group that was sister to two well-supported major subclades. The first major subclade (marine clade) included all members of the speciose marine genus *Anchoa*, the Atlantic and Pacific species of *Cetengraulis* and *Anchovia*, the remaining species of *Engraulis*, and two marine species of *Anchoviella*. Our data indicate that the commercially important genus *Engraulis* is not a monophyletic assemblage; in fact, *E. eurystole* (W. Atlantic), *E. encrasicolus* (E. Atlantic) and *E. japonicus* (W. Pacific) were the only members of *Engraulis* to form a clade. The Eastern Pacific *Anchovia marcolepidota* and Western Atlantic *Anchovia clupeoides* were nested within *Anchoa*, whereas the freshwater lineage *Anchovia surinamensis* was a member of the South American freshwater clade (see below). The genus *Cetengraulis* was recovered as closely related to *A. lyolepis* and *Anchoa nasus*.

The second major clade (freshwater clade) consisted of the South American freshwater species of *Anchoviella*, *Lycengraulis*, *Anchovia*, *Juruengraulis* and *Amazonsprattus*, as well as the coastal marine taxa *Lycengraulis poeyi*, *Lycengraulis grossidens*, *Anchoviella brevirostris* and *Anchoviella lepidentostole*. We refer to this clade as the 'freshwater clade' because although it includes several estuarine or marine taxa, these species were derived from freshwater lineages (see habitat transitions below). Within the freshwater clade, *J. juruensis* was the basal lineage in the combined data, but this relationship was not supported by all partitions (see below). The large predatory species of *Lycengraulis* and *Pterengraulis* were sister lineages and part of clade that included the large-bodied planktivorous *A. surinamensis* and the estuarine species *A. lepidentostole*. This group of large-bodied taxa was in turn closely related to a clade of very small-bodied species including the paedomorphic species *A. scintilla*; however, support for this relationship was low (PP = 0.61). Finally, we found strong support for a clade of widely distributed Amazonian freshwater taxa including *Anchoviella carrikeri*, *A. alleni*, *Anchoviella guianensis* and several lineages that appear to represent undescribed species, as well as the estuarine species *A. brevirostris*.

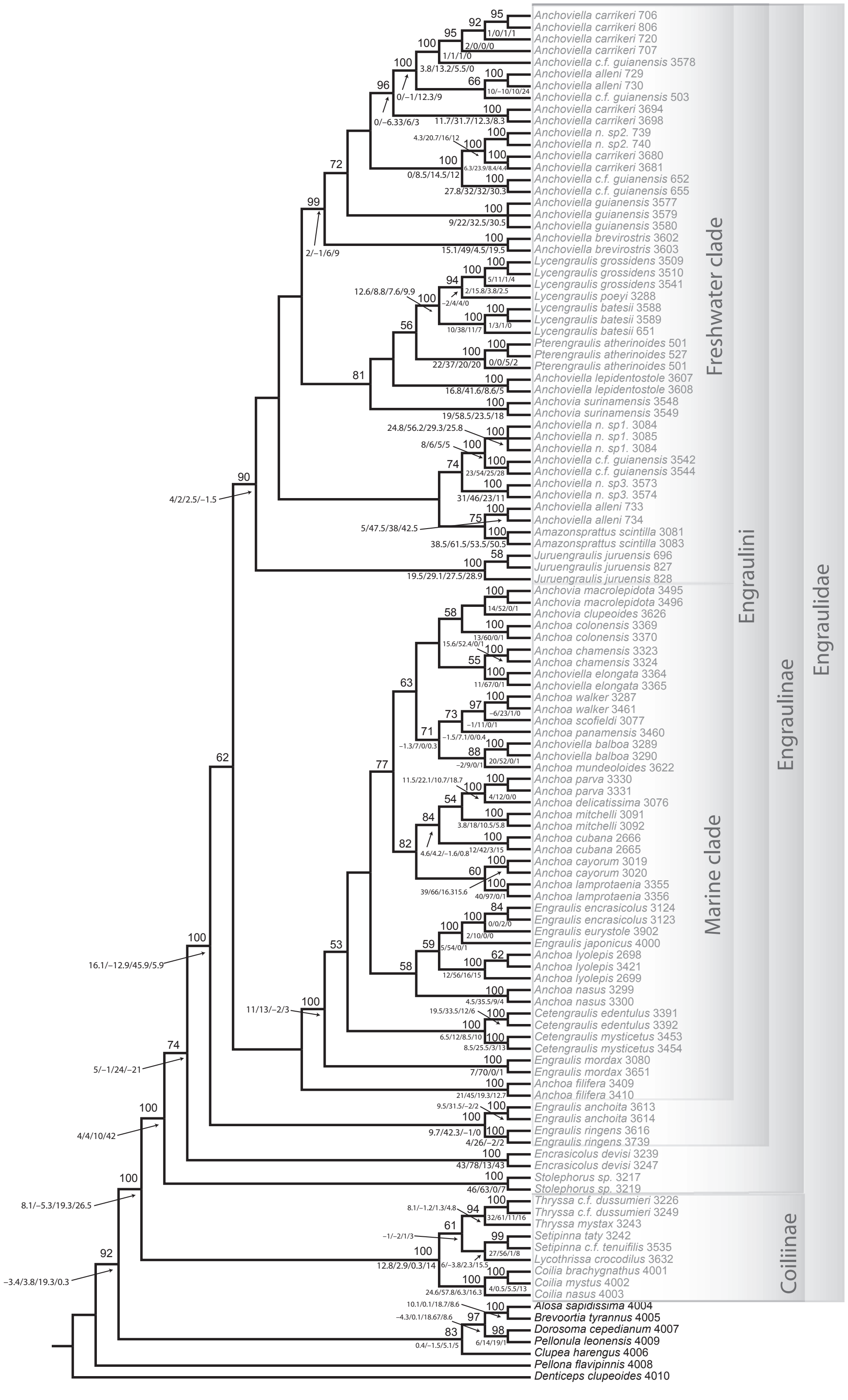

Fig. 2 Strict consensus of three equally parsimonious trees from the combined (cytb, 16s, RAG1, RAG2) data. Numbers above nodes are bootstrap values from 1000 replicates and below indicate decay indices.

