Title
Phylogeny and Divergence Times of Gobiarian Fishes

Author
MCCRANEY, WILLIAM TYLER

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Phylogeny and Divergence Times of Gobiarian Fishes

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

William Tyler McCraney

2019

ABSTRACT OF THE DISSERTATION

Phylogeny and Divergence Times of Gobiarian Fishes

by

William Tyler McCraney

Doctor of Philosophy in Biology

University of California, Los Angeles, 2019

Professor Michael Edward Alfaro, Chair

Gobiarian fishes, exemplified by gobies, sleepers and cardinalfishes, have radiated across coastal marine and aquatic habitats worldwide, yet the biological traits deemed responsible for generating their great diversity, such as small body size, short generation times, and ecomorphological specialization, have also hindered resolution of their phylogeny. The current classification of Gobiaria is based on molecular phylogenetics, and while broad relationships among major groups have largely been settled by independent investigations, the placement of all higher taxa have not been verified with multilocus data. The root topology recovered from different molecular datasets is contentious, with multilocus and phylogenomic studies resolving nursery and cardinalfishes either in reciprocal or sequential sister clades to other gobiarians, and this deep systematic controversy questions the current two-order classification. Here I used two complementary approaches to resolve the phylogeny of Gobiaria. In my first study I mined public databases to assemble a sparse supermatrix of 23 genes and construct a phylogenetic tree with dense taxon sampling, comprised of approximately 30 percent of the more than 2,400

known gobiarian species. In my second study I generated new ultraconserved element sequence data to assemble a phylogenomic matrix of 704 genes, construct a phylogenetic tree with comprehensive sampling of higher taxa, and estimate a timescale for diversification under the relaxed molecular clock. Overall my studies produced more evidence in support of the current two-order classification of Gobiaria, and revealed the root uncertainty is a result of ancient incomplete lineage sorting. I also discovered that collared wrigglers (*Xenisthmus spp.*) form a monophyletic clade separate from sleepers (Eleotridae), and recommend recognizing family Xenisthmidae in the clade-based classification of Gobiaria. I dated origination of Gobiaria in the youngest age of the Early Cretaceous (104 Ma), found major clades of gobies, sleepers, and cardinalfishes diverged in the early Eocene (~50 Ma), and placed goby lineage diversification in the Oligocene and Miocene. In summary, my studies support the current two-order classification placing nursery and cardinalfishes in Kurtiformes and the remaining gobiarian fishes in Gobiiformes, confirm the clade-based phylogenetic classification of gobiarian families, and advance evidence for recognizing collared wrigglers in family Xenisthmidae.

The dissertation of William Tyler McCraney is approved.

Paul Henry Barber

Thomas Welch Gillespie

David K. Jacobs

Christine Thacker

Michael Edward Alfaro, Committee Chair

University of California, Los Angeles

2019

In memory of my father.

William Jeffrey McCraney 1955-2000.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful to my advisor, Michael Alfaro, for his unwavering enthusiasm and support of my research. Thank you for inspiring me with your scholarship and professionalism.

Thank you to my committee members Paul Barber, Dave Jacobs, and Tom Gillespie. You were all there for me when I needed you the most. I am especially grateful to Paul Barber for advising me during the first few years of my dissertation, and providing me with once-in-a-lifetime opportunities for research in Indonesia. Thank you Paul.

I am forever indebted to my external committee member, Christine Thacker, for her exceptional mentorship. I would not have been able to complete my dissertation without the intellectual and material support you provided. Your expertise and passion for ichthyology is awesome.

This dissertation was supported by the National Science Foundation East Asia Pacific Summer Institute, the Society for the Study of Evolution Rosemary Grant, the American Society of Ichthyologists and Herpetologists Raney Award, the Center for Tropical Research Franklin Grant, and several UCLA Ecology and Evolutionary Biology Departmental Travel and Research Awards.

Chapter 1 is adapted from a manuscript that was accepted (with major revisions) for publication in *Molecular Phylogenetics and Evolution*. Christine Thacker and Michael Alfaro are co-authors, and both are credited with developing the study and editing the manuscript.

Chapter 2 is in preparation for submission to *Systematic Biology*, and includes Christine Thacker, Janet Buckner, Brant Faircloth, Richard Harrington, Thomas Near, and Michael Alfaro as co-authors. CT, BF, TN, and MA are credited with developing the study, JB and RH contributed many hours of laboratory work, and all authors are credited with editing the manuscript.

W Tyler McCraney **Curriculum Vitae**

University of California, Los Angeles (UCLA)
Department of Ecology and Evolutionary Biology (EEB)
621 Charles E Young Drive South
Los Angeles, CA 90095

707-497-9309
tmccraney@g.ucla.edu
w.tyler.mccraney@gmail.com

Education

2015	C Phil Biology - UCLA - Alfaro Lab
2009	MS Natural Resources - Humboldt State University (HSU) - Kinziger Lab
2006	BS Fisheries Biology - HSU

Awards

2018	American Society of Ichthyologists and Herpetologists - Raney Award
2018-17	EEB - research and travel award
2016	Western Division American Fisheries Society - conference travel grant
2015	NSF - East Asia Pacific Summer Institute, Taiwan
2014	Society for the Study of Evolution - Rosemary Grant award
2014	Center for Southeast Asian Studies, UCLA - Lemelson fellowship
2014	Center for Tropical Research, UCLA - Betty and EP Franklin grant
2014-13	DOE - Foreign Language and Area Studies - Indonesian
2009	Granite Bay Fly Casters - graduate scholarship
2008	Marin Rod and Gun Club - graduate scholarship

Publications

Kinziger AP, Hellmair M, **McCraney WT**, Jacobs DK, Goldsmith G (2015) Temporal genetic analysis of the endangered tidewater goby: extinction–colonization dynamics or drift in isolation? *Molecular Ecology*, 24, 5544-5560.

McCraney WT (2012) Education, collaboration and friendship at the Indonesian Biodiversity Research Center. *Ocean Lab Magazine*, 3, 20-24.

McCraney WT, Farley EV, Kondzela CM, Naydenko SV, Starovoytov AN, Guyon JR (2012) Genetic stock identification of overwintering chum salmon in the North Pacific Ocean. *Environmental Biology of Fishes*, 94, 663-668.

Kondzela CM, **McCraney WT**, Nguyen HT, Guyon JR (2012) Genetic stock composition analysis of chum salmon bycatch samples from the 2010 Bering Sea groundfish fisheries. *NOAA Technical Memorandum, National Marine Fisheries Service - Alaska Fisheries Science Center*, 233.

McCraney WT, Saski CA, Guyon JR (2011) Isolation and characterization of 12 microsatellites for the commercially important sablefish, *Anoplopoma fimbria*. *Conservation Genetics Resources*, 4, 415-417.

McCraney WT, Kondzela CM, Murphy J, Guyon JR (2010) Genetic stock identification of chum salmon from the 2006 and 2007 Bering-Aleutian Salmon International Survey. *North Pacific Anadromous Fish Commission,* 1288, 1-11.

Publications - continued

McCraney WT, Goldsmith G, Jacobs DK, Kinziger AP (2010) Rampant drift in artificially fragmented populations of the endangered tidewater goby (*Eucyclogobius newberryi*). *Molecular Ecology*, 19, 3315-3327.

Presentations

2018	Oral - American Society of Ichthyologists and Herpetologists - Rochester, NY
2017	Oral - Indo-Pacific Fish Conference - Papeete, Tahiti
2017, 16, 13	Oral - Eco-Evo Pub - UCLA
2017	Oral - Society for Integrative and Comparative Biology - New Orleans, LA
2016	Oral- Western Division American Fisheries Society - Reno, NV
2012	Poster - Northeast Pacific Pink and Chum Salmon Workshop - Juneau, AK
2011	Oral - American Fisheries Society - Seattle, WA
2010	Oral - Alaska Chapter American Fisheries Society - Juneau, AK
2010	Oral - Coastwide Salmonid Genetics Meeting - Boise, ID
2009	Oral - University of Alaska Fairbanks - Fisheries Division - Juneau, AK
2008	Oral- Humboldt Bay Symposium - Eureka, CA
2008	Oral - Gilbert Ichthyological Society - Klamath, OR
2007	Poster - American Fisheries Society - San Francisco, CA
2007	Poster - American Society of Ichthyologists and Herpetologists - St Louis, MO
2006	Oral - Gilbert Ichthyological Society - Newport, OR

Teaching and Research

2019,18,17,13	Teaching Fellow	*Introduction to Marine Science Lab*, UCLA
2019,18,17	Teaching Fellow	*Evolutionary Applications for Fisheries and Wildlife*, UCLA
2019	Teaching Fellow	*Advanced Statistics for Ecology and Evolutionary Biology*
2018,16	Teaching Fellow	*Comparative Biology and Macroevolution*, UCLA
2017,16,15	Teaching Fellow	*Evolution, Ecology, and Biodiversity*, UCLA
2017-16	Teaching Fellow	*Evolution of Cosmos and Life*, UCLA
2016	Teaching Associate	*Evolution*, UCLA
2015,13	Teaching Associate	*Life: Concepts and Issues*, UCLA
2013	Teaching Fellow	*Introduction to Marine Science*, UCLA
2012-09	Genetics Technician	NOAA - Juneau, AK

