Syst. Biol. 73(4):683–703, 2024

https://doi.org/10.1093/sysbio/syae022
Advance Access Publication May 17, 2024

Phylogenomics, Lineage Diversification Rates, and the Evolution of Diadromy in Clupeiformes (Anchovies, Herrings, Sardines, and Relatives)

Joshua P. Egan[1,2,*], Andrew M. Simons[2,3], Mohammad Sadegh Alavi-Yeganeh[4], Michael P. Hammer[5], Prasert Tongnunui[6], Dahiana Arcila[7], Ricardo Betancur-R[7], and Devin D. Bloom[1,8,*]

[1]*Department of Biological Sciences, Western Michigan University, 1903 W Michigan Ave., Kalamazoo, MI 49008, USA*
[2]*Bell Museum of Natural History, University of Minnesota, 100 Ecology, 1987 Upper Buford Circle, Saint Paul, MN 55108, USA*
[3]*Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, 2003 Upper Buford Circle, Saint Paul, MN 55108, USA*
[4]*Department of Marine Biology, Tarbiat Modares University, Nur 4641776489, Iran*
[5]*Museum and Art Gallery of the Northern Territory, GPO Box 4646, Darwin, NT 0801, Australia*
[6]*Department of Marine Science and Environment, Faculty of Science and Fisheries Technology, Rajamangala University of Technology Srivijaya, Sikao, Trang 92150, Thailand*
[7]*Scripps Institution of Oceanography, University of California - San Diego, La Jolla, CA 92093, USA*
[8]*School of the Environment, Geography, and Sustainability, Western Michigan University, 1903 W Michigan Ave, Kalamazoo, MI 49008, USA*
[*]*Correspondence to be sent to: Department of Biological Sciences, Western Michigan University, 1903 W Michigan Ave., Kalamazoo, MI 49008, USA; E-mail: eganx149@umn.edu (J.P.E.); Department of Biological Sciences, Western Michigan University, 1903 W Michigan Ave., Kalamazoo, MI 49008, USA; E-mail: devin.bloom@wmich.edu (D.D.B.).*

Received 16 August 2021; reviews returned 1 May 2024; accepted 15 May 2024
Associate Editor: James Albert

Abstract.—Migration independently evolved numerous times in animals, with a myriad of ecological and evolutionary implications. In fishes, perhaps the most extreme form of migration is diadromy, the migration between marine and freshwater environments. Key and long-standing questions are: how many times has diadromy evolved in fishes, how frequently do diadromous clades give rise to non-diadromous species, and does diadromy influence lineage diversification rates? Many diadromous fishes have large geographic ranges with constituent populations that use isolated freshwater habitats. This may limit gene flow between some populations, increasing the likelihood of speciation in diadromous lineages relative to nondiadromous lineages. Alternatively, diadromy may reduce lineage diversification rates if migration is associated with enhanced dispersal capacity that facilitates gene flow within and between populations. Clupeiformes (herrings, sardines, shads, and anchovies) is a model clade for testing hypotheses about the evolution of diadromy because it includes an exceptionally high proportion of diadromous species and several independent evolutionary origins of diadromy. However, relationships among major clupeiform lineages remain unresolved, and existing phylogenies sparsely sampled diadromous species, limiting the resolution of phylogenetically informed statistical analyses. We assembled a phylogenomic dataset and used multi-species coalescent and concatenation-based approaches to generate the most comprehensive, highly resolved clupeiform phylogeny to date, clarifying associations among several major clades and identifying recalcitrant relationships needing further examination. We determined that variation in rates of sequence evolution (heterotachy) and base-composition (nonstationarity) had little impact on our results. Using this phylogeny, we characterized evolutionary patterns of diadromy and tested for differences in lineage diversification rates between diadromous, marine, and freshwater lineages. We identified 13 transitions to diadromy, all during the Cenozoic Era (10 origins of anadromy, 2 origins of catadromy, and 1 origin of amphidromy), and 7 losses of diadromy. Two diadromous lineages rapidly generated nondiadromous species, demonstrating that diadromy is not an evolutionary dead end. We discovered considerably faster transition rates out of diadromy than to diadromy. The largest lineage diversification rate increase in Clupeiformes was associated with a transition to diadromy, but we uncovered little statistical support for categorically faster lineage diversification rates in diadromous versus nondiadromous fishes. We propose that diadromy may increase the potential for accelerated lineage diversification, particularly in species that migrate long distances. However, this potential may only be realized in certain biogeographic contexts, such as when diadromy allows access to ecosystems in which there is limited competition from incumbent species. [Divergence time; ecological opportunity; exon capture; migration; next-generation sequencing; speciation; State-dependent diversification.]

Migration evolved independently and repeatedly throughout the animal tree of life, with wide-ranging ecological and evolutionary implications (Alerstam et al. 2003; Fudickar et al. 2021). For example, migration influenced the evolution of body size and locomotor traits in fishes (Burns and Bloom 2020; DeHaan et al. 2023; Finnegan et al. 2024), wing morphology in insects (Flockhart et al. 2017), and lineage diversification rates in birds (Claramunt et al. 2012; Rolland et al. 2014). In fishes, perhaps the most extreme form of migration is diadromy, the migration between marine and freshwater environments. There are 3 forms of diadromy: 1) amphidromous fishes undergo migrations between marine and freshwater that are variable in timing and not for reproduction, 2) anadromous fishes reproduce in freshwater and migrate to the ocean for feeding and growth, and 3) catadromous species reproduce in the ocean and migrate to freshwaters for feeding and

growth (McDowall 1988). Diadromous migration distances vary widely among populations and species. For example, American shad (*Alosa sapidissima*) and sockeye salmon (*Oncorhynchus nerka*) ascend rivers up to ~1100 and 1448 km, respectively, although species such as the Atlantic sabretooth anchovy (*Lycengraulis grossidens*) and giant bully (*Gobiomorphus gobioides*) generally travel no more than 100 km upstream (McDowall 1997; Limburg et al. 2003; Kline and Flagg 2014; Mai et al. 2014). Diadromy is present in nearly all major fish lineages and geographically widespread (McDowall 1988; Corush 2019; Alò et al. 2021). Therefore, diadromy may have played a role in shaping patterns of diversity within many clades of fishes (McDowall 1988; Willson and Hulupka 1995; Bloom and Lovejoy 2014; Bloom et al. 2018; Corush 2019; Burns and Bloom 2020; Burridge and Waters 2020; Delgado and Ruzzante 2020; Alò et al. 2021; DeHaan et al. 2023).

A key and long-standing question in evolutionary ecology is the extent to which migration, including diadromy, influences lineage diversification rates (McDowall 1988; Rolland et al. 2014; Corush 2019). Many diadromous species travel long distances during migration, have large geographic ranges, and use isolated freshwater habitats, which may increase the likelihood of both ecological and nonecological speciation (McDowall 1988, 2001). Alternatively, diadromy may reduce lineage diversification rates if migration is associated with the evolution of enhanced dispersal capacity that facilitates gene flow within and between populations (Roff 1991; McDowall 2001; Kisel and Barraclough 2010; Hasselman et al. 2013). Additionally, if the macroevolutionary process of lineage diversification is not always tightly linked to microevolutionary processes such as gene flow, relationships between diadromy and diversification rates may be inconsistent (Nosil 2008). Using a dataset spanning ray-finned fishes (Teleostei), Corush (2019) found that mean lineage diversification rates were faster in diadromous than nondiadromous fishes. However, studies on smaller teleost clades have failed to detect faster rates of lineage diversification associated with diadromy (Burridge and Waters 2020; Thacker et al. 2021). Therefore, diadromy may increase a clade's potential for diversification, but relationships between diadromy and diversification rates may vary widely among teleost clades, mediated by biogeographic settings and phenotypic traits, such as migration distances (Hasselman et al. 2013; Corush 2019; Burridge and Waters 2020).

The evolution of an extreme mode of migration, such as diadromy, may evolve rarely due to physiological constraints and high energetic demands associated with this life history strategy (Roff 1988, 1991; Bowlin et al. 2010). Empirical evidence supports this hypothesis. Only ~450 of more than 35,000 fish species are diadromous (Delgado and Ruzzante 2020). Furthermore, Corush (2019) found that across ray-finned fishes, transition rates to diadromy were considerably slower than transition rates out of diadromy. This suggests that evolving diadromy is unlikely due to the complexity of this behavior, but that once evolved, there is a high propensity for speciation leading to nondiadromous species nested within diadromous clades. This pattern of slower transition rates to migratory, rather than out of migratory life histories, has also been documented in other clades. For example, there is evidence that transitions from migratory to nonmigratory life histories occurred much more frequently than transition from nonmigratory to migratory life histories in birds (Rolland et al. 2014; Winger et al. 2014; Dufour et al. 2020) and bats (family: Vespertilionidae; Bisson et al. 2009). Exploring the relative transition rates between diadromy and nondiadromy and testing for differences in lineage diversification rates between diadromous and nondiadromous lineages are key steps to understanding the macroevolutionary dynamics of this life history strategy.

Clupeiformes is an ecologically diverse clade containing ~443 species of anchovies, sardines, herrings, and their relatives that are globally distributed in freshwater and marine environments (Whitehead 1985; Whitehead et al. 1988; Egan et al. 2018a, 2018b, 2022; Arnette et al. 2024). Clupeiformes include an exceptionally high proportion of diadromous species (at least an order of magnitude higher than other major fish lineages; Delgado and Ruzzante 2020). There is evidence for at least 10 independent evolutionary origins of diadromy in clupeiforms (2 origins of catadromy and 8 origins of anadromy) and there may be additional transitions that have yet to be documented (McDowall 1988; Bloom and Lovejoy 2014; Bloom et al. 2018). Because diadromy evolved repeatedly within Clupeiformes, this study system affords considerable statistical power for conducting phylogenetic comparative analyses of this trait (Felsenstein 2004; Palkovacs et al. 2008; Faria et al. 2012; Bloom and Lovejoy 2014; Twining et al. 2017; Bloom et al. 2018; Velotta et al. 2018). Although Clupeiformes is an excellent study system for testing evolutionary hypotheses about diadromy, relationships among many lineages remain uncertain due to small numbers of loci and insufficient taxon sampling used in previous molecular studies (Bloom and Lovejoy 2012, 2014; Lavoué et al. 2013, 2014, 2017a, 2017b; Bloom and Egan 2018; Egan et al. 2018a; Wang et al. 2022). A recent study sampled 1165 loci and resolved relationships among several major clades, but only sampled 66 species (Fig. 1; Wang et al. 2022). The clupeiform phylogeny with the most comprehensive taxon sampling included 191 species (43% of recognized species and < 50% of diadromous species), but only sampled 4 loci and failed to resolve associations among multiple clades (Egan et al. 2018a). Inferring a more robust clupeiform phylogeny that samples both many loci and species (especially diadromous taxa) will allow for more accurate inferences about the evolutionary patterns of diadromy.

Here, we assembled a phylogenomic dataset and used multi-species coalescent (MSC) and concatenation-based approaches to generate a comprehensive

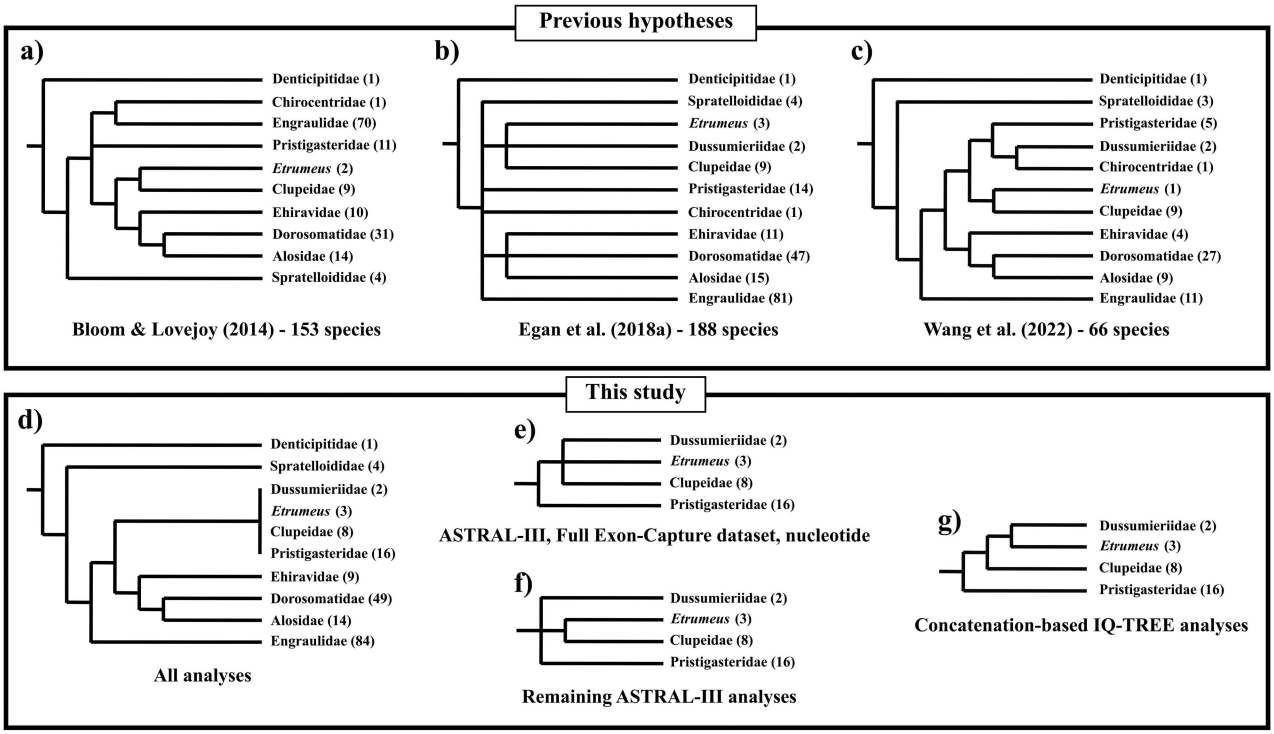

FIGURE 1. Hypothesized relationships among major clupeiform clades inferred by (a–c) selected published studies, and (d–g) this study. We recognize the revised clupeiform classification described by Wang et al. (2022) and the change of *Thryssa* to *Thrissina* (Kottelat 2013). We collapsed branches with low support (bootstrap or posterior probability values < 0.85), resulting in polytomies. Numbers in parentheses following tip labels denote the number of species sampled in each lineage. a) Bloom and Lovejoy (2014) and b) Egan et al. (2018a) used mitochondrial and nuclear loci and Bayesian multi-species coalescent approaches; and c) Wang et al. (2022) used nuclear loci and concatenation-based approaches. The present study recovered 3 different topologies through phylogenetic analyses of the Full Exon-Capture and Reduced Locus datasets. d) All analyses recovered the same relationships among most major lineages. However, there was variation among analyses in the positions of Dussumieriidae, Clupeidae (sans *Etrumeus*), *Etrumeus*, and Pristigasteridae. e) The topology recovered by multi-species coalescent ASTRAL-III analysis of the Full Exon-Capture Dataset using fully resolved nucleotide-based gene trees as input. f) The topology recovered by the 4 remaining ASTRAL-III analyses (Full Exon-Capture Dataset using amino-acid-based gene trees with poorly supported branches collapsed (UFBoot support < 33%), Full Exon-Capture Dataset using fully resolved amino-acid-based gene trees, Full Exon-Capture Dataset using nucleotide-based gene trees with poorly supported branches collapsed, and analysis of Reduced Locus Dataset using fully resolved nucleotide-based gene trees). g) The topology inferred using concatenation-based IQ-TREE analyses of the Full Exon-Capture Dataset, based on both amino acid and nucleotide alignments.

clupeiform phylogeny for testing hypotheses about the evolution of diadromy. We used biogeographic and fossil node calibrations for Bayesian relaxed-clock estimation of clupeiform divergence times. We then used our time-calibrated phylogeny and a suite of phylogenetic comparative methods to accomplish 2 objectives: (1) infer the timing, patterns, and rates of transitions between diadromous, marine, and freshwater life histories; and (2) test for differences in lineage diversification rates between diadromous, marine, and freshwater clades.