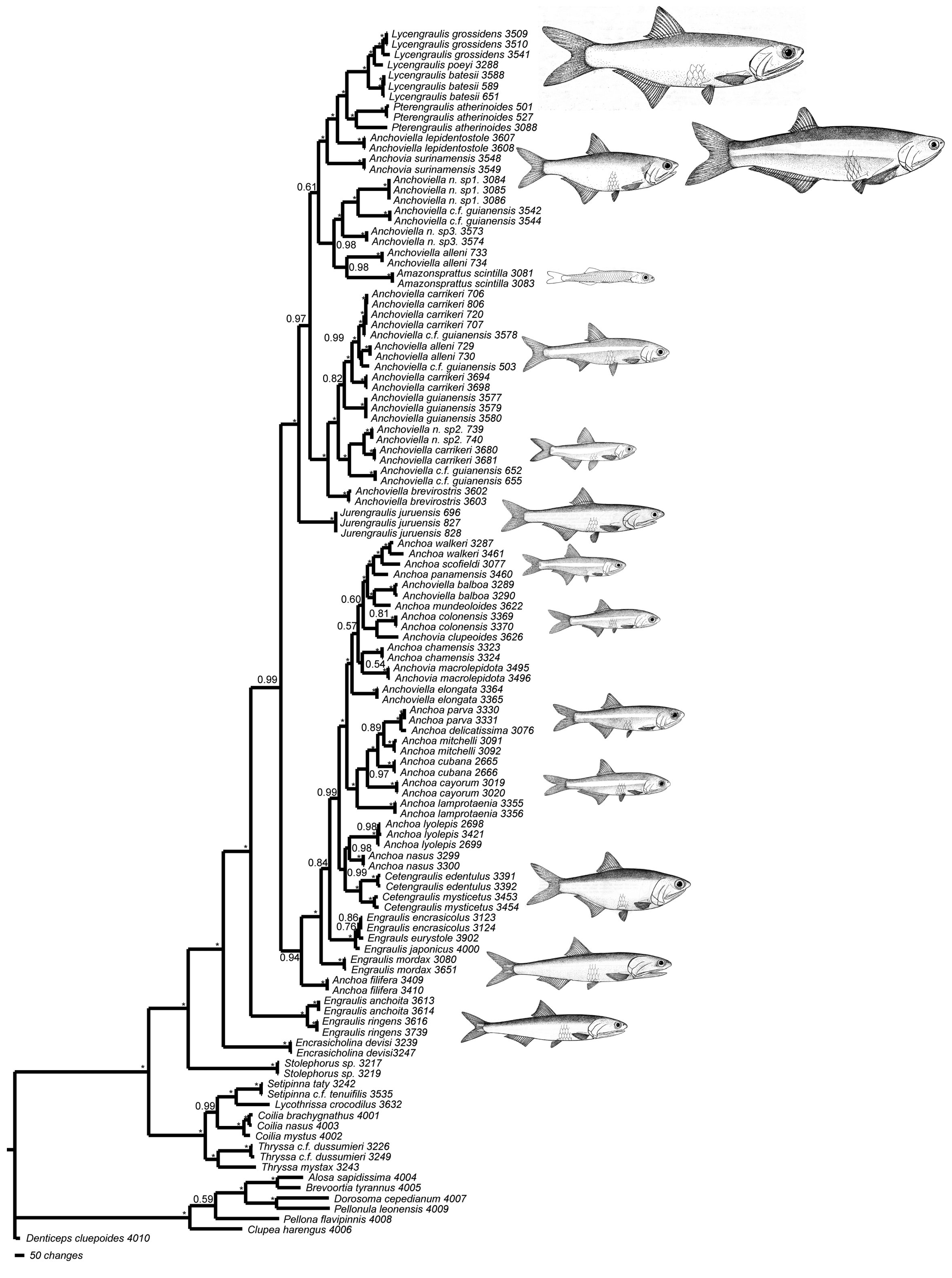

Fig. 3 Bayesian (BI) phylogeny estimated from a partitioned mixed-model analysis of the combined data set. Numbers above nodes represent posterior probabilities (PP). Asterisks above nodes indicate 100% PP values and below indicate nodes with maximum likelihood (ML) bootstrap values >70. The ML topology (not shown) was nearly identical to the BI phylogeny.

Incongruence among analyses and partitions

Maximum likelihood and BI reconstructions of the four-gene data set were nearly identical; although the major findings were consistent between MP and ML+BI, here we report the few notable differences between these methods. In addition, although we consider the four-gene data set the best estimate of anchovy relationships, we explore the relative contribution of each data partition. The major differences between analyses and partitions primarily deal with internal nodes within Engraulini that display very short branches for all genes.

Our BI analysis recovers *Cetengraulis* as sister to a clade consisting of *A. lyolepis* and *A. nasus* (PP = 0.99), whereas MP suggests *Cetengraulis* is sister to a large clade of *Anchoa*, *Engraulis* and some species of *Anchoviella*; however, bootstrap support was low. Further, BI indicates that *Anchovia macrolepidota* is sister to *Anchoa chamensis* and *A. clupeoides* is sister to *Anchoa colonensis* (but PP for both <0.55), whereas MP supports *A. clupeoides* and *A. macrolepidota* as a monophyletic clade (but BS < 50).

Topological structure among major lineages within the South American freshwater clade received poor statistical support and correspondingly suggested different relationships between MP and BI+ML. Most notably, BI suggested a close relationship between a clade of small-bodied species *Anchoviella spp.* and *A. scintilla* with a clade including the piscivorous genera *Lycengraulis* and *Pterengraulis* in addition to *A. surinamensis* and *A. lepidentostole*, whereas MP was unable to resolve these clades.

The nuclear genes, RAG1 and RAG2, showed very similar topologies overall; however, RAG2 had less resolution and lower bootstrap support for most clades. The following were the only two differences between the RAG genes: 1) RAG2 suggested *J. juruensis* was part of the marine clade (but BS < 50) rather than the basal lineage of the freshwater clade and 2) the Indo-Pacific genus *Stolephorus* was recovered as sister to *Engraulis mordax* by RAG2 (BS < 50) and as a basal member of Engraulinae by all other partitions (BS > 85 and PP = 1.0). The mtDNA supports a sister relationship between *Cetengraulis* and *E. mordax* (but BS < 50), whereas both RAG genes indicate *Cetengraulis* form a clade with *A. lyolepis* and *A. nasus* and places *E. mordax* as a basal Engraulini. The mtDNA and RAG1 data sets propose *E.* r*ingens* + *E. anchoita* as basal to the marine clade (mtDNA BS = 72; RAG1 BS < 50), whereas RAG2 and the four-gene data set strongly support the *E. ringens* + *E. anchoita* lineage as the basal Engraulini.