Academic and Community Service

2018	Organizer and Host - EEB Monday social - UCLA
2018-17	Representative - Biological Science Council - UCLA
2018-15	Laboratory Safety Coordinator - 2154 Terasaki Life Science Building - UCLA
2016-15	Organizer and Moderator - Eco-Evo Pub - UCLA
2015-14	Scientific Diver - LA Waterkeeper Kelp Project - Los Angeles, CA

Certifications

2015 Institutional Animal Care and Use Committee protocol
2012 *Scientific Diver* - American Academy of Underwater Sciences

CHAPTER 1

Supermatrix Phylogeny Resolves Goby Lineages and Reveals Unstable Root of Gobiaria

1.1 Abstract

Gobies, sleepers, and cardinalfishes represent major clades of a species rich radiation of small bodied, ecologically diverse percomorphs (Gobiaria). Molecular phylogenetics has been crucial to resolving broad relationships of sleepers and gobies (Gobioidei), but the phylogenetic placements of cardinalfishes and nurseryfishes, as reciprocal or sequential sister clades to Gobioidei are uncertain. In order to evaluate relationships among and within families, we used a phylogenetic data mining approach to generate densely sampled trees inclusive of all higher taxa. We utilized conspecific amino acid homology to improve alignment accuracy, included ambiguously identified taxa to increase taxon sampling density, and resampled individual gene alignments to filter rogue sequences before concatenation. This approach yielded the most comprehensive tree yet of Gobiaria, inferred from a sparse (17 percent-complete) supermatrix of one ribosomal and 22 protein coding loci (18,065 characters), comprised of 50 outgroup and 777 ingroup taxa, representing 32 percent of species and 68 percent of genera. Our analyses confirmed the lineage-based classification of gobies with strong support, identified sleeper clades with unforeseen levels of systematic uncertainty, and quantified competing phylogenetic signals that confound resolution of the root topology. We also discovered that multilocus data completeness was related to maximum likelihood branch support, and verified that the phylogenetic uncertainty of shallow relationships observed within goby lineages could largely be explained by supermatrix sparseness. These results highlight both the potential and limits of

publicly available sequence data for producing densely-sampled phylogenetic trees of exceptionally biodiverse groups.

1.2 Introduction

The current hypothesis for acanthomorph phylogeny resolves Gobiaria, a globally distributed, species rich radiation of small bodied nearshore and freshwater fishes as an early diverging group among percomorphs (Alfaro et al. 2018; Hughes et al. 2018). Accounting for nearly one in twenty described vertebrate species, Gobiaria hosts incredible diversity in morphology, ecophysiology, and behavioral specialization, including the shortest lifespans and smallest body sizes recorded among vertebrates (Depczynski and Bellwood 2005; Watson and Walker 2004). Taxonomy of Gobiaria began with descriptions of *Gobius* and *Apogon* (Linnaeus 1758), yet remains in a discovery phase (Jaafar and Murdy 2017), with a quarter of the 2,412 species in Eschmeyer et al. (2016) described over the past 20 years.

Gobiaria is principally composed of gobies (Gobiidae, Gobionellidae) and sleepers (Eleotridae, Butidae), who share a similar suite of general shape features including separated dorsal fins and a rounded body, pectoral, and caudal fins (Figure 1). The pelvic fins may be fused, often forming a suction disk that holds to diverse substrates ranging from corals and sponges, to rocky temperate reefs and waterfalls atop tropical oceanic islands. A few depauperate taxa also group with gobies and sleepers, including ocean sleepers (Thalasseleotrididae), freshwater sleepers (Odontobutidae), blind cave gobies (Milyeringidae), and loach gobies (Rhyacichthyidae). The eight families of gobies and sleepers collectively comprise the taxon Gobioidei, and their morphological synapomorphies include a gland associated with the sperm duct in males (accessory gonadal structure), the loss of parietal bones in the skull, and a widely displaced symplectic and preopercle in the suspensorium (Miller 1992; Winterbottom 1993).

Skeletal characters are numerous but difficult to interpret among gobioids because they are small fishes that often present variable patterns of loss within and between lineages

(Birdsong et al. 1988; Johnson and Brothers 1993). The number of branchiostegal rays is considered the most reliable character for higher-level classification (Fig. 1). Unlike percomorph fishes or non-gobioid gobiarians, who have seven branchiostegal rays (7BR), gobioids have six or fewer. Six branchiostegal rays (6BR) is the plesiomorphic character state observed among early diverging loach gobies, blind cave gobies, and sleepers, and five branchiostegal rays (5BR) is a synapomorphy for gobies in Gobiidae and Gobionellidae (Fig. 1).

Higher classification of gobioids was originally hypothesized on morphology and subsequently revised with molecules, but the systematic relationships among non-gobioid gobiarians, and between the non-gobioids and gobioids, are entirely based upon analysis of DNA sequences (Smith and Wheeler 2006; Thacker 2009; Thacker et al. 2015). Relationships at the root of Gobiaria are controversial because different taxon and character sampling between studies (Fig. 2) has yielded conflicting topologies (Fig. 3).

Sister to gobioids are sanddivers (Trichonotidae), and sister to gobioids plus sanddivers are cardinalfishes (Apogonidae), and nurseryfishes (Kurtidae) (Fig. 1). Sanddivers were recently identified as the sister taxon to gobioids in a molecular phylogenetic hypothesis of Gobiaria that included comprehensive sampling of families (Thacker et al. 2015), but the affinity of nurseryfishes and cardinalfishes to one another, and to gobioids was noted by Johnson (1993) before a broadly sampled molecular hypothesis drew attention to their systematic relationships (Smith and Wheeler 2006). Subsequent reexamination of sensory and reproductive characters yielded a loosely defined set of shared characters for Gobiaria that includes grid-like patterns of cephalic pores and papillae, egg adhesion, and egg brooding on forehead (nurseryfishes), in buccal cavity (cardinalfishes, sanddivers), or in nests (sleepers, gobies) (Thacker 2009; Thacker et al. 2015), although these characters are not unique to Gobiaria (Nelson et al. 2016).

Recent phylogenomic studies have confirmed gobioid phylogenetic classification with resolution of two goby clades, three sleeper clades (Milyeringidae, Eleotridae, Butidae), and an early-diverging clade of freshwater sleepers (Odontobutidae) and loach gobies (Rhyacichthyidae) (Li et al. 2017; Alfaro et al. 2018; Kuang et al. 2018). However, the two gobioid-specific phylogenomic studies of exon sequences, as well as another broadly-sampled acanthomorph study of ultraconserved element sequences (Alfaro et al. 2018), all resolved a root topology that placed nurseryfishes and cardinalfishes as successive sisters to sanddivers and gobioids (Fig. 3). Curiously, the sequential clade relationship of gobioids, nurseryfishes, and cardinalfishes was initially recovered from early mitochondrial DNA trees that formed the basis of our current phylogenetic classification (Thacker 2009; Chakrabarty et al. 2012).

Within the largest families Gobiidae and Gobionellidae, difficulty in producing a phylogenetic hypothesis robust to sparse, variable taxon sampling has hampered progress in advancing systematic relationships among the nearly 2,000 species of gobies (Fig. 1C). Several molecular phylogenetic studies have converged on a structure including 19 lineages (subfamilial clades), five in Gobionellidae and 14 in Gobiidae (Thacker, 2009; Thacker and Roje 2011; Agorreta et al. 2013; Thacker 2013), although the broad relationships among the lineages remain fluid and largely uncertain (Fig. 1C). Surprisingly, Gobiidae was paraphyletic with respect to Gobionellidae in a recent unconstrained phylogeny of more than 14,000 acanthomorph fishes, which included 605 taxa sampled from Gobiaria (Rabosky et al. 2018). The supermatrix assembly pipeline used by Rabosky et al. (2018), implemented with PHLAWD (Smith et al. 2009), excluded incompletely described (*sp.*) and ambiguously identified (*cf.*) taxa, but these samples are common in molecular phylogenetic studies of Gobiaria (Fig. 2). For instance, over 12 percent of taxa sampled in Agorreta et al.'s (2013) tree were not described to the species level, and another goby tree of 75 taxa inferred from two nuclear loci sampled more than 13

percent ambiguous taxa (Tornabene et al. 2013). One explanation for the unexpected paraphyly of Gobiidae in Rabosky et al.'s (2018) tree could be that exclusion of ambiguously described taxa reduced phylogenetic signal of goby lineages. If this is correct then additional sampling of ambiguously described taxa may provide increased phylogenetic signal to resolve goby lineages and families. Another explanation could be that the automated pipeline used for supermatrix assembly in Rabosky et al. (2018) sampled "dirty data" from GenBank characterized by errors in specimen identification, DNA sequencing, or orthology (Hinchliff and Smith 2014), in which case avoidance of error-suspected data may resolve goby families. Alternatively, lack of resolution of goby families in Rabosky et al.'s (2018) tree may stem from low phylogenetic signal present in the sparse, yet densely taxon-sampled supermatrix, or from biological scenarios such as incomplete lineage sorting that occurred during periods of rapid diversification.