Materials and Methods

DNA Sequencing, Exon Assembly, and Alignment

We extracted genomic DNA from muscle and fin tissue using Qiagen® DNAeasy Blood and Tissue Kits (Qiagen, Valencia, CA) following the manufacturer's protocol. We used the "Backbone 1" probe set designed by Hughes et al. (2018, 2021) to target 1104 exons. Library preparation using a dual-round capture protocol (Li et al. 2013), target enrichment, and Illumina HiSeq 2500 paired-end sequencing was performed by Arbor Biosciences (Arbor Biosciences, Ann Arbor, MI, United States). We used the automated bioinformatics pipeline developed by Hughes et al. (2021) for quality-controlled exon assembly and alignment (see Supplementary Methods for details; Supplementary Material, including data files, can be found in the Dryad data repository at https://doi.org/10.5061/dryad.47d7wm3h8).

Taxonomic Sampling and Phylogenomic Datasets

We assembled 4 datasets for phylogenomic analyses: (i) "Full Exon-Capture Dataset," (ii) "Reduced Locus Dataset," (iii) Expanded Dataset," and (iv) "30-Species Dataset." To resolve clupeiform phylogenetic relationships and estimate divergence times, we assembled the

TABLE 1. Species shown in Fig. 2 with biogeographic and character state information

Family	Species	Diadromy	Biogeographic region
Engraulidae	*Lycengraulis batesii*	Freshwater	WA
Engraulidae	*Lycengraulis grossidens*	Anadromous	WA, SSA
Engraulidae	*Lycengraulis poeyi*	Marine	EP
Engraulidae	*Pterengraulis atherinoides*	Freshwater	WA
Engraulidae	*Anchoviella lepidentostole*	Anadromous	WA
Engraulidae	*Anchovia surinamensis*	Freshwater	WA
Engraulidae	*Anchoviella jamesi*	Freshwater	WA
Engraulidae	*Anchoviella carrikeri*	Freshwater	WA
Engraulidae	*Anchoviella alleni*	Freshwater	WA
Engraulidae	*Anchoviella guianensis*	Freshwater	WA
Engraulidae	*Anchoviella manamensis*	Freshwater	WA
Engraulidae	*Anchoviella brevirostris*	Marine	WA
Engraulidae	*Anchoviella* sp. (Rio Napo)	Freshwater	WA
Engraulidae	*Anchoviella hernanni*	Freshwater	WA
Engraulidae	*Anchoviella* sp. (Lawa River)	Freshwater	WA
Engraulidae	*Anchoviella* sp. 2	Freshwater	WA
Engraulidae	*Amazonsprattus scintilla*	Freshwater	WA
Engraulidae	*Anchoviella juruasanga*	Freshwater	WA
Engraulidae	*Anchoa spinifer*	Marine	WA
Engraulidae	*Jurengraulis juruensis*	Freshwater	WA
Engraulidae	*Anchoa mundeoloides*	Marine	NP
Engraulidae	*Anchoa lucida*	Marine	NP, EP
Engraulidae	*Anchoa compressa*	Marine	NP
Engraulidae	*Anchoviella balboae*	Marine	EP
Engraulidae	*Anchoa panamensis*	Marine	EP
Engraulidae	*Anchoa scofieldi*	Marine	NP
Engraulidae	*Anchoa walkeri*	Marine	NP, EP
Engraulidae	*Anchoa colonensis*	Marine	WA
Engraulidae	*Anchoa ischana*	Marine	NP, EP
Engraulidae	*Anchoa hepsetus*	Marine	WA
Engraulidae	*Anchoviella perfasciata*	Marine	WA
Engraulidae	*Anchoa argentivittata*	Marine	NP, EP
Engraulidae	*Anchovia macrolepidota*	Marine	NP, EP
Engraulidae	*Anchovia clupeoides*	Marine	WA
Engraulidae	*Anchoa chamensis*	Marine	EP
Engraulidae	*Anchoa lamprotaenia*	Marine	WA
Engraulidae	*Anchoa delicatissima*	Marine	NP
Engraulidae	*Anchoa parva*	Marine	WA
Engraulidae	*Anchoa cayorum*	Marine	WA
Engraulidae	*Anchoa cubana*	Marine	WA
Engraulidae	*Anchoviella elongata*	Marine	WA
Engraulidae	*Cetengraulis mysticetus*	Marine	NP, EP
Engraulidae	*Cetengraulis edentulus*	Marine	WA
Engraulidae	*Anchoa nasus*	Marine	NP, EP, SSA
Engraulidae	*Anchoa lyolepis*	Marine	WA
Engraulidae	*Engraulis mordax*	Marine	NP
Engraulidae	*Engraulis albidus*	Marine	NEA
Engraulidae	*Engraulis encrasicolus*	Marine	NEA, EA, SA
Engraulidae	*Engraulis eurystole*	Marine	NWA, WA, SSA
Engraulidae	*Engraulis australis*	Marine	SAU, NZ
Engraulidae	*Anchoa filifera*	Marine	WA
Engraulidae	*Engraulis ringens*	Marine	SSA
Engraulidae	*Engraulis anchoita*	Marine	WA, SSA
Engraulidae	*Encrasicholina purpurea*	Marine	IWP
Engraulidae	*Encrasicholina punctifer*	Marine	IWP
Engraulidae	*Encrasicholina* sp.	Marine	IWP
Engraulidae	*Stolephorus commersonnii*	Marine	IWP
Engraulidae	*Stolephorus bataviensis*	Marine	IWP
Engraulidae	*Stolephorus nelsoni*	Marine	IWP
Engraulidae	*Stolephorus lotus*	Marine	IWP
Engraulidae	*Stolephorus carpentariae*	Marine	IWP
Engraulidae	*Stolephorus brachycephalus*	Marine	IWP
Engraulidae	*Stolephorus bengalensis*	Marine	IWP
Engraulidae	*Stolephorus holodon*	Marine	SA
Engraulidae	*Stolephorus dubiosus*	Marine	IWP
Engraulidae	*Thrissina vitrirostris*	Marine	IWP
Engraulidae	*Thrissina mystax*	Marine	IWP
Engraulidae	*Thrissina hamiltonii*	Marine	IWP
Engraulidae	*Thrissina spinidens*	Marine	IWP

Table 1. Continued

Family	Species	Diadromy	Biogeographic region
Engraulidae	*Thrissina kammalensis*	Marine	IWP
Engraulidae	*Thrissina setirostris*	Marine	IWP
Engraulidae	*Thrissina dussumieri*	Marine	IWP
Engraulidae	*Setipinna crocodilus*	Freshwater	IWP
Engraulidae	*Setipinna melanochir*	Marine	IWP
Engraulidae	*Setipinna taty*	Marine	IWP
Engraulidae	*Setipinna tenuifilis*	Marine	IWP
Engraulidae	*Thrissina nasuta*	Marine	IWP
Engraulidae	*Thrissina rastrosa*	Freshwater	IWP
Engraulidae	*Papuengraulis micropinna*	Marine	IWP
Engraulidae	*Thrissina brevicauda*	Marine	IWP
Engraulidae	*Thrissina scratchleyi*	Freshwater	IWP
Engraulidae	*Thrissina encrasicholoides*	Marine	IWP
Engraulidae	*Thrissina chefuensis*	Marine	IWP
Engraulidae	*Coilia dussumieri*	Marine	IWP
Dorosomatidae	*Opisthonema libertate*	Marine	NP, EP
Dorosomatidae	*Opisthonema medirastre*	Marine	NP, EP
Dorosomatidae	*Opisthonema oglinum*	Marine	NWA, WA
Dorosomatidae	*Dorosoma petenense*	Freshwater	WA
Dorosomatidae	*Dorosoma cepedianum*	Freshwater	NWA, WA
Dorosomatidae	*Harengula humeralis*	Marine	NWA, WA
Dorosomatidae	*Harengula thrissina*	Marine	NP, EP
Dorosomatidae	*Sardinella aurita*	Marine	NWA, NEA, WA, PC, SSA, EA
Dorosomatidae	*Rhinosardinia bahiensis*	Freshwater	WA
Dorosomatidae	*Rhinosardinia amazonica*	Freshwater	WA
Dorosomatidae	*Lile stolifera*	Marine	NP, EP
Dorosomatidae	*Platanichthys platana*	Freshwater	WA, SSA
Dorosomatidae	*Sardinella longiceps*	Marine	IWP
Dorosomatidae	*Sardinella lemuru*	Marine	IWP, NP
Dorosomatidae	*Sardinella maderensis*	Marine	NEA, EA
Dorosomatidae	*Sardinella gibbosa*	Marine	IWP
Dorosomatidae	*Sardinella brachysoma*	Marine	IWP
Dorosomatidae	*Sardinella melanura*	Marine	IWP
Dorosomatidae	*Sardinella marquesensis*	Marine	IWP
Dorosomatidae	*Hilsa kelee*	Anadromous	IWP
Dorosomatidae	*Nematalosa vlaminghi*	Marine	IWP
Dorosomatidae	*Nematalosa erebi*	Freshwater	IWP
Dorosomatidae	*Nematalosa come*	Marine	IWP
Dorosomatidae	*Nematalosa nasus*	Marine	IWP
Dorosomatidae	*Nematalosa japonica*	Marine	IWP
Dorosomatidae	*Nematalosa galatheae*	Marine	IWP
Dorosomatidae	*Tenualosa toli*	Anadromous	IWP
Dorosomatidae	*Tenualosa ilisha*	Anadromous	IWP
Dorosomatidae	*Tenualosa thibaudeaui*	Freshwater	IWP
Dorosomatidae	*Gudusia chapra*	Freshwater	IWP
Dorosomatidae	*Anodontostoma chacunda*	Marine	IWP
Dorosomatidae	*Potamothrissa obtusirostris*	Freshwater	EA
Dorosomatidae	*Potamothrissa acutirostris*	Freshwater	EA
Dorosomatidae	*Limnothrissa miodon*	Freshwater	EA
Dorosomatidae	*Pellonula leonensis*	Freshwater	EA
Dorosomatidae	*Pellonula vorax*	Freshwater	EA
Dorosomatidae	*Ethmalosa fimbriata*	Catadromous	EA
Dorosomatidae	*Herklotsichthys blackburni*	Marine	IWP
Dorosomatidae	*Herklotsichthys gotoi*	Marine	IWP
Dorosomatidae	*Herklotsichthys koningsbergeri*	Marine	IWP
Dorosomatidae	*Herklotsichthys lippa*	Marine	IWP
Dorosomatidae	*Herklotsichthys quadrimaculatus*	Marine	IWP
Dorosomatidae	*Herklotsichthys collettei*	Marine	IWP
Dorosomatidae	*Amblygaster sirm*	Marine	IWP
Dorosomatidae	*Amblygaster leiogaster*	Marine	IWP
Dorosomatidae	*Clupeoides papuensis*	Freshwater	IWP
Dorosomatidae	*Clupeoides venulosus*	Freshwater	IWP
Dorosomatidae	*Escualosa thoracata*	Marine	IWP
Dorosomatidae	*Herklotsichthys dispilonotus*	Marine	IWP
Alosidae	*Alosa braschnikowi*	Marine	PC
Alosidae	*Alosa caspia*	Anadromous	PC
Alosidae	*Alosa fallax*	Anadromous	NEA
Alosidae	*Alosa alosa*	Anadromous	NEA

Table 1. Continued

Family	Species	Diadromy	Biogeographic region
Alosidae	*Alosa sapidissima*	Anadromous	NWA
Alosidae	*Alosa alabamae*	Anadromous	WA
Alosidae	*Alosa chrysochloris*	Freshwater	WA
Alosidae	*Alosa mediocris*	Anadromous	NWA
Alosidae	*Alosa pseudoharengus*	Anadromous	NWA
Alosidae	*Brevoortia smithi*	Marine	WA
Alosidae	*Brevoortia pectinata*	Marine	SSA
Alosidae	*Brevoortia patronus*	Marine	WA
Alosidae	*Sardinops sagax*	Marine	NP, SSA, SA, SAU, NZ
Alosidae	*Sardina pilchardus*	Marine	NEA
Ehiravidae	*Corica soborna*	Freshwater	IWP
Ehiravidae	*Corica laciniata*	Freshwater	IWP
Ehiravidae	*Clupeoides borneensis*	Freshwater	IWP
Ehiravidae	*Clupeichthys aesarnensis*	Freshwater	IWP
Ehiravidae	*Clupeichthys perakensis*	Freshwater	IWP
Ehiravidae	*Sundasalanx mekongensis*	Freshwater	IWP
Ehiravidae	*Gilchristella aestuaria*	Amphidromous	SA
Ehiravidae	*Clupeonella cultriventris*	Anadromous	PC
Ehiravidae	*Clupeonella caspia*	Anadromous	PC
Pristigasteridae	*Pristigaster cayana*	Freshwater	WA
Pristigasteridae	*Pristigaster whiteheadi*	Freshwater	WA
Pristigasteridae	*Ilisha amazonica*	Freshwater	WA
Pristigasteridae	*Pellona flavipinnis*	Freshwater	WA, SSA
Pristigasteridae	*Pellona castelnaeana*	Freshwater	WA
Pristigasteridae	*Pellona harroweri*	Marine	WA
Pristigasteridae	*Ilisha megaloptera*	Marine	IWP
Pristigasteridae	*Ilisha* sp. (Thailand)	Marine	IWP
Pristigasteridae	*Ilisha elongata*	Marine	IWP, NP
Pristigasteridae	*Ilisha melastoma*	Marine	IWP
Pristigasteridae	*Ilisha* sp. (Indonesia)	Marine	IWP
Pristigasteridae	*Pellona ditchela*	Marine	IWP
Pristigasteridae	*Odontognathus mucronatus*	Marine	WA
Pristigasteridae	*Odontognathus compressus*	Marine	WA
Pristigasteridae	*Pliosteostoma lutipinnis*	Marine	EP
Pristigasteridae	*Chirocentrodon bleekerianus*	Marine	WA
Clupeidae	*Clupea pallasii*	Marine	NP
Clupeidae	*Clupea harengus*	Marine	NWA, NEA
Clupeidae	*Sprattus sprattus*	Marine	NEA
Clupeidae	*Ramnogaster melanostoma*	Marine	SSA
Clupeidae	*Ramnogaster arcuata*	Marine	SSA
Clupeidae	*Sprattus fuegensis*	Marine	SSA
Clupeidae	*Potamalosa richmondia*	Catadromous	SAU
Clupeidae	*Hyperlophus vittatus*	Marine	SAU
Clupeidae	*Etrumeus sadina*	Marine	NWA, WA
Clupeidae	*Etrumeus golanii*	Marine	IWP
Clupeidae	*Etrumeus whiteheadi*	Marine	SA
Dussumieriidae	*Dussumieria acuta*	Marine	IWP
Dussumieriidae	*Dussumieria elopsoides*	Marine	IWP
Spratelloididae	*Spratelloides delicatulus*	Marine	IWP
Spratelloididae	*Spratelloides robustus*	Marine	SAU
Spratelloididae	*Jenkinsia stolifera*	Marine	WA
Spratelloididae	*Jenkinsia majua*	Marine	WA
Denticipitidae	*Denticeps clupeoides*	Freshwater	IWP