Habitat reconstructions

Ancestral character reconstruction using both MP and ML approaches yielded identical results and is summarized in Fig. 4. Critical nodes for habitat transitions received high statistical support from posterior probabilities and MP and ML bootstrap support. Our analyses all support the long-standing hypothesis that anchovies are an ancestrally marine clade, including members of the New World Clade Engraulini. Further, our data clearly showed that freshwater anchovies in South America are the result of a single transition from a marine to freshwater biome. The sister lineage to freshwater anchovies was a clade consisting of all remaining Engraulini, except *E. anchoita* and *E. ringens*, which were the basal Engraulini lineage. Subsequently, freshwater lineages made three independent invasions back into marine habitats. These marine/estuarine invaders include (i) the clade of Pacific *L. poeyi* and Caribbean + W. Atlantic *L. grossidens*, (ii) *A. brevirostris* and (iii) *A. lepidentostole*. These results were robust to a range of transition rates; even with a 100 times higher transition rate than the optimal rate estimated by Mesquite, the ancestral states remain in the estimated state. Further, a two-rate model was not significantly better than a 1-rate model (LRT, $P = 0.1835$).

Discussion

Phylogeny of Engraulidae

The higher-level relationships of anchovies recovered in our study are consistent with previous investigations based on both molecular (Lavoue *et al.*, 2007, 2009; Li & Orti, 2007; Wilson *et al.*, 2008) and morphological data (Grande & Nelson, 1985; Di Dario, 2002, 2009). The anchovy family Engraulidae is monophyletic and divided into two major clades that correspond to the subfamilies Coilinae and Engraulinae. The subfamily Coilinae is an entirely Indo-Pacific clade including the genera *Coilia*, *Lycothrissa*, *Setipinna*, *Thryssa* and presumably *Paupengraulis*, which has yet to be included in any phylogenetic study. Engraulinae includes the Indo-Pacific genera *Stolephorus* and *Encrasicholina*, along with widespread *Engraulis* and seven New World genera. The relationship of [*Stolephorus* (*Encrasicholina* (New World taxa))] was also supported by a recent mitogenomic study by Lavoue *et al.* (2009). Within Engraulinae, the New World taxa and *Engraulis* form a clade referred to as Engraulini following Lavoue *et al.* (2009). Several morphological characters support the monophyly of Engraulini, most notably the loss of ventral scutes (Nelson, 1970, 1983; Grande, 1985; Grande & Nelson, 1985), a character present in nearly all other clupeomorph fishes.

No previous phylogenetic study has included sufficient taxon sampling to determine relationships within Engraulini; thus, our study offers the first insight into New World anchovy relationships. All freshwater taxa are the result of a single marine to freshwater transition. A well-supported deep divergence between predominantly marine and freshwater clades indicates this biome transition took place early in the diversification of New World anchovies and that freshwater lineages are nearly as old as New World marine lineages. This phylogenetic