Here we used publicly available multilocus sequence data to investigate the phylogenetic resolution of goby lineages, and the relationships among gobioids, cardinalfishes, and nurseryfishes. We expected that increased taxon sampling, inclusion of ambiguously described species, and application of varied tree construction and summary methods would help clarify conflicts in the phylogenetic classification of Gobiaria. In order to mitigate the presence of "dirty data" in our supermatrix, we developed an analysis pipeline based on amino acid sequence alignments, and filtered rogue and questionable sequences from locus partitions prior to concatenation. Then to distinguish poor resolution due to supermatrix sparseness, low or conflicting phylogenetic signals, or incomplete lineage sorting, we constructed maximum likelihood and multispecies coalescent trees, and compared their relationships and branch support using a variety of measures. This strategy produced the most comprehensive molecular phylogenetic hypothesis yet of Gobiaria, resolved goby lineages with strong support, and

revealed extensive phylogenetic uncertainty within and among sleeper, nurseryfish and cardinalfish clades. Our results provide a summary of the current knowledge of gobiarian relationships and highlight persistent uncertainties in the phylogeny and systematics of this diverse radiation.

1.3 Material and Methods

1.3.1 Alignments

GenBank queries for Gobiaria and representatives of acanthomorph outgroup families Lampridae, Polymixiidae, Percopsidae, Aphredoderidae, Anoplogasteridae, Holocentridae, and Scombridae were downloaded in INSDSeq XML format, and parsed using xml2 and tidyverse packages in R (Wickham et al. 2017; Wickham 2017). To accomodate indexing samples with incorrect, synonymous, or ambiguous taxonomy, unique sample identifiers were created for each GenBank record using isolate and specimen voucher information. NCBI taxonomy was reconciled to FishBase.org using rfishbase (Boettiger et al. 2012; Froese and Pauly 2017) and remaining taxonomic ambiguities were resolved with Catalog of Fishes (Eschmeyer et al. 2016).

In order to assemble a sequence matrix with multilocus scaffolding across deeply-diverged ingroup and outgroup clades, individual loci were selected by identifying genes sequenced in at least two of the three gobiarian groups defined by the number of branchiostegal rays (Fig. 1), and at least three of the seven acanthomorph outgroup families. Coding sequences (CDS) of selected loci were downloaded with reutils (Schöfl 2016), sequences of protein-coding genes from annotated mitochondrial genomes were obtained with AnnotationBustR (Borstein and O'Meara 2017), and fasta files of amino acid (AA) and CDS were formatted with Biostrings (Pagès et al. 2017).

Partial sequences in gene alignments can be more problematic to species tree analyses than entire missing sequences (Hosner et al. 2016), but determination of what constitutes a partial sequence in multiple sequence alignments that sample thousands of taxa can be subjective, and may not be reproducible. To automate identification and removal of partial sequences from locus alignments, AA sequences were preliminarily aligned with MAFFT (Katoh and Standley 2013), and the alignments were screened for spurious sequences with trimAl

(Capella-Gutiérrez et al. 2009). Spurious sequences may appear in multiple sequence alignments as incomplete or poorly aligned when compared to other sequences in the alignment. The trimAl model for spurious sequence screening uses a residue overlap threshold to set the minimum residue overlap score for each residue, and a sequence overlap threshold to set the minimum proportion of residues required of each sequence to pass the residue overlap threshold (Capella-Gutiérrez et al. 2009). Sequences that do not pass the sequence overlap threshold are considered spurious and removed from the preliminary alignments. Residue and sequence overlap threshold parameters were set empirically, and spurious-filtered, preliminary alignments were visually inspected with Geneious (Kearse et al. 2012).

Alignment accuracy of deeply diverged taxa can be improved by including several homologous sequences from each distantly related taxon (Katoh et al. 2005). To improve alignment accuracy with positions of conserved residues and gaps provided by homologs, AA sequences from multiple individuals of each taxon, if available, were aligned with MAFFT (--localpair). Local pairwise alignment is the most accurate method available in MAFFT because it assumes that sequences have a central continuous alignable region (Katoh et al. 2005). After alignment, edges and other poorly-aligned positions of the multi-individual AA sequences were trimmed with trimAl (-automated1), and CDS were back-translated to AA sequences of the trimmed alignments to preserve reading frames.

Alignment, filtering, and trimming methods were modified for non-coding ribosomal sequences. MAFFT (--genafpair) was used to accommodate interspersed unalignable regions of ribosomal sequences (Katoh and Frith 2012), the preliminary alignment was visually inspected with Geneious (Kearse et al. 2012), and poorly-aligned or incomplete spurious sequences were manually removed from the alignment. The filtered ribosomal sequences were realigned with

MAFFT (--genafpair), and the final multi-individual ribosomal alignment was trimmed with trimAl (-automated1; Capella-Gutiérrez et al. 2009).

1.3.2 Gene Trees

Gene trees were inferred using three iterations of tree searching, bootstrapping, and rogue taxon analysis. The first iteration was used to guide selection of single individuals from alignments where multiple individuals of taxa were present, the second iteration was used to filter rogue and erroneous taxa from single-individual alignments, and the third iteration was used to infer gene trees and bootstraps on the final gene alignments.

To begin, best-fitting codon position partitioning schemes of GTR+G models from the five possible arrangements were selected with PartitionFinder2 (Lanfear et al. 2014), gene trees were inferred on multi-individual alignments with RAxML (Stamatakis 2014), and multi-individual gene tree inferences were bootstrapped to standard criteria with autoMRE bootstoping (Stamatakis et al. 2008; Pattengale et al. 2010). Rogue taxa in multi-individual gene trees were identified with RogueNaRok (Aberer et al. 2011), intruder and outlier individuals that violated monophyly at the species level were identified with MonoPhy (Schwery and O'Meara 2016), and relative composition frequency variability (RCFV) values were computed for each sequence with BaCoCa (Kück and Struck 2014). The number of ambiguous and gap sites were tallied for each sequence, and combined with rogue taxon identification, monophyly status, and RCFV scores in decision tables. A single individual was chosen from multiple individuals present in each alignment using the decision tables to guide selection of individuals with the best available data quality. Rogue taxa were not filtered from individual alignments in the first iteration of gene tree inference. Identification of rogue taxa was only used to avoid selecting rogues when there were better quality conspecific sequences available to choose from. For example, of four *Elacatinus*

illecebrosus sequences present in the multi-individual *rag1* alignment, two sequences were identified as rogue taxa and ambiguous characters were present in one of the remaining sequences, therefore the sequence with no ambiguous characters was selected for downstream analysis.

In the next iteration of gene tree analyses, phylogenies were inferred from partitioned single individual alignments, bootstrapped to standard criteria with autoMRE bootstoping, and filtered for rogue taxa identified in dropsets with RAxML (-J) on strict and majority rule consensus (MRC) trees (Pattengale et al. 2011). Rogue-filtered gene trees were inspected and sequences that violated monophyly at the family rank were manually removed from alignments. After filtering rogue and erroneously-placed taxa from alignments, a final iteration of analyses were conducted to infer gene trees and autoMRE bootstraps under GTR+G models selected from best-fitting codon position partitioning schemes (Lanfear et al. 2014).

1.3.3 Concatenated Trees

Phylogeny was inferred from the concatenated alignment using two iterations of ML tree searches, bootstrapping, and rogue taxon analysis. In the first iteration, rogue-filtered, single-individual gene alignments were concatenated with Phyutility (Smith and Dunn 2008), CDS were partitioned by codon positions, and phylogeny and bootstraps were inferred with autoMRE bootstopping on the concatenated alignment with RAxML (-U). The -U option was used to reduce computational burdens arising from tree inference on the large, sparse, concatenated alignment (Izquierdo-Carrasco et al. 2011). Rogue taxa were identified from bootstrap dropsets with RAxML (-J) and removed from the concatenated alignment. In the second iteration, phylogeny and bootstraps were inferred with autoMRE bootstopping and

branch length optimization on the rogue-filtered, concatenated alignment using GTR+G models partitioned by codon position.

The supermatrix approach yields sparse concatenated alignments characterized by a majority of missing data (Thomson and Shaffer 2010). Sparseness may not reduce the accuracy of phylogenetic inference if many characters have been sampled and the missing data are randomly distributed across lineages (Wiens 2006). However, missing data may reduce the reliability of bootstraps for indicating whether the position of a branch is supported in the best ML tree (Hinchliff and Roalson 2013), and a positive relationship of branch support to sparseness may explain poor resolution within clades when taxa are unexpectedly non-monophyletic. Strategies to understand how sparseness is distributed across lineages, and how the missing data affect bootstrap and other measures of branch support, can be useful for determining the extent to which phylogenetic accuracy may have been compromised by missing data (Wiens et al. 2005; Pyron et al. 2011; Hinchliff and Smith 2014). To complement bootstraps and assess the impact of missing data on branch support, confidence in branches was measured with methods based on likelihood and information theory, the depth of locus coverage on branches was computed from taxon occupancy of gene partitions in the concatenated alignment, and the relationships of all branch support measures to locus coverage were tested with simple linear regression.