Notes: Species are listed in their order of appearance along the tips of the phylogeny in Fig. 2, from top to bottom. Diadromy = 3-state diadromy character states. Note that we generated Fig. 2 using the two-state diadromy character: nondiadromous (marine and freshwater species) versus diadromous (anadromous, catadromous, and amphidromous species). Biogeographic region = the biogeographic regions in which each species occurs (Northwest Atlantic = NWA, Northeast Atlantic = NEA, West Atlantic = WA, Ponto-Caspian = PC, Indo-West Pacific = IWP, North Pacific = MP, East Pacific = EP, South South America = SSA, East Atlantic = EA, South Africa = SA, South Australia = SAU, and New Zealand = NZ).

Full Exon-Capture Dataset containing all 1071 target exon loci that were successfully sequenced via target capture and passed quality control steps. This dataset included 190 of 443 clupeiform species, 62 of 82 genera, 6 of 7 families, and 16 of 29 diadromous species (Table 1; Supplementary Table S1). This dataset also contained sequences for 5 outgroup taxa downloaded from NCBI: *Scleropages formosus* (Osteoglossiformes), *Danio rerio* (Cypriniformes), *Chanos chanos* (Gonorynchiformes), *Astyanax mexicanus* (Characiformes), and *Ictalurus punctatus* (Siluriformes). We attempted to include Chirocentridae in the Full Exon-Capture Dataset, but

Table 2. Summary of key phylogenetic analyses used to infer species trees

Dataset	Concatenation-based analyses	Multi-species coalescent analyses	Time-calibrated species tree
Full Exon-Capture (Nt)	Yes	Yes	Yes
Full Exon-Capture (AA)	Yes	Yes	No
Reduced Locus (Nt)	No	Yes	No
Expanded (Nt)	No	Yes	Yes
Expanded (AA)	No	Yes	No
30-Species (Nt)	No	Yes	No

Notes: We inferred species trees using 4 datasets: (1) Full Exon-Capture (190 species, 1,071 loci), (2) Reduced Locus (190 species, 613 loci), (3) Expanded (272 species; 1,076 loci (nuclear markers + 5 additional mitochondrial markers)), and (4) 30-Species (30 species, 252 loci). Concatenation-based species tree analyses were conducted in IQ-TREE. Multi-species coalescent species tree analyses were conducted in ASTRAL-III. Time calibration was done with MCMCTree. Nt = nucleotide alignment and AA = amino acid alignment.

Chirocentrus samples failed to sequence. To assess the impact of missing sequence data in gene alignments on phylogenetic analyses, we assembled the Reduced Locus Dataset containing the same taxon sampling as the Full Exon-Capture Dataset, but only the 613 loci with at least 100 species in alignments.

To generate a phylogeny that maximizes taxonomic sampling for comparative analyses, we assembled the Expanded Dataset containing all 1071 target exon loci and additional sequences downloaded from the National Center for Biotechnology Information (NCBI) sequence database (Sayers et al. 2022) for 7 genetic markers (2 nuclear markers also targeted by the Backbone 1 probe set and 5 additional mitochondrial markers not targeted by the Backbone 1 probe set): *rag1*, *rag2*, *cytb*, *ATP6*, *COI*, *COII*, and *COIII* (Supplementary Table S1). This dataset included 272 clupeiform species, 77 genera, all 7 families (including Chirocentridae), 26 of 29 diadromous clupeiform species (23 anadromous species, 2 catadromous species) (*Potamalosa richmondia* and *Ethmalosa fimbriata*, and one amphidromous species [*Gilchristella aestuaria*]), and the same outgroup taxa as the Full Exon-Capture Dataset and Reduced Locus Dataset.

Heterogeneity in evolutionary processes, such as variation among lineages in rates of sequence evolution (heterotachy) and base-composition (nonstationarity/compositional bias), can confound phylogenetic analyses (Lockhart et al. 2006; Rodriguez-Ezpeleta et al. 2007; Betancur-R et al. 2013). To explore potential spurious inferences due to nonstationarity, we assembled amino acid alignments in addition to nucleotide alignments for the Full Exon-Capture Dataset and Reduced Locus Dataset. Previous research found modest variation in base-composition among clupeiform lineages and that base-composition had little impact on phylogenetic inferences (Wang et al. 2022). Therefore, we analyzed amino acid alignments, which can effectively reduce the impacts of moderate compositional bias on phylogenetic inferences (Hasegawa and Hashimoto 1993; Jeffroy et al. 2006; Supplementary Table S1). We translated nucleotide alignments into amino acid alignments using AliView v1.26 (Larsson 2014). Analyses with the Full Exon-Capture, Reduced Locus, and Expanded datasets used homogeneous models (i.e., assumed homotachous sequence evolution) due to computational constraints. To investigate the impacts of heterotachous sequence evolution on phylogenetic inferences, we assembled the 30-Species Dataset. This dataset was a subset of the Reduced Locus Dataset, containing 30 species representing every major clupeiform lineage and 252 loci with at least 25 species in alignments (Supplementary Table S2). We used this dataset to conduct computationally intensive analyses that accounted for heterotachous sequence evolution through nonhomogeneous models of sequence evolution (see below and Supplementary Methods).

Phylogenetic Inference

We used both concatenation-based (all genes assembled into a single alignment and constrained to a single evolutionary history) and summary MSC (genes have separate alignments and independent evolutionary histories) approaches to infer phylogenetic relationships. We estimated gene trees with maximum likelihood analyses in IQ-TREE v1.6.9 for the Full Exon-Capture, Reduced Locus, Expanded, and 30-Species datasets using both nucleotide alignments and amino acid alignments as input (Nguyen et al. 2015). In addition to estimating individual gene trees for mitochondrial genes, we concatenated mitochondrial gene alignments and inferred a single mitochondrial tree using IQ-TREE. This allowed us to constrain mitochondrial genes to a shared evolutionary history in summary MSC analyses. We also used IQ-TREE to generate species trees by concatenating gene alignments for the Full Exon Capture Dataset, based on both nucleotide and amino acid alignments (Table 2). We conducted 10 independent searches for each concatenation-based analysis. We partitioned nucleotide alignments by codon position and used the built-in ModelFinder TESTMERGE option in IQ-TREE to select homogeneous nucleotide substitution models for each partition (Kalyaanamoorthy et al. 2017). Additionally, for the 30-Species Dataset, we estimated gene trees using the nonhomogeneous GHOST model (Crotty et al. 2020; Supplementary Methods). We assessed branch support for IQ-TREE analyses with 1000 ultra-fast bootstrap (UFBoot) replicates (Minh et al. 2013) and 1000 SH-like approximate likelihood ratio test (SH-aLRT) replicates (Guindon et al. 2010). Values for IQ-TREE and UFBoot metrics range from 0.0 to 1.0 with values close to 1.0 indicating strong branch support (Guindon et al. 2010; Minh et al. 2013).

We used gene trees resulting from IQ-TREE analyses to generate species trees for the Full Exon-Capture, Reduced Locus, Expanded, and 30-Species datasets under the multi-species coalescent in ASTRAL-III v5.6.3 (Mirarab et al. 2014; Zhang et al. 2018; Table 2). We used a single mitochondrial tree that was generated by analyzing concatenated mitochondrial gene alignments as input for ASTRAL-III analyses because mitochondrial genes do not independently segregate. We used ASTRAL-III to generate a distribution of 18 species trees ("18-Tree Distribution"), allowing us to account for topological and divergence time uncertainty in comparative analyses by implementing a gene subset approach used in previous phylogenomic studies (Kolmann et al. 2020; Rincon-Sandoval et al. 2020; Santaquiteria et al. 2021). To accomplish this, we assembled eighteen 65-gene subsets of the Expanded Dataset, ensuring that all species were represented in every subset by including the same 6 genes with high species sampling in every subset. We randomly assigned the remaining 59 genes to each subset. We generated an ultrametric maximum clade credibility species tree from the 18-Tree Distribution in TreeAnnotator v1.8.2 using the "common ancestor heights" option (Rambaut and Drummond 2015). We assessed branch support for ASTRAL-III species trees using local posterior probabilities (LPP; Sayyari and Mirarab 2016). Because ASTRAL-III is sensitive to gene tree estimation error (Zhang et al. 2017, 2018), we conducted a second set of ASTRAL-III analyses after collapsing poorly supported relationships in input gene trees (UFBoot support <33%; Table 2) using TreeCollapserCL (Hodcroft 2021) following Zheng et al. (2017) and Arcila et al. (2021).

Relaxed-Clock Divergence Time Estimation

We time-calibrated species trees generated from ASTRAL-III analyses of the Full Exon-Capture Dataset and the 18-Tree Distribution using Bayesian relaxed-clock methods with the MCMCTree package in PAML v4.9 (Yang 2007; Table 2). MCMCTree computational time is dependent on the number of partitions, making analysis of large datasets with many partitions unfeasible (Dos Reis and Yang 2019). Consequently, we specified 2 partitions for all MCMCtree analyses: (1) first + second and (2) third codon positions. We used 8 priors in divergence time estimation analyses, all with normal distributions and, following the recommendation of Parham et al. (2012), used the youngest age estimates of fossils to set priors:

(1) Root (Otocephala (a.k.a. Ostarioclupeomorpha, Otomorpha))—MRCA: *Thryssa setirostris*, *Chanos chanos*. Used prior from Benton et al. (2015). Hard minimum age: 150.94 Ma. 95% soft maximum age: 228.4 Ma. Prior setting in MCMCTree control file: B(1.509,2.284,1e-300,0.05).

(2) Crown Clupeoidei—MRCA: *Thryssa setirostris*, *Spratelloides gracilis*. We used the crown clupeoid †*Cynoclupea nelsoni* (Malabarba and Di Dario 2017) to set a minimum age of 125 Ma for the most recent common ancestor (MRCA) of crown Clupeoidei and set a conservative soft 95% maximum age of 145 Ma due to the absence of Jurassic clupeoid, clupeiform, or clupeomorph fossils and previous estimates of the timing of Actinopterygian diversification that included broad sampling of actinopterygians and several fossil calibrations (Near et al. 2012a; Broughton et al. 2013; Hughes et al. 2018). This prior is also used in previous studies of clupeiform divergence times (Bloom and Lovejoy 2014; Lavoué et al. 2017a; Bloom and Egan 2018; Egan et al. 2018a). Prior setting in MCMCTree: B(1.250,1.450,1e-300,0.05).

(3) *Dorosoma*—MRCA: *Dorosoma petenense*, *Dorosoma cepedianum*. A *Dorosoma petenense* fossil (Miller 1982) to set a minimum age of 2.5 Ma for the MRCA of *Dorosoma* and set a soft 95% maximum age of 86.3 (Coniacian/Santonian limit) because most crown clupeoid fossils are younger (Grande 1985). This prior is also used in previous studies of clupeiform divergence times (Bloom and Lovejoy 2014; Bloom and Egan 2018; Egan et al. 2018a). Prior setting in MCMCTree: B(0.025,0.863,1e-300,0.05).

(4)–(6) Trans-Isthmus of Panama sister species—a minimum age of 3.0 Ma and soft 95% maximum age of 86.3 Ma for 3 sister pairs of anchovies separated by the Isthmus of Panama (*Cetengraulis edentulus*/*C. mysticetus*, *Anchovia macrolepidota*/*A. clupeoides*, and *Lycengraulis grossidens*/*L. poeyi*). The minimum age for this prior was based upon the estimated timing of closure of the Central American Seaway (Montes et al. 2015) and the maximum was set at 86.3 Ma because most crown clupeoid fossils are younger. This prior is also used in previous studies of clupeiform divergence times (Bloom and Lovejoy 2014; Bloom and Egan 2018; Egan et al. 2018a). Prior setting in MCMCTree: B(0.030,0.863,1e-300,0.05).

(7) Crown Engraulidae—MRCA: *Thryssa setirostris*, *Stolephorus nelsoni*. †*Eoengraulis fasoloi*, sister to the subfamily Engraulinae (Marramà and Carnevale 2016), was used to set a minimum age of 50 Ma for the MRCA of Engraulidae. We set a soft 95% maximum age of 86.3 Ma because most crown clupeoid fossils are younger. This prior is also used in previous studies of clupeiform divergence times (used in Lavoué et al. 2017a; Bloom and Egan 2018; Egan et al. 2018a). Prior setting in MCMCTree: B(0.500,0.863,1e-300,0.05).

(8) Divergence of Engraulidae and Pristigasteridae + Clupeidae—MRCA: *Thryssa setirostris*, *Clupea harengus*. We used the stem engraulid †*Clupeopsis straeleni* (Casier 1946; Capobianco et al. 2020) to set a minimum age

of 54.1 Ma for the divergence of Engraulidae and Pristigasteridae + Clupeidae and the crown clupeoid †*Cynoclupea* to set a soft maximum age of 125 Ma. This calibration has not been used by previous studies. In some of our trees, Spratelloididae was recovered within the Pristigasteridae + Clupeidae clade which prevented us from using this calibration in these cases. Prior setting in MCMCTree: B(0.541,1.25,1e-300,0.05).