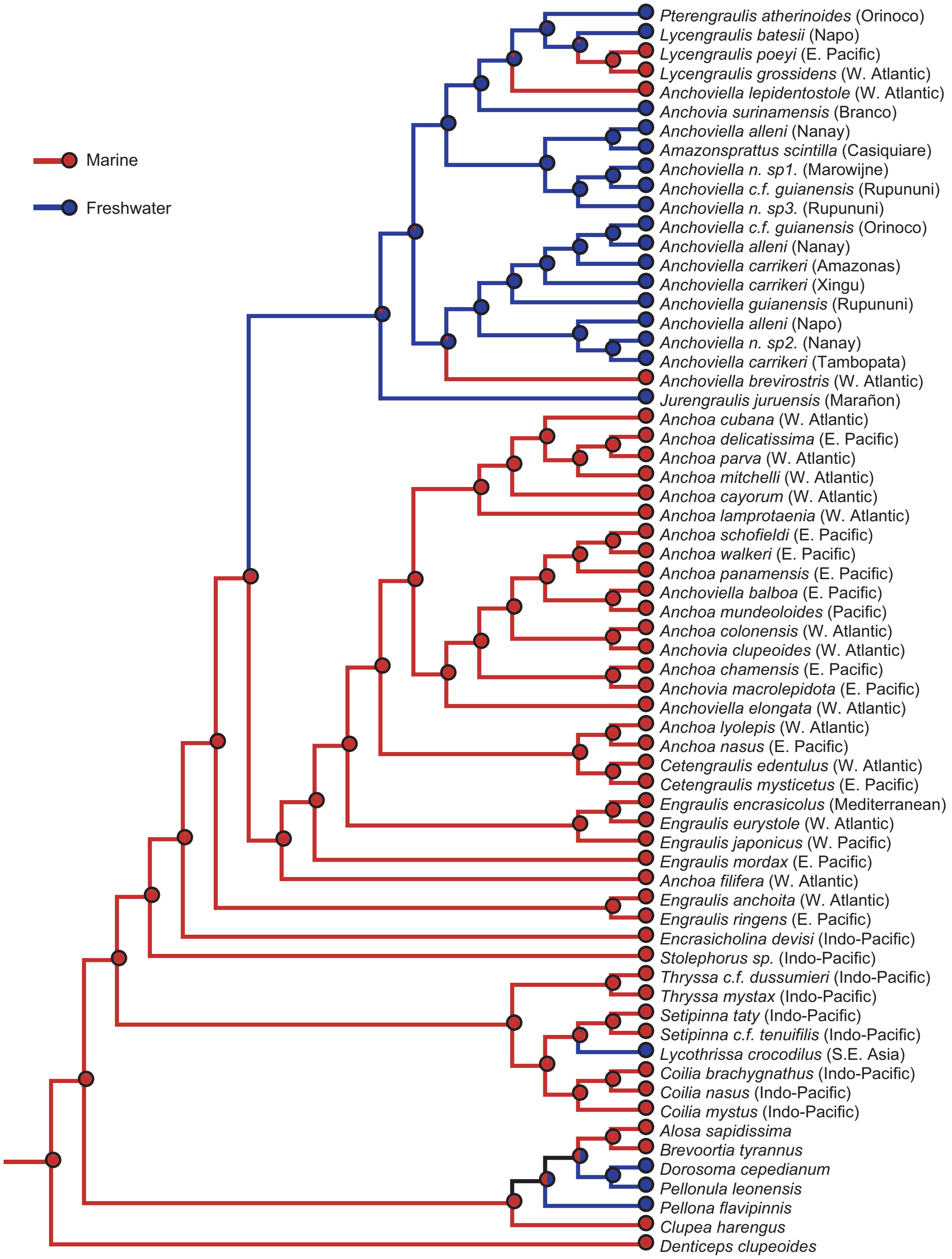

Fig. 4 Ancestral character reconstructions of marine (red) and freshwater (blue) biomes on the Bayesian anchovy phylogeny. Pie charts at nodes show maximum likelihood support for ancestral states, and branch colour indicates maximum parsimony reconstructions.

arrangement differs significantly from the currently accepted taxonomy for New World anchovies where several different genera include both marine and freshwater species, suggesting multiple transitions to freshwater. In the phylogeny presented here, the marine members of these genera are either a result of reversal to marine habitats (*Lycengraulis*, several *Anchoviella spp.*) or members of the large 'marine clade' (*Anchovia*, *Cetengraulis*, remaining *Anchoviella*). In the former case, the relationships previously proposed by taxonomy are consistent with our study, but the inferred phylogenetic pattern requires a different biogeographic interpretation (reversal to marine state versus multiple freshwater invasions; see below). The latter case, marine taxa thought to have close affinities to freshwater lineages (e.g. *A. clupeoides* and *A. macrolepidota*), is likely the result of classifying lineages based on homoplasious morphologically characters. For example, a deep body and high number of fine gill rakers are shared by members currently recognized as *Anchovia* (Whitehead, 1973), and a posteriorly rounded short maxilla is a defining character of all species recognized as in *Anchoviella* (Hildebrand, 1943), but the non-monophyly of these genera suggests that these functional characters likely reflect trophic niche rather than phylogenetic relatedness.

Within the freshwater anchovies, we identified three major multi-species clades, in addition to the basal lineage *J. juruensis*. The first clade was composed entirely of large-bodied anchovies (most larger than 250 mm SL), including the piscivorous genera *Lycengraulis* and *Pterengraulis atherinoides*, as well as the coastal marine–estuarine *A. lepidentostole* and the freshwater species *A surinamensis.* As this clade includes marine lineages (*L. grossidens* + *L. poeyi*, and *A. lepidentostole*) that are well nested within freshwater species, the tree suggests these marine anchovy lineages are the product of two independent re-invasions of coastal habitats. The second clade is a diverse array of *Anchoviella* lineages from the upper Amazon, Orinoco and Essequibo rivers, several of which are likely undescribed species. The inclusion of the coastal marine/estuarine *A. brevirostris* in this clade represents a third re-invasion of the marine environment along the northern South American coast. The third major clade included the paedomorphic *A. scintilla* and several diminutive species of *Anchoviella* (likely including undescribed species), a relationship previously proposed by Nelson (1986). The placement of *A. scintilla* as a member of this clade refines Lavoue *et al.*'s (2009) recent mitogenomic work, confirming this taxon as a member of Engraulinae. The structuring of body size among clades suggests an early and substantial diversification of large- and small-bodied lineages of freshwater anchovies, possibly in response to the ecological opportunity of invading a novel habitat (Yoder *et al.*, 2010). More recently, at least three lineages have re-invaded coastal marine/esturarine habitats along north-eastern South America. Further investigations may reveal why this region has been a fertile ground for re-invasion of marine habitat.

The marine clade includes all species of *Anchoa* and members of *Anchovia*, *Anchoviella* and *Engraulis* from both the Eastern Pacific and Western Atlantic (mostly Caribbean) oceans. The genus *Anchoviella* is clearly in need of revision; *Anchoviella elongata* and *Anchoviella balboa* are nested within a large clade of *Anchoa* and did not group with other marine and freshwater species of *Anchoviella*. Further, we reject the close relationship between *Anchovia* and *Cetengraulis* suggested by similarities in overall appearance and gill raker count and structure (Nelson, 1984). *Cetengraulis* is closely related to tropical Caribbean *A. lyolepis* and eastern tropical Pacific *A. nasu* (but see results for incongruence between MP and BI+ML), contrary to proposed affinities with *Engraulis* and *Jurengraulis* (Nelson, 1970, 1983, 1984, 1986). Our data suggest *E. eurystole* (Western Atlantic), *E. encrasicolus* (Mediterranean and Eastern Atlantic) and *E. japonicus* (Western Pacific) are included in the marine clade, and although our data set did not include *E. australis* and *E. capensis*, these taxa are also likely members of this clade (Whitehead *et al.*, 1988; Grant *et al.*, 2005). *Engraulis anchoita* and *E. ringens* together form the basal Engraulini lineage and thus fall outside the marine clade. In summary, none of the polytypic marine genera within Engraulini were monophyletic, with the exceptions of *Cetengraulis*. This incongruence between phylogenetic relationships and current anchovy taxonomy has significant implications for resource management and conservation (Whitehead, 1985), particularly for species of *Engraulis*, which constitute one of the world's largest fisheries (Whitehead, 1985; Whitehead *et al.*, 1988). Several previous biogeographic and evolutionary studies have assumed that marine anchovy taxonomy adequately reflects phylogeny (Grant & Bowen, 1998, 2006; Grant *et al.*, 2005, 2010). Our results clearly show that this is not the case.