A likelihood-based measure of confidence, Shimodaira Hasegawa-like (SH-like) branch support, was computed on the final ML tree using RAxML (-f J; Anisimova and Gascuel 2006). For each internal branch of a nearest neighbor interchange (NNI) optimized tree, SH-like support is computed with an approximate likelihood ratio test (aLRT) of the ML topology and two NNI topologies with lower log-likelihood scores. The aLRT considers the null hypothesis of no difference in log likelihoods of best ML and NNI topologies. SH-like branch support is drawn as

the difference of the aLRT *p*-value from one, and values of 0.85 or greater are interpreted as strong support for the observed branch (Guindon et al. 2010; Pyron et al. 2011).

The information theory-based measure used a consensus summary of bootstraps to assess branch support. Topology and branch lengths of ML bootstrap inferences were summarized into a MRC tree with consense (Aberer et al. 2014), and internode certainty (IC) scores were computed on internal branches of the bootstrap MRC tree using RAxML (-L MR -C) with verbose splits output (Salichos et al. 2014). IC measures the difference in gene support frequencies of the most common bipartition observed in the bootstrap set to the second-most common bipartition, if there are any observed at a frequency greater than 5% (Salichos et al. 2014). IC is drawn as a proportion ranging from zero, indicating equivalent amounts of conflicting phylogenetic signal, to one, which indicates absence of any conflicting phylogenetic signals informing the branch.

The depth of locus coverage on branches of the best ML tree was computed as the number of loci with decisive taxon sampling (Hinchliff and Smith 2014). Decisive taxon sampling on a branch describes the possibility of a branch to be informed based solely on the taxa present in a locus partition (Steel and Sanderson 2010). A locus has decisive taxon sampling for a branch if the branch can be observed in a gene tree of the locus, but not all loci have the potential to inform every branch of a concatenated tree inference if taxa missing from any partitions of the multilocus alignment (Hinchliff and Smith 2014). Two steps were used to compute the number of loci with decisive taxon sampling on branches. First we used ape to prune the best ML tree to taxa sampled in each locus partition (Paradis et al. 2017). Then we used the exact version of ASTRAL-III (-t 2 -x) to compute the number of loci with decisive taxon sampling on each branch, scored as the total number of quartet trees in all locus-pruned trees concordant with the branch from the best ML tree (Zhang et al. 2018).

1.3.4 Species Tree

To assess gene tree discordance arising from incomplete lineage sorting (ILS), individual gene trees were summarized into a species tree under the multispecies coalescent (MSC) model with ASTRAL-III (Zhang et al. 2018). ASTRAL-III estimates a species tree from gene trees by finding the species tree, within a restricted set of topologies, that maximizes the number of quartet trees shared with gene trees. Gene tree discordance explained by ILS and the MSC model assumes gene trees were estimated without error, therefore in order to mitigate gene tree estimation errors, branches of gene trees with less than ten percent bootstrap support were collapsed into polytomies with Newick Utilities (Junier and Zdobnov 2010). Contracted gene trees were summarized into a MSC tree with ASTRAL-III (-t 2; Zhang et al. 2018), support on branches of the MSC tree was measured with local posterior probability (Sayyari and Mirarab 2016), and the depth of gene tree coverage was summarized on branches of the MSC tree as the effective number of genes (Sayyari and Mirarab 2018). The effective number of genes measures both incomplete taxon sampling among gene trees, and lack of resolution within gene trees, because missing taxa and multifurcating branches in gene tree quartet sets do not inform branches of the MSC tree (Sayyari and Mirarab 2018). Local posterior probability (LPP) is a function of the number of gene trees analyzed and the quartet frequencies of a branch in the species tree. Values of 0.95 or more indicate very high confidence in a branch, and LPP = 0.7 can be considered as a minimum threshold for confidence in a branch (Sayyari and Mirarab 2016).

In order to match taxa present in the concatenated inferences and facilitate comparisons among trees, rogue taxa previously identified from dropsets of the final ML bootstraps were pruned from the MSC tree with ape (Paradis et al. 2017). Plots of the complete ML, MRC, and

MSC trees annotated with branch support and locus coverage, the ML tree with lineages represented by single taxa, and constrained and unconstrained ML trees with families represented by a single taxon were prepared using treeio and ggtree packages in R (Yu et al. 2017). Multilocus coverage, proportion of generic taxon sampling, and branch support summary statistics were computed for each lineage with tidyverse (Wickham 2017).

1.3.5 Topology Tests

Variation in log-likelihood supporting either the sister clade (Fig. 3A) or sequential clade (Fig. 3B) root topology was quantified, and tested for differences in scores between conflicting root topologies at the levels of individual alignment sites and locus partitions. Verbose splits output of RAxML IC scores was examined to identify competing root topologies and their gene support frequencies (GSF) observed in the bootset, and the phylogeny search was replicated to constrain the inference to the minority root topology using RAxML (-g; Stamatakis 2014). A likelihood ratio test (LRT) of the best ML and constrained ML trees was conducted with RAxML (-f H; SH-test of Shimodaira and Hasegawa 1999) to test the null hypothesis of no difference in log-likelihood scores among alternatively-rooted ML trees. Site-wise log-likelihood scores (SLS) for the best ML and constrained tree were computed and written to TREE-PUZZLE format with RAxML (-f G), and SLS were input to CONSEL to conduct Shimodaira's (2002) approximately unbiased (AU) test on alternative root topologies (Shimodaira and Hasegawa 2001). Differences in site-wise log-likelihood scores (∆SLS) between alternative root topologies were computed and summarized into gene-wise log-likelihood scores (∆GLS; Shen et al. 2016), and ∆SLS and rank-ordered ∆GLS were plotted using the tidyverse and cowplot packages in R (Wickham 2017; Wilke 2016).

1.4 Results

1.4.1 Alignments

Sequences of 19 nuclear loci (18 protein coding, 1 ribosomal) and 4 protein-coding mitochondrial genes were selected from data downloaded from GenBank (Table 1). Residue overlap threshold settings for filtering spurious sequences ranged from 0.05 to 0.25 for nuclear loci, from 0.50 to 0.95 for mitochondrial loci, and sequence overlap threshold settings for all loci ranged from 85 to 95 percent (Tab. 1). A total of 2,124 spurious sequences were filtered from preliminary multi-individual alignments with trimAl. Over half of the total spurious sequences removed (*n* = 1,176) were from the *cytb* alignment. More than 100 spurious sequences were also removed from *rag1* (*n* = 339), *co1* (*n* = 239), and *nd2* (*n* = 168) alignments.

1.4.2 Gene Trees

One hundred and thirty sequences representing 92 taxa were pruned from back-translated, trimmed, single-individual alignments. The pruned sequences contained 41 rogue taxa identified in RAxML dropsets among the 23 gene trees, and 51 taxa that were suspected of technical errors or sample misidentification because they violated monophyly at or above the family level. Final taxon sampling ranged from 17 taxa in *svep1*, to 401 taxa in *rag1*, with a mean of approximately 144 taxa sampled per gene alignment (Tab. 1).

1.4.3 Concatenated Trees

Twenty four rogue taxa identified in dropsets of the initial 250 rapid bootstraps were pruned from the concatenated alignment. The final rogue-filtered concatenated alignment consisted of 827 taxa and 18,065 characters, and was composed of 79.4 percent gaps and missing characters. Mean locus counts among ingroup lineages ranged from 1.7 loci per taxon

in the *Gobiodon* lineage (65 taxa), to 12.5 loci per taxon in *Kurtus* (2 taxa), with an average count of 4 loci per taxon over the ingroup and outgroups (Fig. 4).

Measures of branch support were similarly distributed across internal branches of the ML and MRC trees (Fig. 5A, Fig. 5B). On the NNI-optimized ML tree, mean bootstrap support was 75.6, with 65 percent of branches supported by scores greater than 70, and mean SH-like support was 0.82, with 67 percent of branches supported by scores greater than 0.85 (Fig. 5A). On the MRC tree, mean gene support frequency was 0.87, with 66 percent of branches supported by GSF greater than 0.85, and mean internode certainty was 0.69, with 58 percent of branches supported by IC greater than or equal to 0.7 (Fig. 5B; Fig. 6).

Patterns in the depth of locus coverage were similar across internal branches of the ML and MRC trees. On the NNI-optimized ML tree, 50 percent of the internal branches were covered by 2 or fewer loci with decisive taxon sampling, and 75 percent of internal branches were covered by 3 or fewer loci with decisive taxon sampling (Fig. 7). On the MRC tree, 49 percent of internal branches were covered by 2 or fewer loci with decisive taxon sampling, and 73 percent of internal branches were covered by 3 or fewer loci with decisive taxon sampling.