Diadromy Character Data

We collected clupeiform diadromy character state data from the primary literature (Beumer 1978; Chubb and Potter 1984; Lal Mohan 1985; Whitehead 1985; Whitehead et al. 1988; Rainboth 1996; Whitfield and Harrison 1996; Blaber et al. 1998; Allen et al. 2002; Strydom et al. 2002; Ezenwaji and Offiah 2003; Hussain et al. 2005; Remya et al. 2019) and from FishBase (www.fishbase.org). To gain insight into the geographic distribution of diadromous clupeiforms, we collected geographic range data for each species from FishBase and scientific articles (primarily Egan et al. (2022)) and used this information to determine the biogeographic regions (*sensu* Lavoué et al. (2013)) occupied by each species. We documented 29 diadromous clupeiform species (2 catadromous, one amphidromous, and 26 anadromous; Supplementary Tables S1). We coded a two-state diadromy character (diadromous or nondiadromous) and a 3-state diadromy character (marine, freshwater, or diadromous; Supplementary Methods). Several clupeiform species occur in the brackish Caspian and Black seas (Whitehead 1985); we coded these species as marine. We coded *Alosa caspia* as diadromous, but there is uncertainty about whether this species migrates to freshwater or merely low-salinity upper reaches of estuaries (Whitehead 1985; Kottelat and Freyhof 2007; M. S. Alavi-Yeganeh personal observation; see Supplementary Methods for more discussion on this topic).

Evolutionary Patterns of Diadromy and Lineage Diversification Rates

We estimated net diversification rates (speciation minus extinction) with BAMM v.2.5.0 (Rabosky 2014). We did not attempt to separately infer extinction rates, because estimates based on phylogenies largely comprised of extant taxa can be unreliable (Rabosky 2010). For analyses, we set the prior expectation for shifts to one, used default Markov Chain Monte Carlo (MCMC) operators, ran MCMC for $2*10^5$ generations, and set additional priors using the *setBAMMpriors* function in the program R v3.4.0 (R Core Team 2017) BAMMtools package (Rabosky et al. 2014). We checked for convergence of MCMC runs by viewing log-likelihood traces and confirmed sufficient effective sample sizes of the log-likelihood and number of rate shifts using the *effectiveSize* function in the program R coda package (Plummer et al. 2006). We accounted for incomplete taxon sampling in BAMM analyses by compiling a clupeiform species list containing 443 species updated with recent clupeiform taxonomic revisions and species descriptions (e.g., Hata et al. 2020, 2021a, 2021b), then calculating "sampling fractions," the proportion of species sampled in clupeiform clades (Supplementary Methods; Supplementary Tables S1–S3). We accounted for topological uncertainty by estimating net diversification rates for each species tree in the 18-Tree Distribution.

BAMM net diversification rate estimates are robust throughout a wide range of parameter space, but under certain conditions, can produce spurious results (Title and Rabosky 2019; Laudanno et al. 2021). Therefore, we also estimated lineage diversification rates with the Lineage-Specific Birth-Death-Shift model (Martínez-Gómez et al. 2024) implemented in the Bayesian program RevBayes v1.2.1 (Höhna et al. 2016). We ran the Lineage-Specific Birth-Death-Shift model by following the tutorial (https://revbayes.github.io/tutorials/divrate/branch_specific.html) and specified 6 rate categories, a global sampling fraction of 0.61, and a Markov chain length of 2500. We compared the output from RevBayes and BAMM to determine if the methods produced concordant results by plotting RevBayes-estimated branch-specific net diversification rates using the RevGadgets program R package (Tribble et al. 2022).

We tested for differences in lineage diversification rates between diadromous and nondiadromous clupeiforms using the two-state diadromy character, 18-Tree Distribution, and 3 methods to determine whether our findings were robust to both differences in methodology and phylogenetic uncertainty. We selected methods with sufficient power to detect correlations between diadromy and diversification rates on moderately-sized phylogenies, such as our clupeiform phylogeny, and did not use methods, such as Structured Rate Permutations on Phylogenies analyses (STRAPP), that are only suitable for larger phylogenies (i.e., >300 species; Beaulieu and O'Meara 2016; Rabosky and Huang 2016; Rabosky and Goldberg 2017). We used the nonparametric Fast, Intuitive State-Dependent Speciation-Extinction analyses (FiSSE) method, which estimates a parameter correlated with, but not identical to speciation rates (Rabosky and Goldberg 2017). We ran FiSSE analyses in program R using code provided in Rabosky and Goldberg (2017) and set each FiSSE analysis to run 1000 bootstrap replicates. We also used the parametric Binary-State Speciation and Extinction (BiSSE; Maddison et al. 2007) and Hidden-State Speciation and Extinction (HiSSE; Beaulieu and O'Meara 2016) methods to test for differences in net diversification rates between diadromous and nondiadromous lineages. We used the *hisse* function in the hisse program R package (Beaulieu and O'Meara 2016) to fit 4 models with varying complexity, along with 2 null models, and specified the proportion of each character state sampled (nondiadromous = 0.63, diadromous = 0.93) to account for

incomplete taxon sampling: 1) a BiSSE model in which diversification rates vary between diadromous and nondiadromous states, 2) a null BiSSE model in which diversification rates are equal among diadromous and nondiadromous states, 3) a HiSSE model with 2 hidden states in which diversification rates vary between diadromous and nondiadromous states, 4) a HiSSE model with a single hidden state associated with diadromy in which diversification rates vary between diadromous and nondiadromous states, 5) a HiSSE model with a single hidden state associated with nondiadromous states in which diversification rates vary between diadromous and nondiadromous states, and 6) a null HiSSE model (CID-2 model) with 2 hidden states in which diversification rates are independent of character states. We determined the best-fitting HiSSE model using Akaike information criterion (AIC), and we considered models to be significantly better than alternatives when the AIC score was at least 2 units lower, following Burnham and Anderson (2002) and Ng and Smith (2018).

We fit models of discrete trait evolution using each tree in the 18-Tree Distribution and both the 2-state and 3-state diadromy characters with the *fitMK* function in the phytools program R package and identified the best-supported model using Akaike Information Criterion (Lewis 2001; Revell 2012). We fit 4 models of discrete trait evolution to the two-state character data: 1) the equal rates (ER) model, 2) the all-rates-different (ARD) model, 3) an irreversible model in which diadromy can evolve to nondiadromy, but not the reverse, and 4) an irreversible model in which nondiadromy can evolve to diadromy, but not the reverse. We fit 6 models of discrete trait evolution to the 3-state character data: 1) the equal rates (ER) model, 2) the all-rates-different (ARD) model, 3) an irreversible model in which diadromy can evolve to marine, but not the reverse, 4) an irreversible model in which diadromy can evolve to freshwater, but not the reverse, 5) an irreversible model in which marine can evolve to diadromy, but not the reverse, and 6) an irreversible model in which freshwater can evolve to diadromy, but not the reverse. We did not test for irreversible models between freshwater and marine character states because previous studies found strong evidence for transitions from both marine to freshwater and freshwater to marine character states (Bloom and Lovejoy 2012, 2014; Bloom and Egan 2018; Egan et al. 2018a). We used the best-fitting model of discrete trait evolution to estimate evolutionary transition rates (q) and locations of diadromy character state changes on our phylogenies. We inferred the phylogenetic positions of diadromy character state changes using the 18-Tree Distribution and the time-calibrated phylogeny generated from ASTRAL-III analysis of the Full Exon-Capture Dataset via Bayesian stochastic character mapping (SIMMAP) with the phytools *make.simmap* function run for 100 iterations on each tree for both the 2-state and 3-state diadromy characters (Bollback 2006; Revell 2012). We plotted the results of the 100 iterations of stochastic character mapping on each phylogeny using the plot and Phytools densityMap functions to determine the consensus locations of character state changes. We did not conduct SIMMAP analyses using a state-dependent diversification model (e.g., Freyman and Höhna 2019) because FiSSE, BiSSE, and HiSSE analyses did not identify a consistent association between diadromy and lineage diversification rates (see Results section).

Results

Clupeiform Phylogenetic Relationships and Divergence Times

Our various Full Exon-Capture and Reduced Locus analyses generated largely congruent phylogenetic hypotheses for Clupeiformes. Additionally, the 30-Species Dataset with and without implementation of the GHOST model inferred the same relationships among major clupeiform lineages. We documented little interspecific variation in base-composition (GC content mean = 58.03%, standard deviation = 0.01%; Supplementary Tables S1 and S2) and MSC analyses of nucleotide and amino acid alignments produced very similar topologies. Relationships among Dussumieriidae, *Etrumeus,* Clupeidae, and Pristigaseridae were the only instances of disagreement between analyses involving relationships among major lineages (Fig. 1). Among MSC analyses, this disagreement was eliminated when we collapsed poorly-supported relationships in input gene trees. Analyses of the Expanded Dataset generated phylogenies similar to those produced by the Full Exon-Capture and Reduced Locus analyses. Therefore, here we only report results of Expanded Dataset analyses involving Chirocentridae (see Supplementary Figs. S2–S19 and Supplementary tree files for full results from Expanded Dataset analyses).

We resolved, with strong support, relationships among most early-diverging clupeiform lineages. We confirmed the sister-group relationships between *Denticeps clupeoides* and all other crown clupeiforms (Clupeoidei) and Spratelloididae and all remaining clupeoids (Figs. 1 and 2). All analyses supported Engraulidae as sister to a clade containing 2 lineages (LPP values = 0.91–0.97 and UFBoot and SH-aLRT values = 1.00): 1) a clade containing Pristigasteridae, Clupeidae, Dussumieriidae, and *Etrumeus* (incertae cedis Clupeidae; Wang et al. 2022) and 2) a clade containing Dorosomatidae, Ehiravidae, and Alosidae. Analyses of the complete Expanded Dataset recovered Chirocentridae sister to Dorosomatidae + Alosidae, but with little support (nucleotide analysis LPP = 0.41). Analyses of the 65-gene subsets variously recovered Chirocentridae sister to Dussumieriidae, Dussumieriidae + Clupeidae, and Clupeidae. We recovered Clupeidae sister to *Etrumeus*, but did not resolve relationships among Clupeidae + *Etrumeus,*

Pristigasteridae, and Dussumieriidae with strong support. Nucleotide-based MSC analyses in which poorly-supported gene trees were not collapsed and concatenation-based analyses recovered different relationships among these lineages than all other analyses (Fig. 1). Every other analysis recovered the relationships among these taxa shown in Fig. 2, with Pristigasteridae sister to 2 clades: Clupeidae + *Etrumeus* and Dussumieriidae. These results indicate that inferences among these clades were sensitive to methodology. All analyses resolved Ehiravidae (except *Clupeoides venulosus* and *Clupeoides papuensis*; see below) as sister to Alosidae + Dorosomatidae with strong support (LPP values = 1.00 and UFBoot and SH-aLRT values = 1.00; Fig. 1).

All MSC and concatenation-based analyses resolved the 10 clupeiform families proposed by Wang et al. (2022) as monophyletic with strong support, with a single exception (LPP values = 1.00, UFBoot and SH-aLRT values = 1.00; Figs. 1 and 2; we report all branch support values in Supplementary tree files): Two members of Ehiravidae, *Clupeoides venulosus* and *Clupeoides papuensis*, were placed in Dorosomatidae, rendering Ehiravidae polyphyletic and Dorosomatidae paraphyletic (LPP values = 1.00 and UFBoot and SH-aLRT values = 1.00). Our results support recognizing *Etrumeus* as a member of Clupeidae (assigned incertae cedis Clupeidae by Wang et al. (2022); Figs. 1 and 2). We identified several polyphyletic and paraphyletic genera in addition to *Clupeoides*: *Sprattus, Pellona, Ilisha, Herklotsichthys, Sardinella, Thrissina, Engraulis, Anchoa,* and *Anchoviella* (Fig. 2).

MCMCTree estimated a late Triassic/early Jurassic origin of crown Clupeiformes (Full Exon Dataset ~199 Ma; 18-Tree Distribution ~213 Ma) and a Cretaceous origin for crown Clupeoidei (Full Exon-Capture Dataset ~139 Ma; 18-Tree Distribution ~ 137 Ma; Fig. 2; Supplementary Fig. S1). The majority of remaining major clupeiform clades originated during the late Cretaceous (e.g., Spratelloididae, Engraulidae, Ehiravidae, Clupeidae, and Dorosomatidae) and early to middle Cenozoic (e.g., Pristigasteridae, and Alosidae; Fig. 2; Supplementary Fig. S1). Divergence time estimates were similar among phylogenies within the 18-Tree Distribution and were similar to divergence times estimated using the Full Exon-Capture Dataset (Fig. 2; Supplementary Fig. S1).

Evolutionary Patterns of Diadromy and Diversification Rate/Diadromy Associations

We documented the greatest numbers of diadromous species in the Indo-West Pacific and Ponto-Caspian biogeographic regions and the lowest numbers in the East Pacific and New Zealand biogeographic regions (Fig. 3). The ARD model was the best-fit for all trees in the 18-Tree Distribution when using the two-state diadromy character (Supplementary Table S4). The ARD model was the best-fit for 13 of 18 trees when using the 3-state diadromy character. An irreversible model, in which diadromy could not evolve to the marine state, was the best fit for the remaining 5 trees (Supplementary Table S5). *FitMK* analyses estimated asymmetric transition rates between diadromy character states for both the 2-state and 3-state diadromy characters. For the two-state diadromy character, *FitMK* analyses inferred higher rates out of diadromy than to diadromy (Fig 4a). For the 3-state diadromy character, *FitMK* analyses inferred that the highest transition rates were from diadromous and freshwater character states and that the and slowest transition rates were from marine to freshwater character states (Fig. 4b; Supplementary Tables S4 and S5). SIMMAP analyses inferred a mean of 13 origins of diadromy (10 origins of anadromy, 2 origins of catadromy, and one origin of amphidromy with 4 transitions from freshwater to diadromous character states and 9 transitions from marine to diadromous character states), a mean of 7 losses of diadromy (6 transitions from diadromy to freshwater and one transition from diadromy to marine), a mean of 17 transitions from marine to freshwater, and a mean of 4 transitions from freshwater to marine character states (Supplementary Table S6; Supplementary Figs. S2–S38). We found that all diadromous clupeiform lineages originated during the Cenozoic Era, with 5 origins during the middle to late Paleogene Period, 7 during the Neogene Period, and one during the Quaternary Period (Fig. 2; Supplementary Fig. S39). Because transitions between diadromy and nondiadromy occurred relatively recently, uncertainty about the positions of the early-diverging lineages Chirocentridae, Clupeidae, Dussumieriidae, and Pristigasteridae had little impact on the results of our phylogenetic comparative analyses.