Transitions between marine and freshwater biomes

Our molecular phylogeny and habitat reconstruction for anchovies reveals that South American freshwater anchovies are the product of a single evolutionary transition from a marine to freshwater environment. This is a striking result, given that previous taxonomic arrangements suggested six or more invasions of South American freshwaters. Interestingly, several other clades of fishes share a similar pattern of only a single or very few invasions of freshwater by marine lineages into particular geographic area. For example, South American freshwater stingrays are exceptionally diverse (>20 species), found across the entire continent and resulted from a single transition from marine to freshwater (Lovejoy, 1996; Lovejoy *et al.*, 1998). Lovejoy & Collette (2001) argued that needlefishes invaded freshwaters of Amazonia twice, but a single transition has an equal probability under a likelihood model (D. Bloom unpublished data). Herring invaded freshwaters of West Africa between 25 and 50 mya and subsequently spread across the continent to include Lake Tanganyika, and later independently invaded both South Africa and Malagasy (Wilson *et al.*, 2008). Yamanoue *et al.* (2011) found that freshwater pufferfishes in South America, Central Africa and Southeast Asia are each products of single independent invasions. Possible reasons for this pattern are discussed below.

Intriguingly, although a pattern of single invasions of freshwaters is common, anchovies are apparently unique in that they have re-invaded marine habitats. It appears that transitions from marine to freshwater are far more common than freshwater to marine (Vermeij & Dudley, 2000). Indeed, we know of no other instances of reversals back to the marine biome in fishes at similar taxonomic scales (but see Betancur-R, 2010). Although we acknowledge that making these comparisons based on taxonomy is somewhat arbitrary, the absence of a detailed phylogeny for teleost fishes prevents more phylogenetically correct assessments.

Our data on anchovies and evidence from a growing number of studies on other aquatic lineages (Hamilton *et al.*, 2001; Daniels *et al.*, 2006; Heger *et al.*, 2010; Logares *et al.*, 2010; Hou *et al.*, 2011) strongly suggest a pervasive pattern of aquatic biome conservatism, evidenced by evolutionary transitions between biomes occurring far less frequently than lineages remaining in their ancestral biome (Vermeij & Dudley, 2000; Wiens & Donoghue, 2004; Wiens & Graham, 2005; Crisp *et al.*, 2009). However, there is a lack of mechanistic explanations for what might be driving this widespread biogeographic pattern. We suggest that the factors regulating the frequency of transitions between marine and freshwater biomes, and thus the widespread pattern of biome conservatism, include (i) geographic opportunity for invasion, (ii) physiological barriers, (iii) competition and (iv) unique biogeographic events. Below, we discuss these factors with particular reference to the evidence from the freshwater invasion of New World anchovies.

In order for a transition between biomes to occur, there must be geographic opportunity. For example, a lineage occurring in the Antarctic Ocean would be unable to invade the Amazon River, whereas a lineage occurring along of the Atlantic coast of South America would have ample opportunity for invasion (Wiens *et al.*, 2006). Anchovies are found along nearly every coastline in the Western Hemisphere (except polar regions), suggesting that over macroevolutionary time these fishes (and others discussed above) must have had many opportunities to invade freshwater habitats, including those in South America, and yet they failed to do so more than once. Clearly, lack of geographic opportunity does not explain biome conservatism in New World anchovies and many other fishes.

The strongest physiological barrier between marine and freshwater biomes is the salinity gradient – a transition between these biomes requires osmoregulatory adaptations during all stages of a species life cycle (Lee & Bell, 1999). Constraints on the ability to evolve novel osmoregulatory capabilities has likely caused biome conservatism in many clades of fishes (and other aquatic organisms) and resulted in taxa with persistent and ancient associations with either freshwater or marine habitats (Myers, 1949). However, anchovies are members of Clupeiformes, a group that has colonized freshwater habitats numerous times across the globe (Whitehead *et al.*, 1988; Wilson *et al.*, 2008), including several Indo-Pacific anchovies in the subfamily Coilinae. Further, the freshwater invasion of South America occurred early in the diversification of New World anchovies (the second branching event), suggesting that ancestral anchovy lineages were physiologically capable of habitat transitions. Finally, freshwater South American anchovies made three independent re-invasions of marine/estuarine habitats. This evidence suggests that anchovy lineages have long possessed the evolutionary and physiological capacity for adaptation to new salinity regimes. Thus, intrinsic physiological constraints do not offer a good explanation of biome conservatism in anchovies.

Harvey & Pagel (1991) argued that adjacent habitats are rarely invaded because well-adapted incumbent species outcompete invaders. Thus, competition, particularly among closely related species, can drive biome conservatism just as readily as constraints on the fundamental niche (Vermeij & Dudley, 2000; Vermeij & Wesselingh, 2002; Wiens *et al.*, 2006, 2010). If a lineage invades a particular area, diversifies and becomes widespread, open niches will be filled, preventing future invasions (Wiens *et al.*, 2006). The pattern observed in anchovies makes a compelling case for competition, playing a major role in driving biome conservatism. Once anchovies invaded freshwaters of South America, they diversified into a wide array of ecologies, including large-bodied piscivores (*Lycengraulis* and *Pterengraulis*) and several miniature species (*Amazonsprattus* and *Anchoviella spp.*), that are unparalleled by marine engraulid lineages (Whitehead *et al.*, 1988). The freshwater anchovy clade also expanded geographically to nearly every major river basin in South America. Competition as a general driving force behind marine/freshwater biome conservatism is supported by the fact that many other marine-derived freshwater fish lineages show a pattern similar to anchovies. For example, pufferfishes have invaded multiple continents, but never invaded the same continent more than once (Yamanoue *et al.*, 2011). We suggest that the presence of a diverse and widespread anchovy fauna in South America that originated early in the history of New World anchovies has precluded subsequent freshwater invasions by the same clade.

Finally, the conservatism of transitions between habitats may be connected to unique palaeogeographic events. The large numbers of marine-derived freshwater fishes found in South American river systems have been hypothesized to be the product of a continental-scale marine incursion that occurred during the Miocene palaeo-environmental ecosystem known as the Pebas wetland (Lovejoy *et al.*, 1998, 2006; Bloom & Lovejoy, 2011). The Pebas wetland was a spatially and temporally dynamic ecological setting with shifting salinity levels (Hoorn *et al.*, 2010) and may have muted competition with incumbent freshwater lineages, thereby allowing marine invaders to gain a foothold in a new environment (Lovejoy *et al.*, 2006). Further, the fluctuating salinity levels of the Pebas wetland may have provided a fertile landscape for adaptation to freshwater habitats during all life-history stages (Bloom & Lovejoy, 2011). Linking marine invaders with this unique palaeogeographic event requires knowledge of the timing of transition to South American freshwaters (Donoghue & Moore, 2002; Wiens & Donoghue, 2004). A time-calibrated phylogeny was beyond the scope of our study because anchovies have an extremely sparse fossil record (Grande & Nelson, 1985). However, the split between *L. poeyi* (Pacific) and *L. gossidens* (Atlantic) occurs at the tips of the freshwater clade, and the distribution of these taxa on both sides of the Isthmus of Panama requires the age of the freshwater clade to be considerably older than 3.5 mya.