The only confidence measure correlated with the depth of locus coverage was SH-like branch support (r^2 = 0.315, F = 8.29, df = 18, p = 0.01; Fig. 8). No correlation to the number of loci with decisive taxon sampling was found for bootstrap support (r^2 = 0.024, F = 0.44, df = 18, p = 0.52), gene support frequency (r^2 = 0.002, F = 0.03, df = 18, p = 0.86), or internode certainty (r^2 < 0.001, F < 0.01, df = 18, p = 0.95).

1.4.4 Species Tree

Mean local posterior probability was 69 percent (LPP = 0.69) on branches of the MSC tree (Fig. 5C). Forty two percent of MSC tree branches were supported by LPP greater than the

minimum threshold of 70 percent, and 21 percent of branches exceeded the “very high” threshold of 95 percent LPP (Sayyari and Mirarab 2016). The mean effective number of genes was 2.8, with 52 percent of branches represented by 2 or fewer effective genes, and 74 percent of branches represented by 3 or fewer effective genes (Fig. 5C). Mean local posterior probability was not related to the effective number of genes in the MSC tree (r^2 = 0.068, F = 1.31, df = 18, p = 0.27).

1.4.5 Topology Tests

Verbose output of RAxML IC scores from 200 rapid bootstraps was parsed to identify split topologies at the root of Gobiaria and their gene support frequencies corresponding to two conflicting phylogenetic hypotheses (Fig. 3). The majority root topology observed in the MRC tree resolved families Kurtidae and Apogonidae as sister taxa (Order Kurtiformes; Fig. 3A) with a gene support frequency of 56 percent (GSF = 0.56). The alternative, minority root topology was observed with a gene support frequency of 42 percent (GSF = 0.42), with families Kurtidae and Apogonidae placed as successive sister taxa to the remainder of Gobiaria (Order Gobiiformes; Fig. 3B). The equivocal difference in gene support frequencies between conflicting root topologies was expressed in the near-absence of internode certainty computed on the branch subtending Order Kurtiformes (IC = 0.01; Fig. 5B).

A maximum likelihood tree search was constrained to the minority root topology, site-wise log-likelihood scores between the *ad hoc* inference and best ML trees were compared, and minor differences in relative support for conflicting root topologies were found. Approximately 52 percent of individual SLS supported the majority root topology (9,348 of 18,065 sites), whereas SLS supported the minority root topology in the remaining 48 percent of sites (8,717 of 18,065 sites; Fig. 9).

Phylogenetic signal in support of the majority root topology was evident after computing gene-wise log-likelihood scores, with positive ΔGLS recovered from 19 of the 23 loci analyzed (Fig. 10). Of the four loci supporting the minority root topology, only three had ΔGLS greater than one log-likelihood unit (*tbr1*, *myh6*, and *nd1*), while all loci supporting the majority topology had ΔGLS greater than one log-likelihood unit. The only locus with ΔGLS less than one was *hoxc6a* (ΔGLS = -0.27), indicating a non-significant difference in phylogenetic signal of the gene supporting either root topology.

Two hypothesis tests provided additional evidence to support the majority root topology. A SH-test in RAxML rejected the null hypothesis of no difference in log-likelihood scores between alternatively rooted trees with an observed difference of -267.5 log-likelihood units ($p < 0.01$). The AU-test conducted with CONSEL was also significant ($p < 0.001$), and rejected the null hypothesis of no difference in site-wise log-likelihood scores supporting either root topology (Fig. 9).

1.5 Discussion

This study presents a comprehensive analysis of publicly available multilocus sequence data to infer the phylogeny of Gobiaria. A custom pipeline allowed sampling of ambiguously identified taxa and selection of best quality sequences from species represented by multiple individuals. Data-driven techniques were applied to strengthen phylogenetic signal of single gene and concatenated alignments, and multiple tree construction methods were used to assess the robustness of systematic relationships. The phylogenetic hypothesis presented here includes approximately 32 percent of species, 68 percent of genera, and all higher taxa of Gobiaria, and is the first densely-sampled tree to resolve the diverse family Gobiidae without use of topological constraints (Rabosky et al. 2018). We did not find any notable differences with respect to phylogenetic arrangements presented by previous gobiarian-specific studies. Our results confirm uncertainty at the root of Gobiaria, and along the backbone branches subtending Eleotridae and Gobiidae taxa, and provide the most complete appraisal yet of the phylogenetic systematics of this diverse radiation of fishes.

1.5.1 Root Topology

Results from three phylogeny construction methods indicated topological instability at the root of Gobiaria (Fig. 5). The rooting topologies in best ML and bootstrap MRC trees matched Thacker et al.'s (2015) clade-based phylogenetic classification at higher taxonomic ranks, although statistical confidence in the observed branches were low (Fig. 5A; Fig. 5B). SH-like support on the branch subtending Kurtidae and Apogonidae in the best ML tree was 25 percent (SH-like = 0.25; Fig. 5A), gene support frequency on the same branch recovered in the bootstrap MRC tree was six percent above the majority-rule threshold (GSF = 0.56; Fig. 5B), and internode certainty was one percent (IC = 0.01; Fig. 5B). The near-absence of IC on the

root of Kurtiformes indicated conflicting phylogenetic signals in support of the observed MRC topology (Fig. 6).

The sister clade relationship of Kurtidae and Apogonidae inferred in best ML and bootstrap MRC trees, which previously informed a major taxonomic revision splitting Gobiaria into Kurtiformes and Gobiiformes (Thacker et al. 2015), was not robust to the MSC species tree summary of 23 gene trees (Fig. 5C). The unstable root topology of Gobiaria was apparent with recovery of the conflicting sequential clade relationship in the MSC tree, where Kurtidae and Apogonidae resolved as successive sister taxa to the remainder of Gobiaria, with 84 percent local posterior probability on the root (LPP = 0.84; Fig. 5C). The sequential root topology observed in the MSC tree matches the topology of recent phylogenomic trees (Li et al. 2017; Alfaro et al. 2018; Kuang et al. 2018) and was also evident as the minority root topology in 200 bootstrap trees (Fig. 10). Nevertheless, 83 percent of ∆GLS favored the sister group relationship (Fig. 10), as did the SH-test and AU-test on total and site-wise log-likelihoods (Fig. 9), respectively.

1.5.2 Cardinalfish Tribes

Within Apogonidae, most of the cardinalfish tribal lineages delineated by Mabuchi et al. (2014) with morphology, mitochondrial DNA, and two nuclear markers (*rag1*, *enc1*) were resolved. However, poor SH-like support (SH-like = 0.25) and low internode certainty (IC = 0.22) were found on the backbone branch separating the cardinalfish subfamily Apogoninae from earlier-diverging subfamilies Amioidinae and Pseudaminae. Backbone support within the subfamily Apogoninae was similarly weak but in accordance with Mabuchi et al.'s (2014) recovery of earliest diverging *Apogonichthys* (*Foa*, *Fowleria, Apogonichthys perdix*; tribe Apogonichthyini), *Gymnapogon* (*Gymnapogon*, *Cercamia*, *Lachneratus*; tribe Gymnapogonini),

and *Apogon* (*Apogon*, *Zapogon*, *Phaeoptyx*, *Astrapogon*; tribe Apogonini) lineages. The *Glossamia* lineage (tribe Glossamiini), including the genera *Glossamia* and *Yarica*, is supported by morphology and is unique among cardinalfishes because it inhabits freshwater environments. Our results confirm the previous molecular hypothesis that found the two genera *Yarica* and *Glossamia* split apart, proximate to the species-rich *Apogon* and *Cheilodipterus* lineages, respectively (tribes Apogonini and Cheilodipterini in Mabuchi et al. 2014). The remainder of the diversity within subfamily Apogoninae resolves into a series of clades largely consistent with generic boundaries, with lineages containing bioluminescent genera *Jaydia*, *Taeniamia*, *Rhabdamia*, *Verulux*, and *Siphamia* (tribes Sphaeramiini, Archamiini, Rhabdamiini, Veruluxini, and Siphamiini) distributed throughout the topology, concordant with the results of (Thacker and Roje 2009). The distinctiveness of three subgroups of *Ostorhinchus* depicted in Mabuchi et al. (2014) was confirmed, with a clade of *Ostorhinchus rueppellii*, *O. ishigakiensis*, and *O. hoevenii* placed within the *Sphaeramia* lineage (tribe Sphaeramiini; the Barred group of Mabuchi et al. 2006), and *O. margaritophorus* recovered with *Rhabdamia*, *Fibramia,* and *Zoramia* (tribes Zoramiini and Rhabdamiini). The unstable placement of *O. margaritophorus*, as well as *Pterapogon* and *Vincentia,* was also confirmed. *Vincentia* was identified in RAxML dropsets and subsequently pruned from all three trees. *Pterapogon* was recovered near the base of the clade containing the bulk of the *Ostorhinchus* species (tribe Ostorhinchini; the Striped group of Mabuchi et al. 2006).