For brevity, we only report results of BAMM-estimated net diversification rates in the main text because BAMM and the Lineage-Specific Birth-Death-Shift model yielded concordant results (Supplementary Fig. S40). We detected relatively uniform diversification rates across most clupeiforms, but found elevated net diversification rates within 2 lineages: 1) a largely tropical, marine clade of anchovies primarily distributed in the Gulf of Mexico, Caribbean Sea, and the northern Atlantic coasts of South America and 2) *Brevoortia* + *Alosa*, a clade containing a transition to diadromy (Supplementary Figs. S40 and S41). The distribution of BAMM-estimated tip-specific net diversification rates did not suggest a categorical association between diadromy and lineage diversification rates (Fig. 3). FiSSE analyses found no associations between diadromy and lineage diversification rates (P-value = 0.867 ± 0.105; Supporting Table S7). In contrast, HiSSE analyses found that diadromous fishes exhibited significantly faster net diversification rates than nondiadromous fishes, a result driven by the rapid diversification rates of diadromous *Alosa,* with AIC preferring a state-dependent HiSSE model for 17/18 species trees (Supporting Table S8). For 15/17 cases in which AIC selected a HiSSE state-dependent model, a

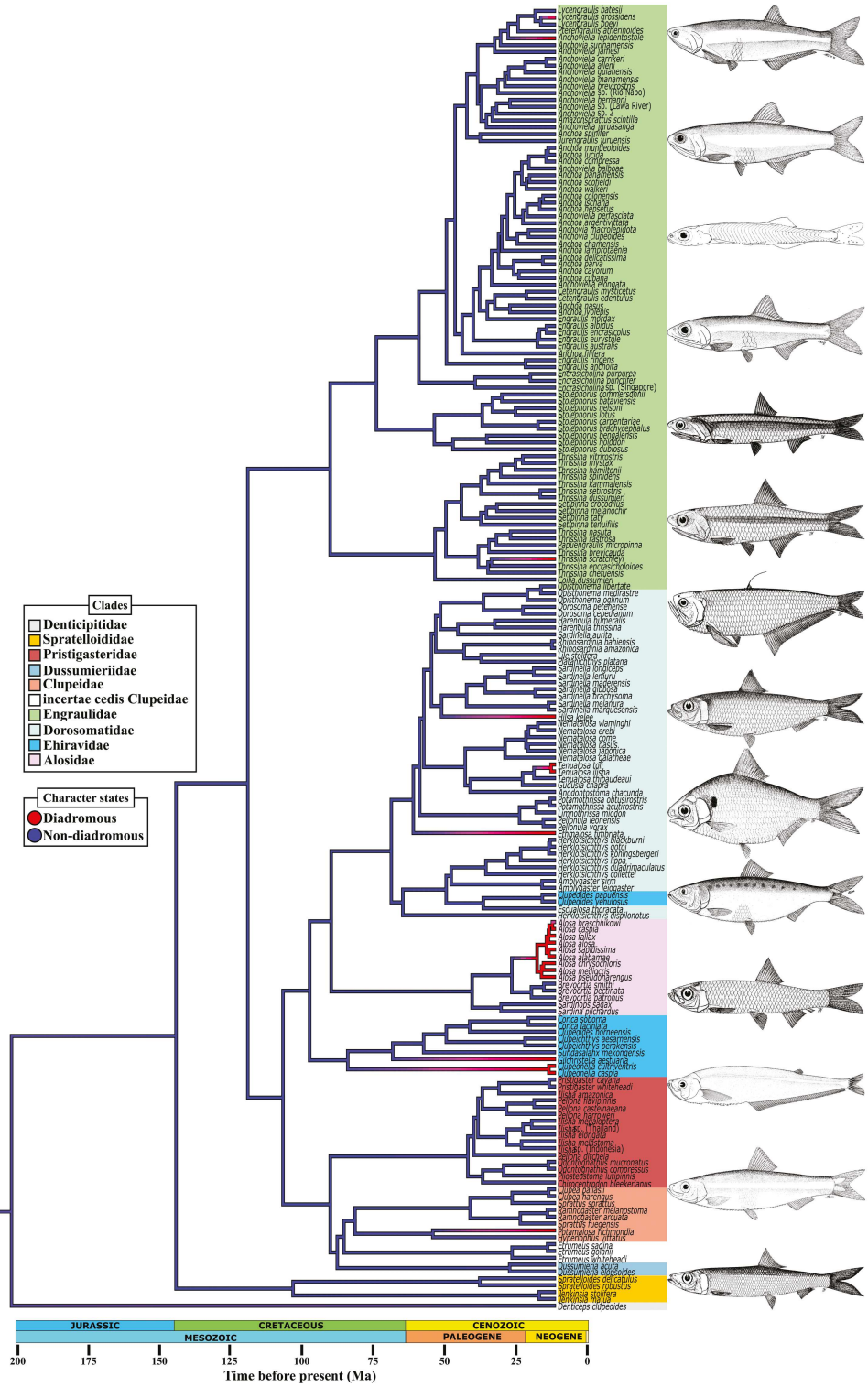

FIGURE 2. Time-calibrated phylogeny generated from ASTRAL-III analysis of the Full Exon-Capture dataset (input gene tress with poorly supported relationships were not collapsed for this analysis; outgroup taxa not shown). Shading behind species names along the phylogeny tips demarcates currently recognized families. The posterior densities of character states inferred by 100 SIMMAP iterations using the two-state diadromy character and ARD model of discrete character evolution are shown along branches. All diadromous clupeiform lineages are anadromous with the exception of 2 catadromous lineages (*Potamalosa richmondia* and *Ethmalosa fimbriata*) and one amphidromous lineage (*Gilchristella aestuaria*). Line drawings depict representative clupeiforms (listed from top to bottom): *Pterengraulis atherinoides*, *Anchoviella lepidentostole*, *Amazonsprattus scintilla*, *Anchoa hepsetus*, *Engraulis encrasicolus*, *Stolephorus nelsoni*, *Papuengraulis micropinna*, *Sardinella gibbosa*, *Anodontostoma chacunda*, *Alosa fallax*, *Corica laciniata*, *Odontognathus mucronatus*, *Hyperlophus vittatus*, and *Dussumieria elopsoides*. Illustrations modified and reproduced from Whitehead (1985) and Whitehead et al. (1988) with permission. See Table 1 for a list of all species included in Fig. 2, along with diadromy character state and biogeographic data, shown in the same order that they appear along the tips of this phylogeny, from top to bottom.

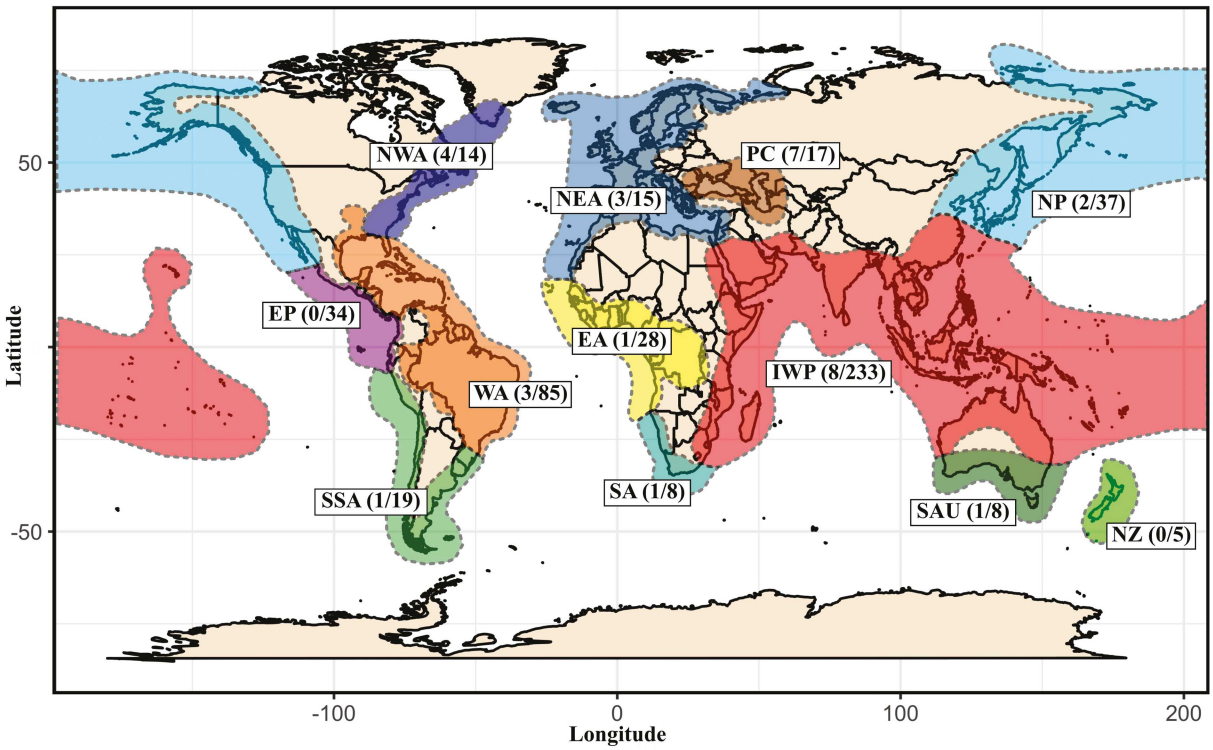

FIGURE 3. Number of diadromous clupeiform species and total clupeiform species richness (diadromous species/total species richness) in major biogeographic regions (*sensu* Lavoué et al. (2013)): Northwest Atlantic = NWA, Northeast Atlantic = NEA, West Atlantic = WA, Ponto-Caspian = PC, Indo-West Pacific = IWP, North Pacific = MP, East Pacific = EP, South South America = SSA, East Atlantic = EA, South Africa = SA, South Australia = SAU, and New Zealand = NZ. Several species occur in more than one biogeographic region. Map projection: geographical.

model with a hidden state associated with diadromy was the best fit. This indicates that diadromy and additional, unmeasured factors may have led to elevated net diversification rates (Supporting Table S8).

DISCUSSION

Biogeography and Migration Distance May Govern Relationships Between Diadromy and Diversification Rates

We found that the largest lineage diversification rate increase in clupeiforms was associated with a transition to diadromy in the most recent common ancestor of temperate shads (*Alosa* spp.), but uncovered little statistical support for categorically faster lineage diversification rates in diadromous fishes. Lineage diversification rates overlapped substantially among freshwater, marine, and diadromous lineages and the nonparametric (more conservative) FiSSE analyses did not identify statistically significant differences in diversification rates between diadromous and nondiadromous clupeiforms. In contrast, the parametric (more powerful) HiSSE analyses identified significantly faster net diversification rates associated with diadromy. The AIC nearly always selected a HiSSE model containing a hidden state associated with diadromy, indicating that diadromy in combination with additional, unmeasured factors may have led to increased net diversification rates in clupeiforms. HiSSE can report statistically significant correlations without multiple, phylogenetically independent associations between traits and rates (Maddison and FitzJohn 2014; Rabosky and Goldberg 2015, 2017; Harvey and Rabosky 2018). This appeared to be the case in clupeiforms, with BAMM and Lineage-Specific Birth-Death-Shift model analyses indicating that the relationship between elevated diversification rates and diadromy was driven by a single clade: *Alosa*. *Alosa* contained one transition to diadromy and exhibited net diversification rates > 4x faster than all other clupeiform lineages. The remaining 12 diadromous clupeiform clades did not exhibit accelerated diversification rates.

Previous research found elevated lineage diversification rates in some, but not all diadromous clades, a result consistent with our study. This suggests that diadromy only facilitates rapid diversification in specific ecological and paleogeographic scenarios (Corush 2019; Burridge and Waters 2020; Thacker et al. 2021). We hypothesize that accelerated diversification rates in *Alosa* stem from the combination of long-distance migration and biogeographic context. Diadromous species that migrate long distances and have large range sizes may have more opportunities to colonize novel

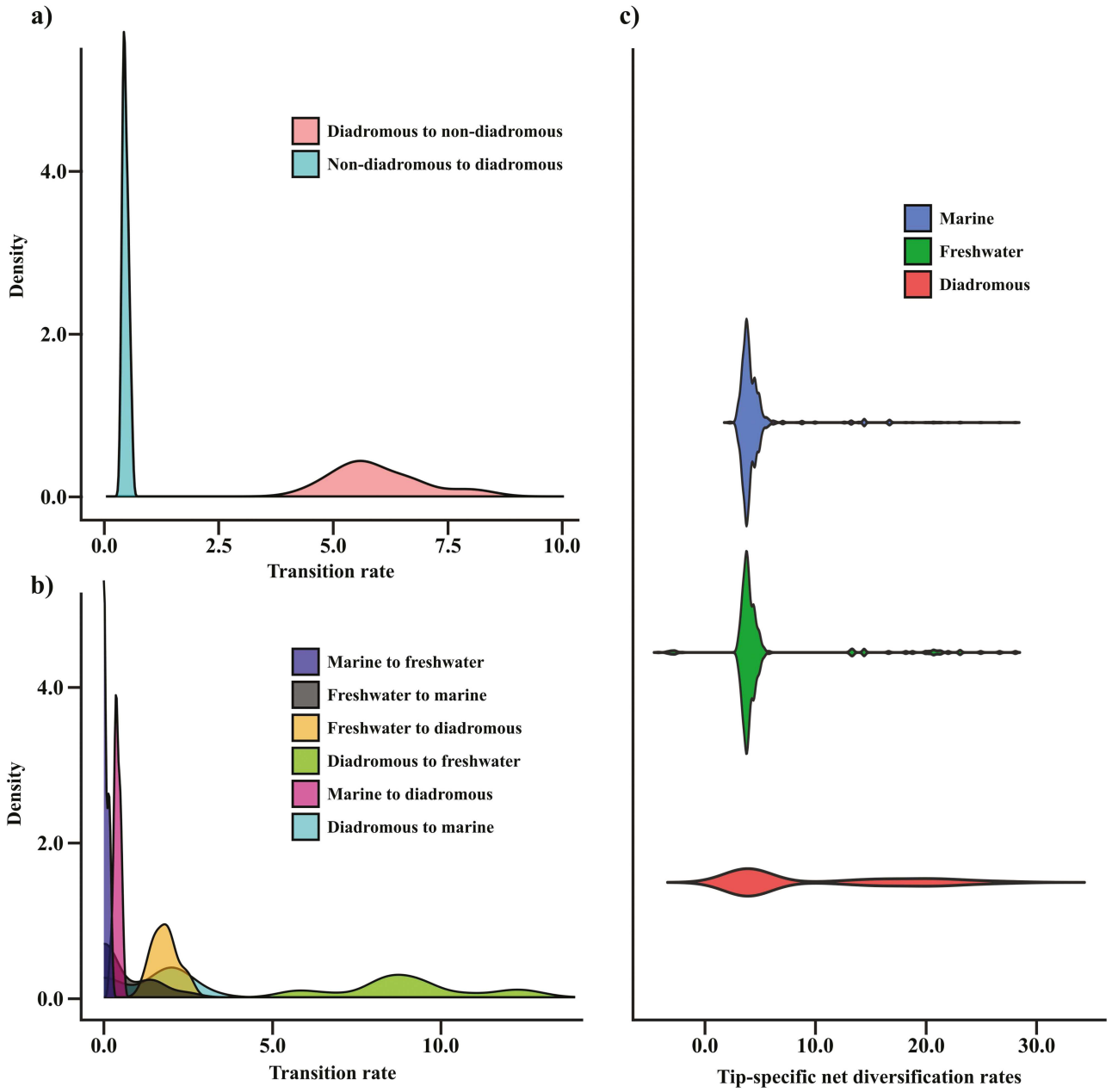

FIGURE 4. a) Probability density plots of transition rates between character states estimated using the two-state diadromy character, fitMK phytools function, each tree in the 18-Tree Distribution, and the best-supported model of discrete character evolution for each tree. b) Probability density plots of transition rates between character states estimated using the 3-state diadromy character, fitMK phytools function, each tree in the 18-Tree Distribution, and the best-supported model of discrete character evolution for each tree. c) Probability density violin plots of BAMM-estimated tip-specific net diversification rates for diadromy character states. Probability densities were generated using the Expanded Dataset posterior distribution of 18 time-calibrated phylogenies. Thus, this figure summarizes 18 tip-specific net diversification rate estimates for each of the 272 species in the Expanded Dataset, for a total of 4896 tip-specific net diversification rate estimates.