Conclusions

Here, we have demonstrated the effectiveness of phylogenetic approaches for studying the evolution and ecology of biome conservation at a continental scale. Our study highlights the importance of robust phylogenetic inference and cautions against uncritical use of taxonomy for inferring macroevolutionary patterns of biome transition. In contrast to taxonomic expectations, we determined that the remarkable freshwater anchovies of South America are the product of a single evolutionary transition from marine habitats. However, we also found evidence for three independent re-invasions of marine habitats. We propose that the rarity of biome shifts is due neither to limited geographic opportunity nor to physiological constraint. Rather, we believe that competition and palaeogeographic events are the principal factors affecting anchovy habitat evolution. Similar habitat patterns in other aquatic taxa should be investigated with competition and palaeogeography in mind.

Acknowledgments

We thank the following people for generously providing tissue samples and specimens: Hernan Lopez-Fernandez (ROM), Mark Sabaj (ANSP), Kyle Piller (SELU), Sabastien Lavoue, Edward Wiley (KHM), H.J. Walker (SIO), Kevin Lim (RMBR), Anthony Wilson (University of Zurich) and Barbara Brown (AMNH). For assistance in the field, we thank Hernan Lopez-Fernandez, Don Taphorn, Calvin Bernard, Javier Maldonado-Ocampo, Dawn Phillip, Eric Lewallen, Mike Littmann, Alex Flecker, Patrick Traynor and Josh Lovejoy. Chenhong Li offered helpful suggestions for sequencing RAG genes. Hernan Lopez-Fernandez, Belinda Chang and Allan Baker contributed insightful discussions of this project. Belinda Chang graciously helped with the PAML analysis. G. Nelson offered helpful suggestions for identifying anchovy specimens. Members of the Lovejoy Lab and Jason Weir provided comments that greatly improved an earlier version of this manuscript. Financial support was provided by Sigma Xi, Lerner Gray grant (AMNH), Centre for Global Change (University of Toronto), the NSERC Discovery Grant Program and National Science Foundation grant DEB-0614334.

References

Anderson, J.A. & Sues, H.D. 2007. *Major Transitions in Vertebrate Evolution*. Indiana University Press, Bloomington, IN.

Augusto, A., Pinheiro, A.S., Greene, L.J., Laure, H.J. & McNamara, J.C. 2009. Evolutionary transition to freshwater by ancestral marine palaemonids: evidence from osmoregulation in a tide pool shrimp. *Aquatic Biology* **7**: 113–122.

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

Bloom, D.D. & Lovejoy, N.R. 2011. The biogeography of marine incursions in South America. In: *Historical Biogeography of Neotropical Freshwater Fishes* (J.S. Albert & R.E. Reis, eds), pp. 137–144. University of California Press, Berkeley.

Bornbusch, A.H. 1988. Gill raker morphologies of anchovies (Teleostei, Engraulidae) from the Rio Orinoco, Venezuela. *Copeia* **1**: 174–182.

Brandley, M.C., Schmitz, A. & Reeder, T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst. Biol.* **54**: 373–390.

Brown, J.M. & Lemmon, A.R. 2007. The importance of data partitioning and the utility of Bayes Factors in Bayesian phylogenetics. *Syst. Biol.* **56**: 643–655.

Bruen, T., Philippe, H. & Bryant, D. 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**: 2665–2681.

Cassens, I., Vicario, S., Wadde, V.G., Balchowsky, H., Balle, D.V., Wang, D. *et al.* 2000. Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. *Proc. Natl Acad. Sci. USA* **97**: 11343–11347.

Crisp, M. 2006. Biome assembly: what we know and what we need to know. *J. Biogeogr.* **33**: 1332–1333.

Crisp, M.D., Arroyo, M.T.K., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S. *et al.* 2009. Phylogenetic biome conservatism on a global scale. *Nature* **458**: 754–U790.

Daeschler, E.B., Shubin, N.H. & Jenkins, F.A. 2006. A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. *Nature* **440**: 757–763.

Daniels, S.R., Cumberlidge, N., Perez-Losada, M., Marijnissen, S.A.E. & Crandall, K.A. 2006. Evolution of Afrotropical freshwater crab lineages obscured by morphological convergence. *Mol. Phylogenet. Evol.* **40**: 227–235.

Di Dario, F. 2002. Evidence supporting a sister-group relationship between Clupeoidea and Engrauloidea (Clupeomorpha). *Copeia* **2002**: 496–503.

Di Dario, F. 2009. Chirocentrids as engrauloids: evidence from suspensorium, branchial arches, and infraorbital bones (Clupeomorpha, Teleostei). *Zool. J. Linn. Soc.* **156**: 363–383.

Donoghue, M.J. & Moore, B.R. 2002. Toward an integrative historical biogeography. *Integr. Comp. Biol.* **43**: 261–270.

Gingerich, P.D., Haq, Mu., Zalmout, I.S., Khan, I.H. & Malkani, M.S. 2001. Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. *Science* **293**: 2239–2242.

Grande, L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. *Bull. Am. Mus. Nat. His.* **181**: 231–372.

Grande, L. & Nelson, G. 1985. Interrelationships of fossil and recent anchovies (Teleostei: Engrauloidea) and description of a new species from the Miocene of Cyprus. *Am. Mus. Novit.* **2826**: 1–16.

Grant, W.S. & Bowen, B.W. 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J. Hered.* **89**: 415–426.

Grant, W.S. & Bowen, B.W. 2006. Living in a tilted world: climate change and geography limit speciation in Old World anchovies (Engraulis; Engraulidae). *Biol. J. Linn. Soc.* **88**: 673–689.