1.5.3 Gobioid Families

The placement of Trichonotidae as sister to Gobioidei was robust to different phylogeny construction methods, with strong support estimated on branches subtending the clade recovery in all trees (Fig. 5). SH-like branch support in the best ML tree was complete (SH-like = 1.0), as

was local posterior probability measured in the MSC tree (LPP = 1.0). Gene support frequency in the MRC tree was high (GSF = 0.96), and internode certainty indicated a clear phylogenetic signal supporting the branch (IC = 0.89). *Trichonotus* was only recently identified as the sister taxon to gobioids by Thacker et al. (2015) with broad sampling and ten nuclear loci (*enc1*, *glyt*, *myh6*, *plagl2*, *ptr*, *rag1*, *sh3px3*, *sreb2*, *tbr1*, *zic1*. All of our results confirm the systematic placement of Trichonotidae as sister to goboids (Thacker et al. 2015).

The varied phylogenetic inference methods and analysis of branch support statistics from different trees revealed unexpected instability within Eleotridae, and between Eleotridae and Odontobutidae, Milyeringidae, and Rhyacichthyidae (Fig. 5B; Fig. 5C). Results from the best ML tree indicated these taxa were monophyletic and strongly supported (Fig. 5A). For example, SH-like support on branches subtending crown groups ranged from a low of 89 percent for Odontobutidae, to 97 percent for both Milyeringidae and Eleotridae, to 100 percent for Rhyacichthyidae, and SH-like support on backbone branches subtending these clades was complete (SH-like = 1.0). However, computation of the MRC tree revealed relatively low support on the branch subtending Odontobutidae (IC = 0.09, GSF = 0.61), Eleotridae (IC = 0.43, GSF = 0.73), and Milyeringidae (IC = 0.44, GSF = 0.78), with similarly weak support computed on the backbone branches subtending the clades (Fig. 6). For instance, internode certainty was 30 percent (IC = 0.30) on the backbone branch subtending Rhyacichthyidae and Odontobutidae (Fig. 6). Furthermore, monophyly of Eleotridae and Odontobutidae in the MSC tree was violated by *Xenisthmus* and *Micropercops*, respectively, and Odontobutidae resolved as a sister clade to Milyeringidae in the MSC tree with very high local posterior probability (LPP = 0.96; Fig. 5C). The outlier genera *Xenisthmus* and *Micropercops* were both represented by a single taxon and mitochondrial DNA sequences in our supermatrix, which may explain the systematic

discrepancies, although the varying position of *Xenisthmus* among ML and MSC trees warrants further study.

Eleotridae, as in previous studies based on mitochondrial markers (Thacker and Hardman 2005; Thacker 2017), was comprised of several clades with poor support on the backbone branches. We recover clades including *Ratsirakia*, *Tateurndina*, *Xenisthmus*, *Giurus* and *Mogurnda*; *Gobiomorus*, *Hemieleotris*, *Dormitator*, and *Guavina*; *Gobiomorphus*, *Philypnodon*, *Calumia* and *Microphilypnus*; and separate lineages for each of the genera *Hypseleotris*, *Eleotris*, and *Leptophilypnus*. These results accord partially with the earlier studies, in that similar groups are identified, but their interrelationships are completely uncertain. In particular, the placements of Indo-West Pacific *Hypseleotris* and *Philypnodon*, as well as the dwarf neotropical genera *Microphilypnus* and *Leptophilypnus* remain questionable. In contrast, we provide strong support for Butidae and the relationships of the genera within it, and confirm its placement as sister to Thalasseleotrididae plus Gobionellidae and Gobiidae in all trees (Fig. 5).

1.5.4 Goby Lineages

Nineteen monophyletic goby lineages were resolved, with two outliers that were previously noted for rogue taxon behavior in several studies (Fig. 4; Thacker and Roje 2011; Thacker 2013; Agorreta et al. 2013; Tornabene et al. 2013; Thacker 2015). Among the five lineages of Gobionellidae, we recovered well-supported groupings concordant with the results of Agorreta et al. (2013), resolving the *Periophthalmus* lineage sister to the *Stenogobius* lineage, reciprocal to a clade composed of the *Pomatoschistus, Mugilogobius*, and *Acanthogobius* lineages (Fig. 4). Relationships among the 14 lineages of Gobiidae were generally poorly supported, and were not robust to different tree construction methods, although a few clades

appeared evident. Similar to trees inferred by Agorreta et al. (2013), Thacker and Roje (2011); Thacker (2015), and Tornabene et al. (2013), we resolved clades grouping the *Gobius* and *Gobiosoma* lineages, the *Asterropteryx* and *Lophogobius* lineages, the *Aphia* and *Valenciennea* lineages, and the *Glossogobius*, *Kraemeria*, *Cryptocentrus* and *Gobiopsis* lineages (Fig. 4). Placements of the *Callogobius*, *Gunnellichthys*, *Priolepis* and *Gobiodon* lineages were more variable in our hypothesis (Fig. 4; Fig. 6), as they were in previous trees, and remain to be confidently resolved.

Within lineages of both Gobionellidae and Gobiidae, most of the genera were recovered as monophyletic, but a few of the inter- and intra-generic relationships were questionable. The gobionellid genera *Taenioides*, *Oligolepis*, *Oxyurichthys*, *Ctenogobius*, *Stenogobius*, *Stiphodon*, *Sicydium*, *Eugnathogobius*, and *Pomatoschistus* all resolved as paraphyletic, but generally in situations of shallow resolution and poor support, and likely due to insufficient data, as discussed below. The gobiid genera *Cryptocentrus*, *Acentrogobius*, *Drombus*, *Bryaninops*, *Trimma*, *Gobius*, and *Tigrigobius* were also variously paraphyletic, but we suspect that in most of these cases, the relationships may reflect that actual taxonomic adjustments are needed.

We highlight two outliers of putative goby lineages that warrant further study. *Schismatogobius* resolved as sister to *Kraemeria* (Gobiidae), although it is currently classified in Gobionellidae. *Schismatogobius* was not placed into a suprageneric group in Birdsong et al. (1988) global survey of axial skeletal morphology, but its shape (slender elongate body) and ecology (inhabits freshwater) to *Gobionellus* were used to associate *Schismatogobius* with Gobionellidae. *Schismatogobius* was included in the *Stenogobius* lineage by Larson (2001), but two recent molecular phylogenetics studies have resolved *Schismatogobius* among Gobiidae, either as sister to the *Glossogobius* lineage with two nuclear genes (*rag1*, *rhod*) sequenced from 75 taxa (Tornabene et al. 2013), or as sister to *Kraemeria* with whole mitochondrial genomes

sequenced from 40 taxa (Maeda et al. 2018). *Schismatogobius* resolved with the *Kraemeria* lineage with strong support in the best ML and bootstrap MRC trees (Fig. 4), but was split between the *Kraemeria* and *Gobiodon* lineages in the MSC tree (Fig. 6). The question of whether *Schismatogobius* is properly included in Gobiidae, which are mostly marine, or if its placement is influenced by the presence of the rogue taxon *Kraemeria* remains to be answered (Agorreta et al. 2013; Tornabene et al. 2017). Some aspects of the morphology, ecology, and behavior of *Schismatogobius* are similar to *Kraemeria* (Tsubaki and Kato 2009). Both taxa are characterized by a slender body with no scales, a narrow habitat preference associated with the intertidal (*Schismatogobius* in streams just above, *Kraemeria* in swash zone just below intertidal), and a behavior of darting into coarse-grained, sandy substrate when alarmed (Kottelat and Pethiyagoda 1989), although these characters are found in other gobies and therefore not unique to *Schismatogobius* or the *Kraemeria* lineage.

The paedomorphic genus *Schindleria* resolved by itself in all trees (Fig. 4), but the position of *Schindleria* relative to other goby lineages was variable; as sister to the *Gobiodon* lineage with 100 percent SH-like support in the best ML tree, as part of an unresolved polytomy shared with the *Gunnellichthys*, *Priolepis*, *Gobius*, *Gobiosoma, Aphia* and *Valenciennea* lineages in the bootstrap MRC tree (Fig. 6), or sequential between the *Aphia* and *Priolepis* lineages with low local posterior probability (LPP = 0.26) in the MSC tree. Due to reductions in the skeletal and soft tissue structures of *Schindleria*, this goby has traditionally been difficult to classify with morphology (Johnson and Brothers 1993). *Schindleria* was identified as a rogue taxon in the gobioid molecular phylogeny of (Agorreta et al. 2013), and further questioned by Tornabene et al. (2017), who reanalyzed their matrix with additional taxon sampling and documented instability of *Schindleria* among the *Gunnellichthys* and *Gobiodon* lineages when *Kraemeria* was included. In both of those hypotheses, the branches subtending *Schindleria*

were notably long, and it is likely that *Schindleria* is distinct genetically as well as morphologically, to the point that its relationships may simply be unresolvable with the data at hand. Phylogenomic datasets with fewer missing data may provide more robust resolution for *Schindleria* as well as the other questionably resolved taxa in our hypothesis. However, we note that the majority of genera we sampled were resolved as monophyletic, and all except *Schismatogobius* were assigned to their expected lineage. Overall, we recovered the expected topology for interlineage relationships in Gobionellidae, but the broad relationships among Gobiidae lineages remain unresolved, and warrant further study.