environments than nondiadromous species or diadromous species that migrate short distances (McDowall 1988, 2001; Griffiths 2006). *Alosa* contains some of the longest clupeiform migrations. For example, *Alosa sapidissima*, *Alosa pontica*, and *Alosa fallax* have been reported to ascend rivers up to ~ 1100, 600, and 400 km, respectively (Whitehead et al. 1988; Manyukas 1989; Limburg et al. 2003). In contrast, diadromous clupeiforms such as *Lycengraulis grossidens*, *Anchoviella lepidentostole*, and *Gilchristella aestuaria*, migrate considerably shorter distances (typically <100 km; Whitehead 1985; Strydom et al. 2002; Mendonça and Sobrinhob 2013; Mai et al. 2014) and did not experience increased diversification rates. Furthermore, because many diadromous species use multiple habitats, particularly species that migrate long distances, they are predicted to evolve

generalist ecological niches (Futuyma and Moreno 1988; Van Tienderen 1991). Generalism is thought to impart species with an enhanced capacity to colonize new areas because they can cope with a wide range of environmental conditions (Futuyma and Moreno 1988). However, increases in net diversification rates following colonization of new areas by diadromous species may only occur when there is sufficient ecological opportunity (McDowall 2001; Betancur-R et al. 2012; Bloom and Lovejoy 2014; Bloom and Egan 2018; Burridge and Waters 2020). For example, colonization of environments with few competitors, such as recently deglaciated areas, might provide ideal conditions for diadromous fishes to rapidly diversify (Griffiths 2006; Betancur-R. et al. 2012; Bloom and Egan 2018). This combination of ecological opportunity stemming from biogeographic context and phenotypes that facilitate colonization of new areas may be a general feature of rapidly diversifying diadromous clades such as *Alosa* and Salmonidae (Whitehead 1985; Alexandrou et al. 2013; Macqueen and Johnston 2014). Diadromous *Alosa* and Salmonidae are largely found in north temperate rivers, with a small number of lineages occurring in subtropical regions (Whitehead 1985; Nelson et al. 2016). Both clades diversified within the past 20 Ma, especially after the Miocene Climatic Optimum (<14.5 Ma). The diversification of both clades may have been driven by ecological opportunity in temperate and subtropical aquatic ecosystems due to the recession of tropical climates, expansion of temperate climates, and glacial cycles that occurred during this period and produced temperate aquatic habitats in which early colonists, such as *Alosa* and Salmonidae, faced little competition from incumbent species (Raymo and Ruddiman 1992; Zachos et al. 2001; Near et al. 2012b; April et al. 2013). Although Corush (2019) identified statistically significant increases in speciation rates associated with diadromy using a large teleost dataset, this pattern may have been driven by a subset of diadromous clades, such as *Alosa* and Salmonidae. This would indicate that relationships between diadromy and net diversification rates are context dependent and explain why some studies have not found increases in lineage diversification rates associated with diadromy (Burridge and Waters 2020; Thacker et al. 2021).

There are only ~450 species of diadromous fishes, despite this life history strategy spanning the tree of fishes, from lampreys to gobies. Given this remarkable amount of evolutionary time, why are there so few diadromous fishes? We found that 12 of the 13 diadromous clupeiform clades did not exhibit accelerated lineage diversification rates, which may be due to evolutionary constraints (i.e., phenotypes poorly suited for colonizing or diversifying in novel ecosystems), lack of ecological opportunity, or high rates of gene flow among populations. Our results are consistent with a recent study on Galaxiidae, which also found no correlations between diadromy and diversification rates (Burridge and Waters 2020). Several tropical and subtropical lineages of diadromous clupeiforms, such as Indo-Pacific *Tenualosa* spp. and *Thrissina scratchleyi* and South American *Anchoviella lepidentostole* and *Lycengraulis grossidens* occur in ecosystems with high species richness, such as the Ganges and Amazon River systems (Whitehead 1985; Whitehead et al. 1988). High incumbent species diversity may have resulted in intense competition for resources, subverting increased potential for speciation conferred by diadromy (Betancur-R et al. 2012; Bloom and Egan 2018). This may be particularly relevant for species such as the Indo-West Pacific *Tenualosa* spp. and *Hilsa kelee.* Like some *Alosa* spp., these species can migrate hundreds of kilometers in freshwater, a trait that may promote speciation in certain biogeographic contexts (Whitehead 1985; Bhaumik 2013). Diadromous clupeiform species such as *Lycengraulis grossidens*, *Anchoviella lepidentostole*, and *Gilchristella aestuaria* migrate shorter distances, which may have limited their ability to colonize new areas and exposure to ecological opportunities (Whitehead 1985; Strydom et al. 2002; Mendonça and Sobrinhob 2013; Mai et al. 2014). Although migration may create opportunity for colonizing new areas, many migratory lineages also have high rates of gene flow among populations (McDowall 1999; Goodman et al. 2008; Hasselman et al. 2013), which can constrain local adaptation and decrease the probability of allopatric speciation. Our results contribute to an emerging pattern suggesting that diadromy may increase a clade's potential for diversification, but that relationships between diadromy and lineage diversification rates are modulated by additional factors (Beaulieu and O'Meara 2016; Burridge and Waters 2020). Additional research is needed to further characterize how often diadromy leads to increased lineage diversification rates and the ecological and biogeographic contexts in which this occurs.

Cenozoic Origins of Diadromy and the Frequency and Rate of Character State Transitions

Our expansive taxonomic sampling revealed 13 origins of diadromy (10 origins of anadromy, 2 origins of catadromy (*Potamalosa richmondia* and *Ethmalosa fimbriata*), and one origin of amphidromy (*Gilchristella aestuaria*)) and 7 losses of diadromy in clupeiforms, providing additional insight regarding the evolutionary patterns of this life history strategy (Bloom and Lovejoy 2014; Bloom et al. 2018; Egan et al. 2018a; Corush 2019). Our study sampled 26 of 29 diadromous species. We were unable to include the diadromous species *Alosa kessleri*, *Alosa tanaica*, and *Alosa volgensis* in phylogenetic comparative analyses. However, because these species are nested in the diadromous clade *Alosa*, it is unlikely that there were additional transitions to diadromy within clupeiforms that were not detected by this study. We did not find support for models of irreversible evolution of diadromy (i.e., once evolved, never lost) and we found 7 transitions from diadromy to either marine or freshwater states. This suggests that,

contrary to previous claims (Feutry et al. 2013; Bloom and Lovejoy 2014), diadromy is not necessarily an evolutionary dead-end.

We found that there were a greater number of transitions from nondiadromy to diadromy than from diadromy to nondiadromy. However, like Corush (2019), we inferred that evolutionary transition rates out of diadromy were higher than transition rates between all other character states. In other words, during the ~200 myr history of crown Clupeiformes, diadromy evolved twice the number of times it was lost. Although, because all extant diadromous lineages evolved within the past 50 myr, and 2 of these clades (*Alosa* and *Tenualosa*) relatively quickly gave rise to nondiadromous lineages, transition rates from diadromy to nondiadromy character states were higher than transition rates from nondiadromy to diadromy. Across fishes, most nondiadromous lineages derived from diadromous ancestors have low species diversity and limited distributions on the fringes of the diadromous ancestral range, a pattern also exhibited by Clupeifomes that points to vicariant speciation as a major driver of diadromous to nondiadromous transitions (McDowall 1988; Kottelat and Freyhof 2007). For example, we found that in <5 myr there were 5 transitions to nondiadromy in *Alosa* that include the monotypic lineages *Alosa vistonica*, restricted to Lake Vistonis, Greece, and *Alosa macedonica*, found only in Lake Volvi, Greece (Whitehead 1985; Kottelat and Freyhof 2007). In contrast, only 2 of 13 diadromous clupeiform lineages gave rise to nondiadromous species (*Alosa* and *Tenualosa*), indicating that in most lineages, once diadromy evolves, there are rarely reversals to nondiadromy (Inoue et al. 2010; Alexandrou et al. 2013; Corush 2019; Burridge and Waters 2020).

We inferred that all diadromous clupeiform lineages originated during the Cenozoic Era, with 5 origins during the middle to late Paleogene Period, 7 during the Neogene Period, and one during the Quaternary Period. Diadromy appears to have also originated in several other teleost lineages during the Cenozoic Era, including Salmonidae (Alexandrou et al. 2013; Macqueen and Johnston 2014), Galaxiidae (Burridge and Waters 2020), and *Takifugu* (Yang and Chen 2008; Yamanoue et al. 2009). It is possible the Cretaceous-Paleogene (K-Pg) extinction event and expansion of temperate environments following the Early Eocene Climatic Optimum created ecological opportunities that fostered the proliferation of diadromous lineages (Raymo and Ruddiman 1992; Zachos et al. 2001; Near et al. 2012b; April et al. 2013; Alfaro et al. 2018). Several diadromous clupeiform lineages have persisted for millions of years (e.g., ~48, 33, 26, and 25 myr in *Clupeonella*, *Ethmalosa*, *Hilsa*, and *Potamalosa*, respectively), providing further evidence that, although diadromous lineages can give rise to nondiadromous lineages, diadromy is not usually a short-lived transient state between fully marine and fully freshwater life histories (Corush 2019).

Clupeiform Phylogenetic Relationships

Our phylogenomic analyses included the most comprehensive clupeiform taxon sampling to date and clarified relationships among major clades, revealing taxa requiring revision and identifying recalcitrant relationships needing further investigation. The limited interspecific variation in the GC content among Clupeiformes and widespread agreement among analyses regarding topology, including between analyses with and without the GHOST model and nucleotide- and amino acid-based analyses, indicates that our major findings are generally robust to methodology, missing data, base-composition heterogeneity, and heterotachy (Figs. 1 and 2; Supplementary tree files). However, minor disagreements between MSC analyses in which input gene trees with poorly-supported relationships were collapsed and were not collapsed indicate gene tree estimation error impacted some of our findings.

Our study largely supports the revised family-level classification proposed by Wang et al. (2022). Due to limited taxonomic sampling, Wang et al. (2022) designated *Etrumeus* as Clupeidae incertae cedis. Our study supports the inclusion of *Etrumeus* as a member of Clupeidae. In agreement with Wang et al. (2022), we recovered Chirocentridae sister to Dussumieriidae in several trees generated from analysis of 65-gene subsets of the Expanded Dataset, albeit with low support. Like many previous molecular studies, we were unable to resolve relationships among Dussumieriidae, Pristigasteridae, and Clupeidae (including *Etrumeus*) with strong support (Lavoué et al. 2013; Bloom and Lovejoy 2014; Egan et al. 2018a; Wang et al. 2022). However, most of our analyses recovered Pristigasteridae sister to Dussumieriidae + Clupeidae with low statistical support. In contrast, Wang et al. (2022) resolved Pristigasteridae sister to Dussumieriidae + Chirocentri edae and Clupeidae sister to all 3 of these clades with strong support. Although Wang et al. (2022) used many loci (1165) for phylogenetic analyses, they included few species and used concatenation-based, but not MSC approaches. Concatenation-based approaches can infer erroneous relationships with strong support because these methods do not account for gene tree discordance (Degnan and Rosenberg 2009). For this reason, and because our MSC analyses failed to recover the same relationships among Pristigasteridae, Dussumieriidae, and Clupeidae as Wang et al. (2022), the phylogenetic positions of these lineages remain unresolved. Our results do not suggest that the lingering uncertainty about the relationships among these taxa can be attributed to base-composition heterogeneity or a lack of sequence data in alignments (see Supplementary Materials for additional discussion about the impact of missing sequence data on phylogenetic analyses). Lack of resolution among Pristigasteridae, Clupeidae, and Dussumieriidae, even in analyses with poorly supported branches in input gene trees collapsed, suggests that gene tree conflict and low information content in individual loci may contribute to uncertainty regarding

these relationships (McCormack et al. 2013; Zhang et al. 2017). Determining the phylogenetic positions of these taxa may remain difficult because the relatively short time span during which these lineages diverged may have produced high levels of incomplete lineage sorting and low signal-to-noise ratios, hindering phylogenetic inference (Degnan and Rosenberg 2009). See Supplementary Materials for additional discussion about phylogenetic relationships.

Clupeiform divergence times

We inferred a Late Triassic/Early Jurassic origin of crown Clupeiformes, a Cretaceous origin for crown Clupeoidei, and Late Cretaceous to middle Cenozoic for the majority of the remaining major clupeiform clades. Our divergence time estimates were similar to previous studies, with a few notable differences (Lavoué et al. 2013; Bloom and Lovejoy 2014; Bloom and Egan 2018; Egan et al. 2018a; Wang et al. 2022). Our Late Triassic/Early Jurassic (199 Ma) estimate for the origin of crown clupeiforms is older than previous studies focusing on Clupeiformes (Lavoué et al. 2013; Bloom and Lovejoy 2014; Bloom and Egan 2018; Egan et al. 2018a; Wang et al. 2022) and studies spanning Teleostei that included several fossil calibrations outside Clupeiformes (Near et al. (2012a) = 175 Ma, Betancur-R. et al. (2013) = 189 Ma, and Rabosky et al. (2018) = 154 Ma). We estimated earlier origins (~10–20 ma) for a few clades, including Engraulidae and Spratelloididae. Potential explanations for the differences between our study and previous studies are our use of novel node calibrations, exclusion of fossils used for past calibrations that have been determined to be unsuitable (e.g., †*Nolfia riachuelensis* and †*Gasteroclupea branisai*; De Figueiredo 2009; Marramà and Carnevale 2017), expanded taxonomic sampling of Clupeiformes, and increased phylogenetic resolution.

Supplementary Material

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.47d7wm3h8.