Grant, W.S., Leslie, R.W. & Bowen, B.W. 2005. Molecular genetic assessment of bipolarity in the anchovy genus Engraulis. *J. Fish Biol.* **67**: 1242–1265.

Grant, W.S., Lecomte, F. & Bowen, B.W. 2010. Biogeographical contingency and the evolution of tropical anchovies (genus Cetengraulis) from temperate anchovies (genus Engraulis). *J. Biogeogr.* **37**: 1352–1362.

Gray, J. 1988. Evolution of the freshwater ecosystem: the fossil record. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **62**: 1–214.

Griffith, R.W. 1987. Fresh-water or marine origin of the vertebrates. *Comp. Biochem. Physiol. A Comp. Physiol.* **87**: 523–531.

Halstead, L.B. 1985. The vertebrate invasion of fresh water. *Philos. Trans. R. Soc. B-Biol. Sci.* **309**: 243–258.

Hamilton, H., Caballero, S., Collins, A.G. & Brownell, R.L. 2001. Evolution of river dolphins. *Proc. R. Soc. Lond. B Biol. Sci.* **268**: 549–556.

Harvey, P.H. & Pagel, M.D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, UK.

Heger, T.J., Mitchell, E.A.D., Todorov, M., Golemansky, V., Lara, E., Leander, B.S. *et al.* 2010. Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Mol. Phylogenet. Evol.* **55**: 113–122.

Hendry, A.P., Lohmann, L.G., Conti, E., Cracraft, J., Crandall, K.A., Faith, D.P. *et al.* 2010. Evolutionary biology in biodiversity science, conservation, and policy: a call to action. *Evolution* **64**: 1517–1528.

Hildebrand, S.F. 1943. A review of the American anchovies (family Engraulidae). *Bull. Bingham oceanogr. Colln* **8**: 1–165.

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J. *et al.* 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.

Hou, Z., Sket, B., Fiser, C. & Li, S. 2011. Eocene habitat shift from saline to freshwater promoted Tethyan amphipod diversification. *Proc. Natl Acad. Sci. USA* **108**: 14533–14538.

Kawahara, R., Miya, M., Mabuchi, K., Near, T.J. & Nishida, M. 2009. Stickleback phylogenies resolved: evidence from mitochondrial genomes and 11 nuclear genes. *Mol. Phylogenet. Evol.* **50**: 401–404.

Lavoue, S., Miya, M., Saitoh, K., Ishiguro, N.B. & Nishida, M. 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. *Mol. Phylogenet. Evol.* **43**: 1096–1105.

Lavoue, S., Miya, M. & Nishida, M. 2009. Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. *Mol. Phylogenet. Evol.* **56**: 480–485.

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

Li, C.H. & Orti, G. 2007. Molecular phylogeny of clupeiformes (Actinopterygii) inferred from nuclear and mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **44**: 386–398.

Logares, R., Shalchian-Tabrizi, K., Boltovskoy, A. & Rengefors, K. 2007. Extensive dinoflagellate phylogenies indicate infrequent marine-freshwater transitions. *Mol. Phylogenet. Evol.* **45**: 887–903.

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.

Logares, R., Brate, J., Heinrich, F., Shalchian-Tabrizi, K. & Bertilsson, S. 2010. Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). *Mol. Biol. Evol.* **27**: 347–357.

Lopez, J.A., Chen, W.J. & Orti, G. 2004. Esociform phylogeny. *Copeia* **44**: 9–464.

Losos, J.B., Leal, M., Glor, R.E., de Queiroz, K., Hertz, P.E., Schettino, L.R. *et al.* 2003. Niche lability in the evolution of a Caribbean lizard community. *Nature* **424**: 542–545.

Lovejoy, N.R. 1996. Systematics of myliobatoid elasmobranchs: with emphasis on the phylogeny and historical biogeography of neotropical freshwater stingrays (Potamotrygonidae:Rajiformes). *Zool. J. Linn. Soc.* **117**: 207–257.

Lovejoy, N.R. & Collette, B.B. 2001. Phylogenetic relationships of new world needlefishes (Teleostei : Belonidae) and the biogeography of transitions between marine and freshwater habitats. *Copeia* **32**: 4–338.

Lovejoy, N.R., Bermingham, E. & Martin, A.P. 1998. Marine incursion into South America. *Nature* **396**: 421–422.

Lovejoy, N.R., Albert, J.S. & Crampton, W.G.R. 2006. Miocene marine incursions and marine/freshwater transitions: evidence from Neotropical fishes. *J. S. Am. Earth Sci.* **21**: 5–13.

Maddison, D.R. & Maddison, W.P. 2001. *MacClade 4: analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA.

Maddison, W.P. & Maddison, D.R. 2009. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org.

Mank, J.E. & Avise, J.C. 2006. Supertree analyses of the roles of viviparity and habitat in the evolution of atherinomorph fishes. *J. Evol. Biol.* **19**: 734–740.

Maynard-Smith, J. & Szathmary, E. 1997. *The Major Transitions in Evolution*. Oxford University Press, New York, NY.

Myers, G.S. 1949. Salt-tolerance of fresh-water fish groups in relation to zoogeographical problems. *Bijdragen tot de Dierkunde* **28**: 315–364.

Nelson, G. 1970. The hyobranchial apparatus of teleostean fishes of the families Engraulidae and Chirocentridae. *Am. Mus. Novit.* **2410**: 1–30.

Nelson, G. 1983. Anchoa-argentivittata, with notes on other Eastern Pacific anchovies and the Indo-Pacific genus Encrasicholina. *Copeia* **1983**: 48–54.

Nelson, G. 1984. Identity of the anchovy Hildebrandichthys-setiger with notes on relationships and biogeography of the genera Engraulis and Cetengraulis. *Copeia* **1984**: 422–427.

Nelson, G. 1986. Identity of the anchovy Engraulis-clarki with notes on the species-groups of Anchoa. *Copeia* **1986**: 891–902.

Niedzwiedzki, G., Szrek, P., Narkiewicz, K., Narkiewicz, M. & Ahlberg, P.E. 2010. Tetrapod trackways from the early Middle Devonian period of Poland. *Nature* **463**: 43–48.

Nielsen, R. & Yang, Z.H. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929–936.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P. & Nieves-Aldrey, J.L. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**: 47–67.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L. & Swofford, D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24**: 581–583.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**: 612–622.