1.5.5 Supermatrix Sparseness and Phylogenetic Resolution

The distribution of decisive loci across internal branches of the ML tree revealed that shallow-level relationships were disproportionately affected by missing data (Fig. 7). This type of pattern is expected when using a sparse supermatrix comprised of publicly available data to infer phylogeny of a diverse radiation, as it reflects the broad taxon sampling strategy used by past researchers who generated the original sequence data (Hinchliff and Smith 2014). Previous studies have not found any correlation of terminal locus sampling to both bootstrap and SH-like support, although those studies were compromised with use of pseudoreplicated data that violated assumptions of the statistical tests (Wiens 2006; Pyron et al. 2011). In contrast, (Hinchliff and Smith 2014) found a weak but significant relationship of an IC-variant (internode certainty considering all conflicting bipartitions; ICA) to the number of loci with decisive taxon sampling, using a subset of their data (5 < decisive loci < 25) that excluded the most extreme values of locus coverage. We found a strong and significant correlation of SH-like branch support to the number of loci with decisive taxon sampling (r^2 = 0.315, F = 8.29, df = 18,

p = 0.01), which provides evidence that supermatrix sparseness was related to reduced confidence in branches of the best ML tree (Fig. 8).

The ML and MRC trees resolved all goby lineages (Fig. 4; Fig 6), but resolution was lacking among genera within several lineages. We suspect that the generic paraphyly observed in our best ML tree can largely be explained by absence of decisive loci. The gobionellid genera recovered as non-monophyletic (detailed above) all lack decisive taxon sampling on subtending branches. Several of the questionably resolved gobiid genera are also without decisive loci on internal branches, but not all, and we suspect that in some of those cases, taxonomic adjustments are needed.

In addition to assessing the impact of ILS on phylogeny, the MSC tree provided another glimpse into the impact of missing data on shallow resolution within goby lineages. In contrast to the concatenated ML and MRC trees inferred with linked alignment partitions, the gene trees we summarized into the MSC tree were considered independent, and taxa with non-overlapping locus coverage did not necessarily cluster as they did when linked by concatenation. Therefore, broad resolution of goby lineages in the MSC tree was poor compared to ML and MRC trees, and genera without decisive locus coverage from the *Stenogobius* and *Gobiopsis* lineages were scattered throughout the MSC tree.

1.6. Conclusion

We present the most comprehensive gobiarian phylogeny to date. Our supermatrix was constructed with data-driven filtering and assembly of existing mitochondrial and nuclear sequence data, and our analyses yielded a phylogenetic hypothesis of Gobiaria that was in line with previous studies and existing classifications. We resolved the primary groups Kurtidae, Apogonidae, Trichonotidae plus Gobioidei, and all the sampled families and lineages. The phylogenetic trees we present are unstable at the root of Gobiaria, and resolve Kurtidae and Apogonidae as a sister clade (ML and MRC trees), or as sequential clades (MSC tree), with poor support on the root in every tree. Backbone relationships among families were strong for gobioids, but within the largest families Eleotridae and Gobiidae, interrelationships among lineages and subclades remained uncertain, characterized by poor resolution and short internodes. This pattern may be due to ILS stemming from an explosive burst of diversification (Alfaro et al. 2018; Alda et al. 2019), or to the inherent limitations of multilocus, single-gene sequencing data (Parker et al. 2019).

Our phylogenetic hypotheses provide a comprehensive summary of the molecular data available for Gobiaria, and support the current classification at the family and intra-familial lineage level. At the generic level, the majority of genera were resolved with expected relationships; the few exceptions were largely in areas of low data coverage, and may also reflect real taxonomic inaccuracies. Expanding existing phylogenomic matrices of ultraconserved element loci (UCEs) with additional taxon sampling could improve resolution, particularly along the backbone of the larger families and potentially at the root (Tornabene et al. 2013; Kuang et al. 2018). A larger dataset that samples widely across the genome may provide strong branch support even where internode distances are shallow, and hopefully yield the

robust lineage relationships that are crucial for understanding the evolution and biogeography of this diverse, cosmopolitan radiation of vertebrates.

1.7 Funding

Tyler McCraney was supported by a Rosemary Grant Award from the Society for the Study of Evolution, a Betty and EP Franklin Grant in Tropical Biology and Conservation from the Center for Tropical Research at UCLA, and Research and Travel Awards from the Department of Ecology and Evolutionary Biology at UCLA.

1.8 Acknowledgements

We are grateful to Jonathan Chang and Francesco Santini for providing technical advice. Computational and storage services associated with the Hoffman2 Shared Cluster were provided by UCLA Institute for Digital Research and Education Research Technology Group.

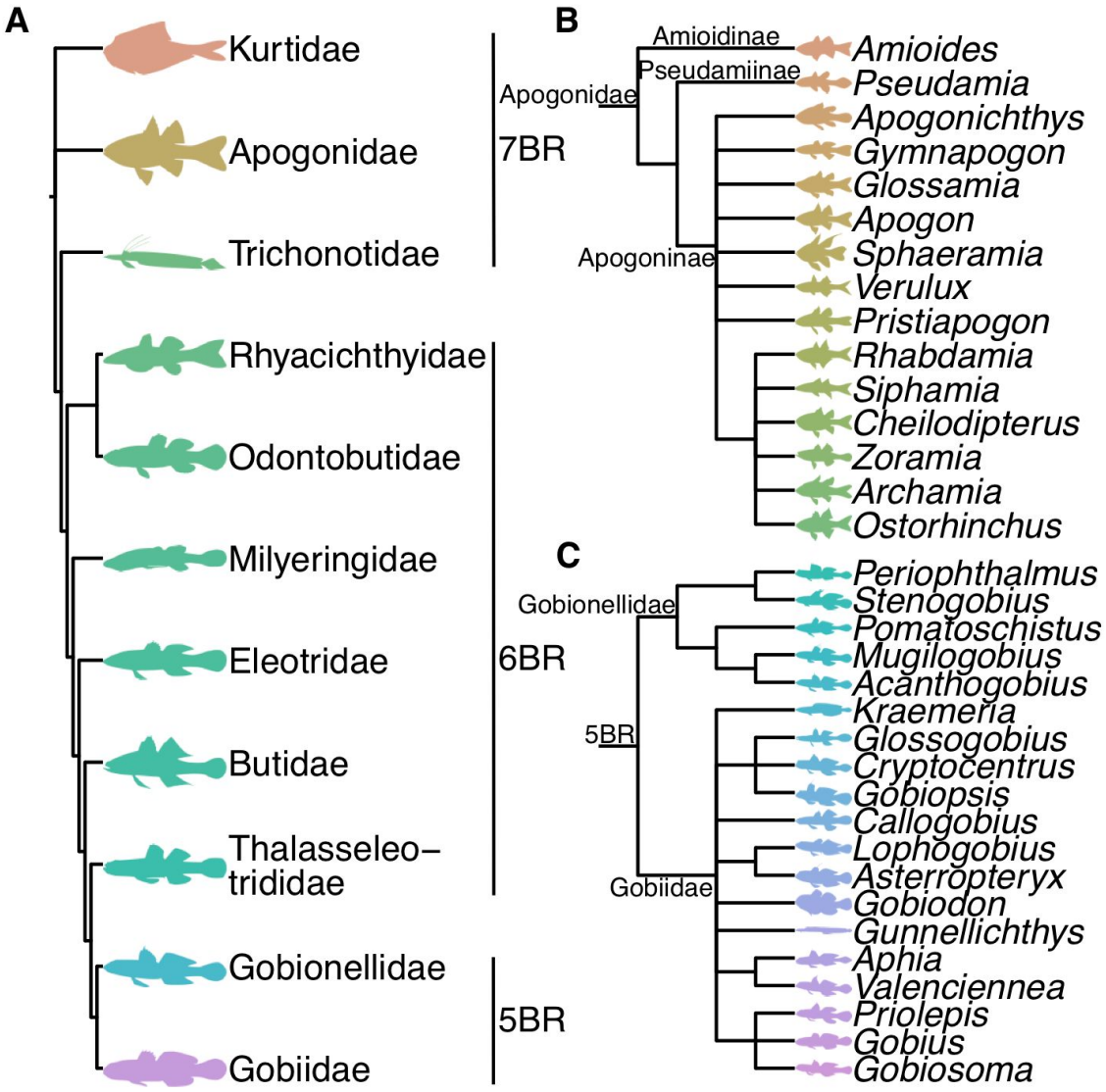


Fig. 1. Consensus trees of (A) 11 gobiarian families labeled with number of branchiostegal rays (BR), (B) 15 cardinalfish lineages, and (C) 19 goby lineages; derived from Thacker (2009), Agorreta et al. (2013), Tornabene et al. (2013), Mabuchi et al. (2014), Thacker (2015), Thacker et al. (2015), Li et al. (2017), Alfaro et al. (2018), and Kuang et al. (2018).