Acknowledgments

We thank P.N. Kaeoprakan and C.C. Phetkong (Rajamangala University), G. Dally and S. Horner (MAGNT), Dr. M. Sheaves (James Cook University), U.-S. Chew and Dr. C.-H. Kuo (National Chiayi University, Taiwan), Dr. B. Bowen, M. Hoban, and Dr. K. Hurley (University of Hawai'i at Mānoa), Dr. Kim Peyton (State of Hawaii Department of Land and Natural Resources), and Dr. A. Schreier (University of California, Davis) for assisting with fieldwork and logistics. We thank Dr. S. Lavoué (School of Biological Sciences, Universiti Sains Malaysia), Dr. M. Stiassny and Dr. B Brown (American Museum of Natural History), V. Magath (University of Hamburg Zoological Museum), Dr. K.-T. Shao (Academia Sinica Biodiversity Research Museum), Dr. L. Page (Florida Museum of Natural History), A. Storey (Wetland Research & Management), K. Hunnam (Charles Darwin University), A. Hay (Australian Museum), A. Graham and W. White (CSIRO), G. Moore (Western Australian Museum) and M. Adams (South Australian Museum) for providing tissue samples. We thank C. Pedraza, Dr. M. Rincon-Sandoval, A. Santaquiteria, and Dr. L. Hughes for help with bioinformatics. Computing for this project was performed at the University of Oklahoma Supercomputing Center for Education & Research (OSCER). OSCER Director H. Neeman and OSCER Senior System Administrators J. Speckman and H. Severini provided valuable technical expertise.

Funding

This work was supported by through grants and followships to JPE from the Lerner-Gray Memorial Fund for Marine Research (American Museum of Natural History), Dayton Research Fund (JFBM, University of Minnesota), the National Science Foundation, USA (No. 00039202), and the East Asia and Pacific Summer Institutes Program sponsored by the National Science Council of Taiwan, ROC and National Science Foundation, USA (No. 1316912). AMS received support from the Minnesota Agricultural Experiment Station. DDB, RBR, and DA received funding from the National Science Foundation, USA (DDB: No. DEB-1754627, RBR: Nos. DEB-1932759 and DEB-2225130, and DA: Nos. DEB-2015404 and DEB-2144325).

Data Availability

Raw sequence data are available for download from the NCBI Sequence Read Archive (SRA; BioProject PRJNA1076878).

References

Alerstam T., Hedenström A., Åkesson S. 2003. Long-distance migration: evolution and determinants. Oikos 103:247–260.

Alexandrou M.A., Swartz B.A., Matzke N.J., Oakley T.H. 2013. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Mol. Phylogenet. Evol. 69:514–523.

Alfaro M.E., Faircloth B.C., Harrington R.C., Sorenson L., Friedman M., Thacker C.E., Oliveros C.H., Černý D., Near T.J. 2018. Explosive diversification of marine fishes at the Cretaceous–Palaeogene boundary. Nat. Ecol. Evol. 2:688–696.

Allen G.R., Midgley S.H., Allen M. 2002. Field guide to the freshwater fishes of Australia. Perth, Western Australia: Western Australian Museum. p. 394.

Alò D., Lacy S.N., Castillo A., Samaniego H.A., Marquet P.A. 2021. The macroecology of fish migration. Glob. Ecol. Biogeogr. 30:99–116.

April J., Hanner R.H., Dion-Côté A.M., Bernatchez L. 2013. Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. Mol. Ecol. 22:409–422.
Arcila D., Hughes L.C., Meléndez-Vazquez B., Baldwin C.C., White W.T., Carpenter K.E., Williams J.T., Santos M.D., Pogonoski J.J., Miya M., Ortí G. 2021. Testing the utility of alternative metrics of branch support to address the ancient evolutionary radiation of tunas, stromateoids, and allies (Teleostei: Pelagiaria). Syst. Biol. 70(6):1123–1144.
Arnette S.D., Simonitis L.E., Egan J.P., Cohen K.E., Kolmann M.A. 2024. True grit? Comparative anatomy and evolution of gizzards in fishes. J. Anat. 244:260–273.
Beaulieu J.M., O'Meara B.C. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.
Benton, M.J., P.C. Donoghue, R.J. Asher, M. Friedman, T.J. Near, and J. Vinther. 2015. Constraints on the timescale of animal evolutionary history. Palaeontol Electron. 18:1-106.
Betancur-R R., Ortí G., Stein A.M., Marceniuk A.P., Pyron R.A. 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., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton Ii J.C., Zhang F., Buser T., Campbell M.A., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Ortí G. 2013. The tree of life and a new classification of bony fishes. PLoS Curr. 5:1–45.
Beumer J.P. 1978. Feeding ecology of four fishes from a mangrove creek in north Queensland, Australia. J. Fish Biol. 12:475–490.
Bhaumik, U. 2013. Decadal studies on Hilsa and its fishery in India-a review. J Interacad 17:377-405.
Bisson I.A., Safi K., Holland R.A. 2009. Evidence for repeated independent evolution of migration in the largest family of bats. PLoS One 4:e7504–e7506.
Blaber S.J.M., Staunton-Smith J., Milton D.A., Fry G., Velde T. Van der, Pang J., Wong P., Boon-Teck O. 1998. The Biology and Life-history Strategies of *Ilisha* (Teleostei: Pristigasteridae) in the Coastal Waters and Estuaries of Sarawak. Estuar. Coast. Shelf Sci. 47:499–511.
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., Egan J.P. 2018. Systematics of Clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. Neotrop Ichthyol. 16:e180095.
Bloom D.D., Burns M.D., Schriever T.A. 2018. Evolution of body size and trophic position in migratory fishes: a phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). Biol. J. Linn. Soc. 125:302–314.
Bollback, J.P. 2006. Stochastic character mapping of discrete traits on phylogenies. BMC Bioinform 7:88.
Bowlin M.S., Bisson I.A., Shamoun-Baranes J., Reichard J.D., Sapir N., Marra P.P., Kunz T.H., Wilcove D.S., Hedenström A., Guglielmo C.G., Åkesson S., Ramenofsky M., Wikelski M. 2010. Grand challenges in migration biology. Integr. Comp. Biol. 50:261–279.
Broughton R.E., Betancur-R R., Li C., Arratia G., Ortí G. 2013. Multilocus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 5.
Burnham, K.P., and D.R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd edn. New York, NY: Springer.
Burns M.D., Bloom D.D. 2020. Migratory lineages rapidly evolve larger body sizes than non-migratory relatives in ray-finned fishes. Proc. Biol. Sci. 287:20192615.
Burridge C.P., Waters J.M. 2020. Does migration promote or inhibit diversification? A case study involving the dominant radiation of temperate Southern Hemisphere freshwater fishes. Evolution. 74:1954–1965.
Capobianco A., Beckett H.T., Steurbaut E., Gingerich P.D., Carnevale G., Friedman M. 2020. Large-bodied sabre-toothed anchovies reveal unanticipated ecological diversity in early Palaeogene teleosts. R. Soc. Open Sci. 7:192260.
Casier, E. 1946. La faune ichthyologique de l'Yprésien de la Belgique. Mém Mus R Hist Natl Belg. 104:1-267.
Chubb C.F., Potter I.C. 1984. The reproductive biology and estuarine movements of the gizzard shad, *Nematalosa vlaminghi* (Munro). J. Fish Biol. 25:527–543.
Claramunt S., Derryberry E.P., Remsen J.V. Jr, Brumfield R.T. 2012. High dispersal ability inhibits speciation in a continental radiation of passerine birds. Proc. Biol. Sci. 279:1567–1574.
Corush J.B. 2019. Evolutionary patterns of diadromy in fishes: more than a transitional state between marine and freshwater. BMC Evol. Biol. 19:1–13.
Crotty S.M., Minh B.Q., Bean N.G., Holland B.R., Tuke J., Jermiin L.S., Haeseler A.V. 2020. GHOST: recovering historical signal from heterotachously evolved sequence alignments. Syst. Biol. 69:249–264.
De Figueiredo F.J. 2009. A new marine clupeoid fish from the Lower Cretaceous of the Sergipe-Alagoas Basin, northeastern Brazil. Zootaxa. 2164:21–32.
Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.
DeHaan L.M., Burns M.D., Egan J.P., Bloom D.D. 2023. Diadromy drives elevated rates of trait evolution and ecomorphological convergence in clupeiformes (Herring, Shad, and Anchovies). Am. Naturalist. 202:830–850.
Delgado M.L., Ruzzante D.E. 2020. Investigating diadromy in fishes and its loss in an -omics era. Iscience. 23:101837.
Dos Reis M., Z. Yang. 2019. Bayesian molecular clock dating using genome-scale datasets. In: Anisimova, M., editor. Evolutionary Genomics. New York, NY: Humana. p. 309–330.
Dufour P., Descamps S., Chantepie S., Renaud J., Guéguen M., Schiffers K., Thuiller W., Lavergne S. 2020. Reconstructing the geographic and climatic origins of long-distance bird migrations. J. Biogeogr. 47:155–166.
Egan J.P., Bloom D.D., Kuo C.H., Hammer M.P., Tongnunui P., Iglésias S.P., Sheaves M., Grudpan C., Simons A.M. 2018a. Phylogenetic analysis of trophic niche evolution reveals a latitudinal herbivory gradient in Clupeoidei (herrings, anchovies, and allies). Mol. Phylogenet. Evol. 124:151–161.
Egan J.P., Gibbs S., Simons A.M. 2018b. Trophic niches through ontogeny in 12 species of Indo-Pacific marine Clupeoidei (herrings, sardines, and anchovies). Mar. Biol. 165:1–13.
Egan J.P., Bloom D.D., Simons A.M. 2022. Time for speciation and niche conservatism explain the latitudinal diversity gradient in clupeiform fishes. J. Biogeogr. 49:1952–1966.
Ezenwaji H.M.G., Offiah F.N. 2003. The biology of Pellonula Leonensis Boulenger, 1916 (Osteichthyes: Clupeidae) in Anambra River, Nigeria. J. Bio. Res. Biotech. 1:33-50.
Faria R., Weiss S., Alexandrino P. 2012. Comparative phylogeography and demographic history of European shads (*Alosa alosa* and *A. fallax*) inferred from mitochondrial DNA. BMC Evol. Biol. 12:194–120.
Felsenstein, J. 2004 Inferring phylogenies, *vol.* 2. Sunderland, MA: Sinauer Associates.
Feutry P., Castelin M., Ovenden J.R., Dettai A., Robinet T., Cruaud C., Keith P. 2013. Evolution of diadromy in fish: insights from a tropical genus (*Kuhlia* species). Am. Nat. 181:52–63.
Finnegan D.L., Egan J.P., Bloom D.D. 2024. Dine and dash: how trophic ecology and migration shape functional locomotory traits in clupeiform fishes. Biol. J. Linn. Soc.:blae046.
Flockhart D.T., Fitz-Gerald B., Brower L.P., Derbyshire R., Altizer S., Hobson K.A., Wassenaar L.I., Norris D.R. 2017. Migration distance as a selective episode for wing morphology in a migratory insect. Mov. Ecol. 5:1–9.
Freyman W.A., Höhna S. 2019. Stochastic character mapping of state-dependent diversification reveals that the tempo of evolutionary decline in self-compatible Onagraceae lineages. Syst. Biol. 58:505–519.
Fudickar A.M., Jahn A.E., Ketterson E.D. 2021. Animal migration: an overview of one of nature's great spectacles. Annu. Rev. Ecol. Evol. Syst. 52:479–497.
Futuyma D.J., Moreno G. 1988. The evolution of ecological specialization. Annu. Rev. Ecol. Syst. 19:207–233.