Parry, G. 1966. Osmotic adaptation in fishes. *Biol. Rev. Camb. Philos. Soc.* **41**: 392–444.

Pearse, A.S. 1927. The migration of animals from the ocean into freshwater and land habitats. *Am. Nat.* **61**: 466–476.

Peng, Y.B. & Zhao, Z.R. 1988. A new genus and new species of Chinese anchovies. *Journal of Fisheries of China* **12**: 355–358.

Ramon, M.L. & Knope, M.L. 2008. Molecular support for marine sculpin (Cottidae; Oligocottinae) diversification during the transition from the subtidal to intertidal habitat in the Northeastern Pacific Ocean. *Mol. Phylogenet. Evol.* **46**: 475–483.

Rheindt, F.E., Christidis, L. & Norman, J.A. 2008. Habitat shifts in the evolutionary history of a Neotropical flycatcher lineage from forest and open landscapes. *BMC Evol. Biol.* **8**: 1193.

Ribera, I. & Vogler, A.P. 2000. Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biol. J. Linn. Soc.* **71**: 33–52.

Ribera, I., Barraclough, T.G. & Vogler, A.P. 2001. The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. *Mol. Ecol.* **10**: 721–735.

Roberts, T.R. 1984. *Amazonsprattus scintilla*, a new genus and species from the Rio Nego, Brazil, the smallest known clupeomorph fish. *Proc. Calif. Acad. Sci.* **43**: 317–321.

Robertson, J.D. 1957. The habitat of the early vertebrates. *Biol. Rev. Camb. Philos. Soc.* **32**: 156–187.

Romer, A.S. & Grove, B.H. 1935. Environment of early vertebrates. *Am. Midl. Nat.* **16**: 805–856.

Ronquist, F. & Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Schluter, D. 2000. *The Ecology of Adaptive Radiations*. Oxford University Press, Oxford.

Schluter, D., Price, T., Mooers, A.O. & Ludwig, D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* **51**: 1699–1711.

Shubin, N.H., Daeschler, E.B. & Jenkins, F.A. 2006. The pectoral fin of Tiktaalik roseae and the origin of the tetrapod limb. *Nature* **440**: 764–771.

Stahl, D.A., Key, R., Flesher, B. & Smit, J. 1992. The phylogeny of marine and freshwater caulobacters reflects their habitat. *J. Bacteriol.* **174**: 2193–2198.

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

Sumida, S.S. & Martin, K.L.M., eds 1997. *Amniote Origins: Completing the Transition to Land*. Academic Press, San Diego, CA.

Swofford, D.L. 2002. *PAUP* Phylogeny Analysis Using Parsimony (*and Other Methods), Version 40b10*. Sinauer Association, Inc, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.

Vermeij, G.J. 2005. Invasion as expectation: a historical fact of life. In: *Species Invasions: Insights into Ecology, Evolution, and Biogeography* (D.F. Sax, J.J. Stachowicz & S.D. Gaines, eds), pp. 315–339. Sinauer Associates, Sunderland, MA.

Vermeij, G.J. 2006. Historical contingency and the purported uniqueness of evolutionary innovations. *Proc. Natl Acad. Sci. USA* **103**: 1804–1809.

Vermeij, G.J. & Dudley, R. 2000. Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? *Biol. J. Linn. Soc.* **70**: 541–554.

Vermeij, G.J. & Wesselingh, F.P. 2002. Neogastropod molluscs from the Miocene of western Amazonia, with comments on marine to freshwater transitions in molluscs. *J. Paleontol.* **76**: 265–270.

Weitzman, S.H. & Vari, R.P. 1988. Miniaturization in South American freshwater fishes; an overview and discussion. *Proc. Biol. Soc. Wash.* **101**: 444–465.

Whitehead, P.J. 1973. The clupeoid fishes of the Guianas. *Bull. Brit. Mus. Nat. His.* **5**: 1–227.

Whitehead, P.J.P. 1985. King Herring: his place amongst the clupeoids. *Can. J. Fish. Aquat. Sci.* **42**: 3–20.

Whitehead, A. 2010. The evolutionary radiation of diverse osmotolerant physiologies in killifish (Fundulus Sp.). *Evolution* **64**: 2070–2085.

Whitehead, P.J.P., Nelson, G.J. & Wongratana, T. 1988. *FAO Species Catalogue, Vol. 7. Clupeoid Fishes of the World (Suborder Clupeoidei)*. UNDP FAO, Rome.

Wiens, J.J. & Donoghue, M.J. 2004. Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19**: 639–644.

Wiens, J.J. & Graham, C.H. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Ann. Rev. Ecol. Evol. Syst.* **36**: 519–539.

Wiens, J.J., Brandley, M.C. & Reeder, T.W. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution* **60**: 123–141.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V. *et al.* 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **13**: 1310–1324.

Wilson, A.B., Teugels, G.G. & Meyer, A. 2008. Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. *PLoS ONE* **3**: e1979.

Winemiller, K.O. & Leslie, M.A. 1992. Fish assemblages across a complex, tropical freshwater/marine ecotone. *Environ. Biol. Fish* **34**: 29–50.

Yamanoue, Y., Miya, M., Doi, H., Mabuchi, K., Sakai, H. & Nishida, M. 2011. Multiple invasions into freshwater by pufferfishes (teleostei: tetraodontidae): a mitogenomic perspective. *PLoS ONE* **6**: e17410.

Yang, Z.H. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.

Yang, Z.H., Nielsen, R., Goldman, N. & Pedersen, A.M.K. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.

Yoder, J.B., Clancey, E., Roches, S.D., Eastman, J.M., Gentry, L., Godsoe, W. *et al.* 2010. Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.* **23**: 1581–1596.

Yokoyama, R. & Goto, A. 2005. Evolutionary history of freshwater sculpins, genus Cottus (Teleostei : Cottidae) and related taxa, as inferred from mitochondrial DNA phylogeny. *Mol. Phylogenet. Evol.* **36**: 654–668.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List of specimens included in the study with corresponding GenBank accession and museum catalogue numbers.

Appendix S2 Sequence length, parsimony-informative sites, nucleotide substitution model, descriptive statistics and MP score for each gene and the combined data set.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 9 September 2011; revised 20 December 2011; accepted 3 January 2012