Goodman D.H., Reid S.B., Docker M.F., Haas G.R., Kinziger A.P. 2008. Mitochondrial DNA evidence for high levels of gene flow among populations of a widely distributed anadromous lamprey *Entosphenus tridentatus* (Petromyzontidae). J. Fish Biol. 72:400–417.
Grande L. 1985. Recent and fossil clupeomorph fishes with materials for revision of the subgroups of clupeoids. Bull. Am. Mus. Nat. Hist. 181:2.
Griffiths D. 2006. Pattern and process in the ecological biogeography of European freshwater fish. J. Anim. Ecol. 75(3):734–751.
Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.
Harvey M.G., Rabosky D.L. 2018. Continuous traits and speciation rates: alternatives to state-dependent diversification models. Methods Ecol. Evol. 9(4):984–993.
Hasegawa M., Hashimoto T. 1993. Ribosomal RNA trees misleading? Nature 361:23.
Hasselman D.J., Ricard D., Bentzen P. 2013. Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. Mol. Ecol. 22:1558–1573.
Hata H., Lavoué S., Motomura H. 2020. *Stolephorus acinaces*, a new anchovy from northern Borneo, and redescription of *Stolephorus andhraensis* Babu Rao, 1966 (Clupeiformes: Engraulidae). Mar. Biodivers. 50:1–11.
Hata H., Lavoué S., Motomura H. 2021a. Redescriptions of *Dussumieria acuta* Valenciennes 1847 and *Dussumieria albulina* (Fowler 1934), two valid species of rainbow sardines (Clupeiformes: Dussumieriidae). Ichthyol. Res. 68:126–138.
Hata H., Lavoué S., Motomura H. 2021b. Taxonomic status of nominal species of the anchovy genus *Stolephorus* previously regarded as synonyms of *Stolephorus commersonnii* Lacepède 1803 and Stolephorus indicus (van Hasselt 1823), and descriptions of three new species (Clupeiformes: Engraulidae). Ichthyol. Res. 68:1–46.
Hodcroft, E. TreeCollapser CL4. Available from: http://emmahodcroft.com/TreeCollapseCL.html (accessed April 2021).
Höhna S., Landis M.J., Heath T.A., Boussau B., Lartillot N., Moore B.R., Huelsenbeck J.P., Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65:726–736.
Hughes L.C., Ortí G., Huang Y., Sun Y., Baldwin C.C., Thompson A.W., Arcila D., Betancur-R R., Li C., Becker L., Bellora N., Zhao X., Li X., Wang M., Fang C., Xie B., Zhou Z., Huang H., Chen S., Venkatesh B., Shi Q. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc. Natl. Acad. Sci. U.S.A. 115:6249–6254.
Hughes L.C., Ortí G., Saad H., Li C., White W.T., Baldwin C.C., Crandall K.A., Arcila D., Betancur-R R. 2021. Exon probe sets and bioinformatics pipelines for all levels of fish phylogenomics. Mol Ecol Res. 21:816–833.
Hussain N.A., Al-Okailee M.T., Ahmed S. 2005. Food and feeding activity of *Ilisha megaloptera* larvae in Shatt Al-Arab Estuary. Iraqi J Aquacult. 2:100–108.
Inoue J.G., Miya M., Miller M.J., Sado T., Hanel R., Hatooka K., Aoyama J., Minegishi Y., Nishida M., Tsukamoto K. 2010. Deep-ocean origin of the freshwater eels. Biol. Lett. 6:363–366.
Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–589.
Kisel Y., Barraclough T.G. 2010. Speciation has a spatial scale that depends on levels of gene flow. Am. Nat. 175:316–334.
Kline P.A., Flagg T.A. 2014. Putting the red back in Redfish Lake, 20 years of progress toward saving the Pacific Northwest's most endangered salmon population. Fisheries. 39:488–500.
Kolmann M.A., Hughes L.C., Hernandez L.P., Arcila D., Betancur-R R., Sabaj M.H., López-Fernández H., Ortí G. 2020. Phylogenomics of piranhas and pacus (Serrasalmidae) uncovers how dietary convergence and parallelism obfuscate traditional morphological taxonomy. Syst. Biol. 70:576–592.
Kottelat M. 2013. The fishes of the inland waters of Southeast Asia: a catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. Raffles Bull. Zool. (Supplement 27):1–633.
Kottelat M., Freyhof J. 2007. Handbook of European freshwater fishes. Cornol, Switzerland: Publications Kottelat. p. 646.
Lal Mohan R.S. 1985. Spawning of *Nematalosa nasus* Bloch in Piollaimadom lagoon at Mandapam. Mar Fish Infor Serv, T and E Ser. 61:13–14.
Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 30:3276–3278.
Laudanno G., Haegeman B., Rabosky D.L., Etienne R.S. 2021. Detecting lineage-specific shifts in diversification: a proper likelihood approach. Syst. Biol. 70:389–407.
Lavoué S., Miya M., Musikasinthorn P., Chen W.J., Nishida M. 2013. Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). PLoS One 8:e56485.
Lavoué, S., Konstantinidis P., Chen W.J. 2014. Progress in clupeiform systematics. In: Ganias, K., editor. Biology and ecology of sardines and anchovies. Boca Raton: CRC Press. p. 3–42.
Lavoué S., Bertrand J.A., Wang H.Y., Chen W.J., Ho H.C., Motomura H., Hata H., Sado T., Miya M. 2017a. Molecular systematics of the anchovy genus *Encrasicholina* in the Northwest Pacific. PLoS One. 12:e0181329.
Lavoué S., Bertrand J.A., Chen W.J., Ho H.C., Motomura H., Sado T., Miya M. 2017b. Phylogenetic position of the rainbow sardine *Dussumieria* (Dussumieriidae) and its bearing on the early evolution of the Clupeoidei. Gene. 623:41–47.
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.
Li C., Hofreiter M., Straube N., Corrigan S., Naylor G.J.P. 2013. Capturing protein-coding genes across highly divergent species. Biotechniques. 54:321–326.
Limburg K.E., Hattala K.A., Kahnle A. 2003. American shad in its native range. Am. Fish. Soc. Symp. 35:125–140.
Lockhart P., Novis P., Milligan B.G., Riden J., Rambaut A., Larkum T. 2006. Heterotachy and tree building: a case study with plastids and eubacteria. Mol. Biol. Evol. 23:40–45.
Macqueen D.J., Johnston I.A. 2014. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc. Biol. Sci. 281:20132881.
Maddison W.P., Midford P.E., Otto S.P. 2007. Estimating a binary character's effect on speciation and extinction. Syst. Biol. 56:701–710.
Maddison W.P., FitzJohn R.G. 2014. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.
Mai A.C.G., Condini M.V., Albuquerque C.Q., Loebmann D., Saint'Pierre T.D., Miekeley N., Vieira J.P. 2014. High plasticity in habitat use of *Lycengraulis grossidens* (Clupeiformes, Engraulidae). Estuar. Coast. Shelf Sci. 141:17–25.
Malabarba M.C., Di Dario F. 2017. A new predatory herring-like fish (Teleostei: Clupeiformes) from the early Cretaceous of Brazil, and implications for relationships in the Clupeoidei. Zool. J. Linn. Soc. 180:175–194.
Manyukas Y.L. 1989. Biology of the Atlantic shad, *Alosa fallax fallax*, in Kurshskiy Bay. J Ichthyol. 29:125–128.
Marramà G., Carnevale G. 2016. An Eocene anchovy from Monte Bolca, Italy: the earliest known record for the family Engraulidae. Geol. Mag. 153:84–94.
Marramà G., Carnevale G. 2017. The relationships of *Gasteroclupea branisai* Signeux, 1964, a freshwater double-armored herring (Clupeomorpha, Ellimmichthyiformes) from the Late Cretaceous-Paleocene of South America. Hist. Biol. 29:904–917.
Martínez-Gómez J., Song M.J., Tribble C.M., Kopperud B.T., Freyman W.A., Höhna S., Specht C.D., Rothfels, C.J. 2024. Commonly used Bayesian diversification methods lead to biologically meaningful differences in branch-specific rates on empirical phylogenies. Evol. Lett. 8(2):189–199.
McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. 2013. A phylogeny of birds based on over 1,500

loci collected by target enrichment and high-throughput sequencing. PLoS One. 8:e54848.
McDowall, R.M. 1988. Diadromy in fishes: migrations between freshwater and marine environments. London: Croom Helm. p. 308.
McDowall R.M. 1997. Note on the conservation status of the giant bully, *Gobiomorphus gobioides* (Teleostei: Eleotridae). J. Royal Soc. New Zealand. 27:163–172.
McDowall R.M. 1999. Diadromy and genetic diversity in Nearctic and Palearctic fishes. Mol. Ecol. 8(4).
McDowall R.M. 2001. Diadromy, diversity and divergence: implications for speciation processes in fishes. Fish Fisheries. 2:278–285.
Mendonça J.T., Sobrinhob R.P. 2013. Management of fishing of the Broadband Anchovy (*Anchoviella lepidentostole*) (Fowler, 1911) in south São Paulo State, Brazil. Brazilian J. Biol. 73:691–697.
Miller R.R. 1982. First fossil record (Plio-Pleistocene) of threadfin shad, *Dorosoma petenense*, from the Gatuna Formation of southeastern New Mexico. J. Paleontol. 56:423–425.
Minh B.Q., Nguyen M.A.T., von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30:1188–1195.
Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 30:i541–i548.
Montes C., Cardona A., Jaramillo C., Pardo A., Silva J.C., Valencia V., Ayala C., Pérez-Angel L.C., Rodriguez-Parra L.A., Ramirez V., Niño H. 2015. Middle miocene closure of the Central American seaway. Science. 348:226–229.
Near T.J., Eytan R.I., Dornburg A., Kuhn K.L., Moore J.A., Davis M.P., Wainwright P.C., Friedman M., Smith W.L. 2012a. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. U.S.A. 109:13698–13703.
Near T.J., Dornburg A., Kuhn K.L., Eastman J.T., Pennington J.N., Patarnello T., Zane L., Fernández D.A., Jones C.D. 2012b. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Natl. Acad. Sci. U.S.A. 109:3434–3439.
Nelson J.S., Grande T.C., Wilson M.V. 2016. Fishes of the World. Hoboken, New Jersey: John Wiley & Sons. p. 752.
Ng J., Smith S.D. 2018. Why are red flowers so rare? Testing the macroevolutionary causes of tippiness. J. Evol. Biol. 31:1863–1875.
Nguyen L.T., Schmidt H.A., Von Haeseler A., Minh B.Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.
Nosil P. 2008. Speciation with gene flow could be common. Mol. Ecol. 17:2103–2106.
Palkovacs E.P., Dion K.B., Post D.M., Caccone A. 2008. Independent evolutionary origins of landlocked alewife populations and rapid parallel evolution of phenotypic traits. Mol. Ecol. 17:582–597.
Parham J.F., Donoghue P.C., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S., Smith N.D., Tarver J.E., van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L., Makovicky P.J., Müller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J. 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61:346–359.
Plummer M., Best N., Cowles K., Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News. 6:7–11.
R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/
Rainboth W.J. 1996. Fishes of the Cambodian Mekong. United Nations Food and Agriculture Organization species identification field guide for fishery purposes. Rome: United Nations Food and Agriculture Organization. p. 265.
Rabosky D.L. 2010. Extinction rates should not be estimated from molecular phylogenies. Evolution 64:1816–1824.
Rabosky D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity dependence on phylogenetic trees. PLoS One. 9:e89543.
Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H., Larson J.G. 2014. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods Ecol. Evol. 5:701–707.
Rabosky D.L., Goldberg E.E. 2015. Model inadequacy and mistaken inference of trait-dependent speciation. Syst. Biol. 64:340–355.
Rabosky D.L., Huang H. 2016. A robust semi-parametric test for detecting trait-dependent diversification. Syst. Biol. 65:181–193.
Rabosky D.L., Goldberg E.E. 2017. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. Evolution 71:1432–1442.
Rabosky D.L., Chang J., Title P.O., Cowman P.F., Sallan L., Friedman M., Kaschner K., Garilao C., Near T.J., Coll M., Alfaro M.E. 2018. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 559:392–395.
Rambaut A., Drummond A.J. 2015. TreeAnnotator v1.8.2. MCMC Output analysis.
Remya V.C., Benakappa S., Mahesh V., Naik A.S., Anjanayappa H.N., Vijaykumar M.E. 2019. Breeding seasonality of *Anodontostoma chacunda* (Hamilton, 1822) off Mangalore coast, Karnataka, India. Indian J. Mar. Sci. 48:628–634.
Raymo M.E., Ruddiman W.F. 1992. Tectonic forcing of late Cenozoic climate. Nature 359:117–122.
Revell L.J. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.
Rincon-Sandoval M., Duarte-Ribeiro E., Davis A.M., Santaquiteria A., Hughes L.C., Baldwin C.C., Soto-Torres L., Acero A., Walker H.J., Carpenter K.E., Sheaves M. 2020. Evolutionary determinism and convergence associated with water-column transitions in marine fishes. PNAS 117:33396–33403.
Rodriguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389–399.
Roff D.A. 1988. The evolution of migration and some life history parameters in marine fishes. Environ. Biol. Fishes 22:133–146.
Roff D.A. 1991. Life history consequences of bioenergetic and biomechanical constraints on migration. Am. Zool. 31:205–216.
Rolland J., Jiguet F., Jønsson K.A., Condamine F.L., Morlon H. 2014. Settling down of seasonal migrants promotes bird diversification. Proc. Biol. Sci. 281:20140473.
Santaquiteria A., Siqueira A.C., Duarte-Ribeiro E., Carnevale G., White W.T., Pogonoski J.J., Baldwin C.C., Ortí G., Arcila D., Betancur-R R. 2021. Phylogenomics and historical biogeography of seahorses, dragonets, goatfishes, and allies (Teleostei: Syngnatharia): assessing factors driving uncertainty in biogeographic inferences. Syst. Biol. 70:1145–1162.
Sayers E.W., Bolton E.E., Brister J.R., Canese K., Chan J., Comeau D.C., Connor R., Funk K., Kelly C., Kim S., Madej T., Marchler-Bauer A., Lanczycki C., Lathrop S., Lu Z., Thibaud-Nissen F., Murphy T., Phan L., Skripchenko Y., Tse T., Wang J., Williams R., Trawick B.W., Pruitt K.D., Sherry S.T. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res. 50:D20–D26.
Sayyari E., Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33:1654–1668.
Strydom N.A., Whitfield A.K., Paterson A.W. 2002. Influence of altered freshwater flow regimes on abundance of larval and juvenile *Gilchristella aestuaria* (Pisces: Clupeidae) in the upper reaches of two South African estuaries. Mar Freshwater Res. 53:431–438.
Thacker C.E., Shelley J.J., McCraney W.T., Unmack P.J., McGee M.D. 2021. Delayed adaptive radiation among New Zealand stream fishes: joint estimation of divergence time and trait evolution in a newly delineated island species flock. Syst. Biol. 71:13–23.
Title P.O., Rabosky D.L. 2019. Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? Methods Ecol. Evol. 10:821–834.
Tribble C.M., Freyman W.A., Landis M.J., Lim J.Y., Barido-Sottani J., Kopperud B.T., Höhna S., May M.R. 2022. RevGadgets: an R package for visualizing Bayesian phylogenetic analyses from RevBayes. Meth Ecol Evol. 13:314–323.
Twining C.W., Palkovacs E.P., Friedman M.A., Hasselman D.J., Post D.M. 2017. Nutrient loading by anadromous fishes: species-specific contributions and the effects of diversity. Can. J. Fish. Aquat. Sci. 74:609–619.
Van Tienderen P.H. 1991. Evolution of generalists and specialists in spatially heterogeneous environments. Evolution 45:1317–1331.
Velotta J.P., McCormick S.D., Jones A.W., Schultz E.T. 2018. Reduced swimming performance repeatedly evolves on loss of migration

in landlocked populations of alewife. Physiol. Biochem. Zool. 91:814–825.
Wang Q., Dizaj L.P., Huang J., Sarker K.K., Kevrekidis C., Reichenbacher B., Esmaeili H.R., Straube N., Moritz T., Li C. 2022. Molecular phylogenetics of the Clupeiformes based on exon-capture data and a new classification of the order. Mol. Phylogenet. Evol. 175:107590.
Whitehead P.J.P. 1985. Clupeoid fishes of the World (Suborder Clupeoidei): An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf herrings. Part 1. Chirocentridae, Clupeidae and Pristigasteridae. FAO Fish Synop. 125:1–303.
Whitehead P.J.P., Nelson G.J., Wongratana T. 1988. Clupeoid fishes of the World (Suborder Clupeoidei): an annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf herrings. Part 2. Engraulididae. FAO Fish Synop. 125:305–579.
Whitfield A.K., Harrison T.D. 1996. *Gilchristella aestuaria* (Pisces: Clupeidae) biomass and consumption of zooplankton in the Sundays Estuary. S. Afr. J. Mar. Sci. 17(1):49–53.
Willson M.F., Hulupka K.C. 1995. Anadromous fish as keystone species in vertebrate communities. Cons Biol. 9:489–497.
Winger B.M., Barker F.K., Ree R.H. 2014. Temperate origins of long-distance seasonal migration in New World songbirds. Proc. Natl. Acad. Sci. U.S.A. 111:12115–12120.
Yamanoue Y., Miya M., Matsuura K., Miyazawa S., Tsukamoto N., Doi H., Takahashi H., Mabuchi K., Nishida M., Sakai H. 2009. Explosive speciation of *Takifugu*: another use of fugu as a model system for evolutionary biology. Mol. Biol. Evol. 26:623–629.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.
Yang Z., Chen Y.F. 2008. Differences in reproductive strategies between obscure puffer *Takifugu obscurus* and ocellated puffer *Takifugu ocellatus* during their spawning migration. J. Appl. Ichthyol. 24:569–573.
Zachos J., Pagani M., Sloan L., Thomas E., Billups, K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. science. 292(5517):686–693.
Zhang C., Sayyari E., and Mirarab S. 2017. ASTRAL-III: increased scalability and impacts of contracting low support branches. In: Meidanis J., Nakhleh L., editors. RECOMB international workshop on comparative genomics. Cham: Springer. p. 53–75.
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform. 19:15–30.