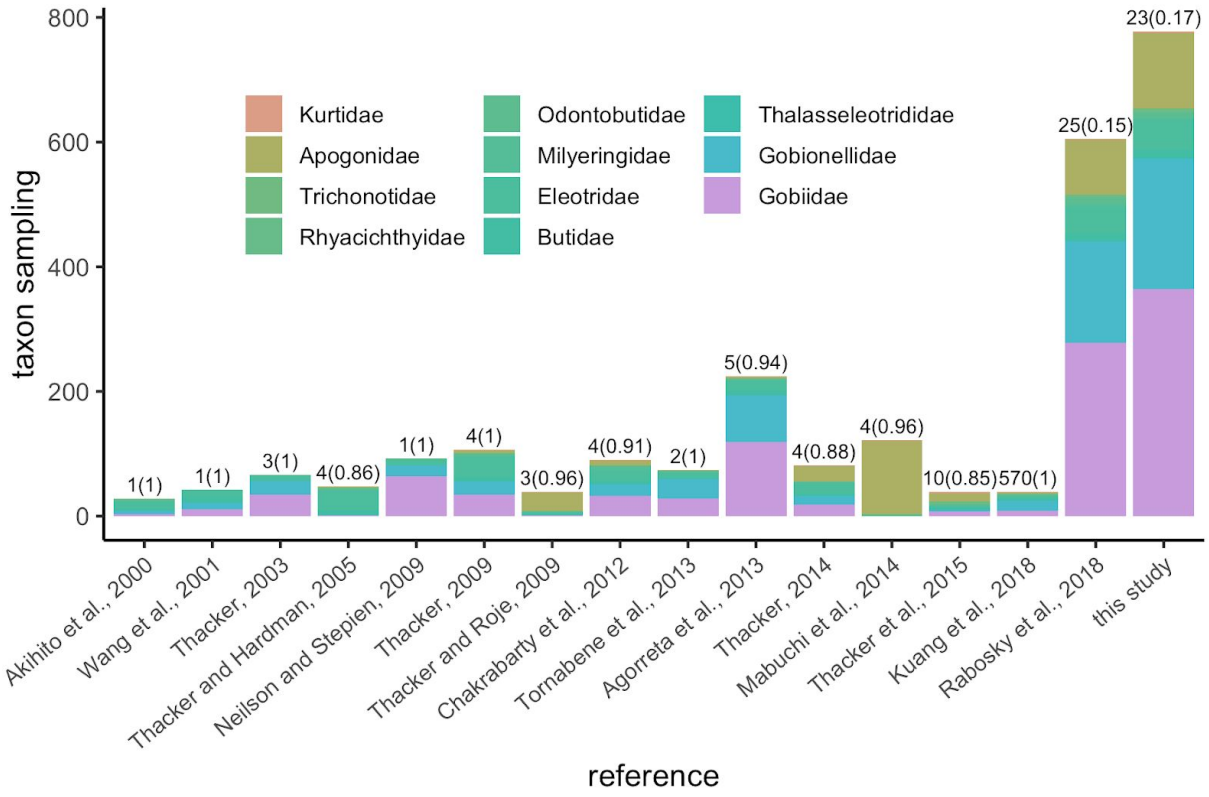


Fig. 2. Taxon sampling of previous molecular phylogenetic studies of Gobiaria. Bars are labeled with the number of loci, and matrix completeness is in parentheses.

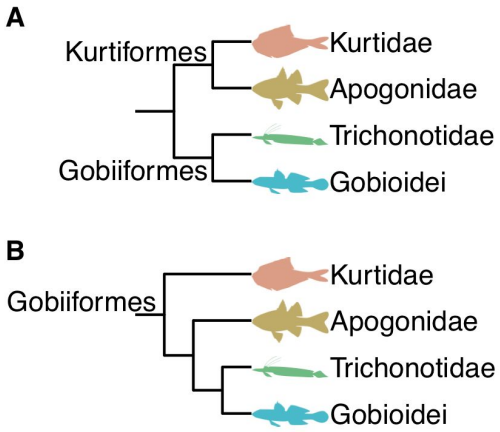


Fig. 3. Controversy in phylogenetic classifications of major groups within Gobiaria. The sister clade hypothesis (A) splits Kurtiformes from Gobiiformes and is supported by studies that used nuclear DNA sequences (Near et al. 2013; Thacker et al. 2015; Betancur-R et al. 2017). The sequential clade hypothesis (B) is supported by all phylogenomic studies (Li et al. 2017; Alfaro et al. 2018; Kuang et al. 2018) and earlier studies that used mitochondrial DNA sequences (Thacker 2009; Chakrabarty et al. 2012).

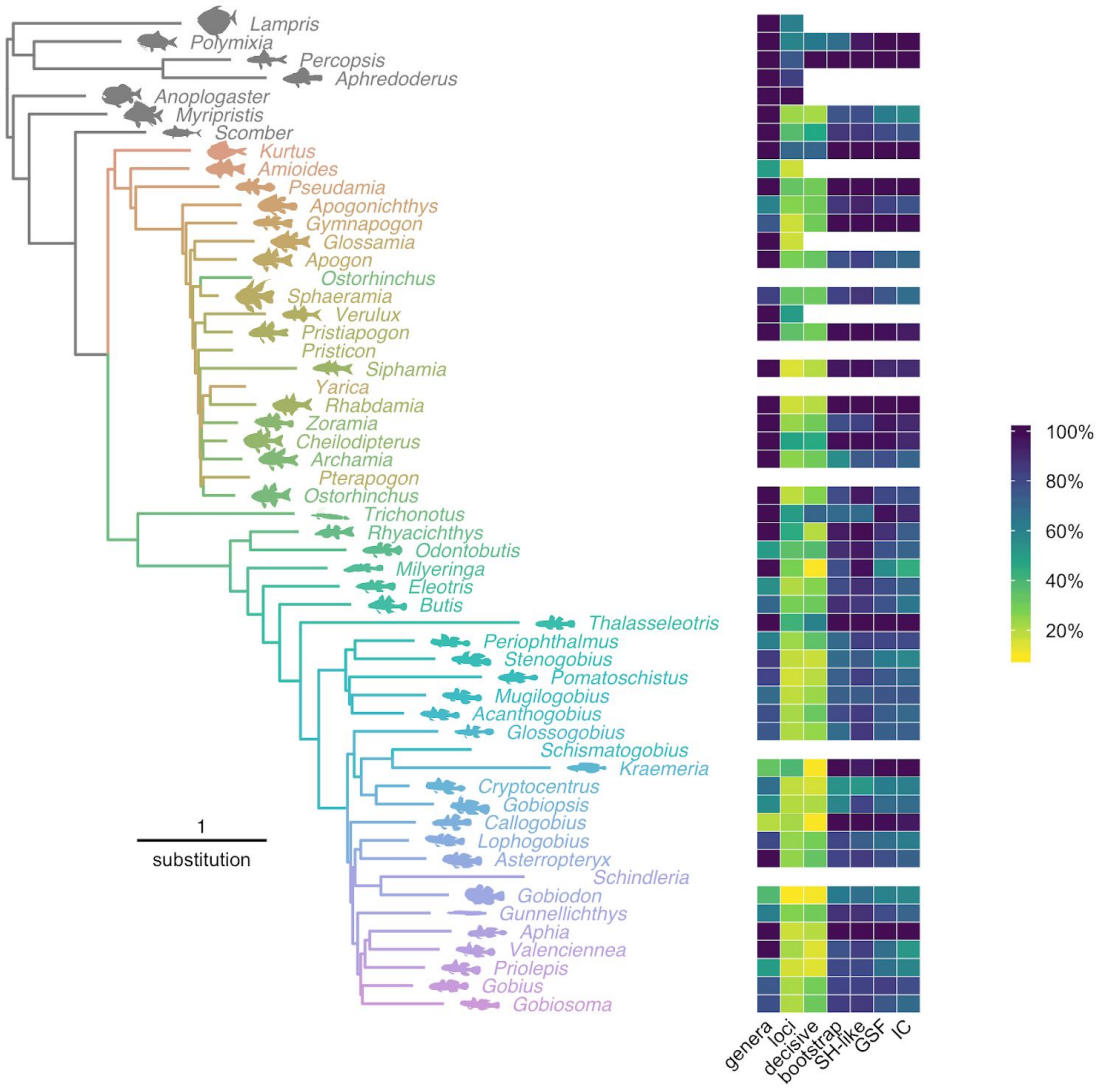


Fig. 4. ML hypothesis of gobiarian lineages inferred from 23 locus supermatrix. Lineage outliers are labeled by genus and shown without a silhouette image or accompanying data. Heatmap includes mean taxon sampling, relative locus coverage, and branch support by lineage, including proportions of sampled genera, loci, decisive loci, bootstrap support, SH-like branch support, and from the MRC tree, gene support frequency (GSF) and internode certainty (IC). Decisive loci and support values are missing from the heatmap in cases where a lineage was represented by a single taxon.

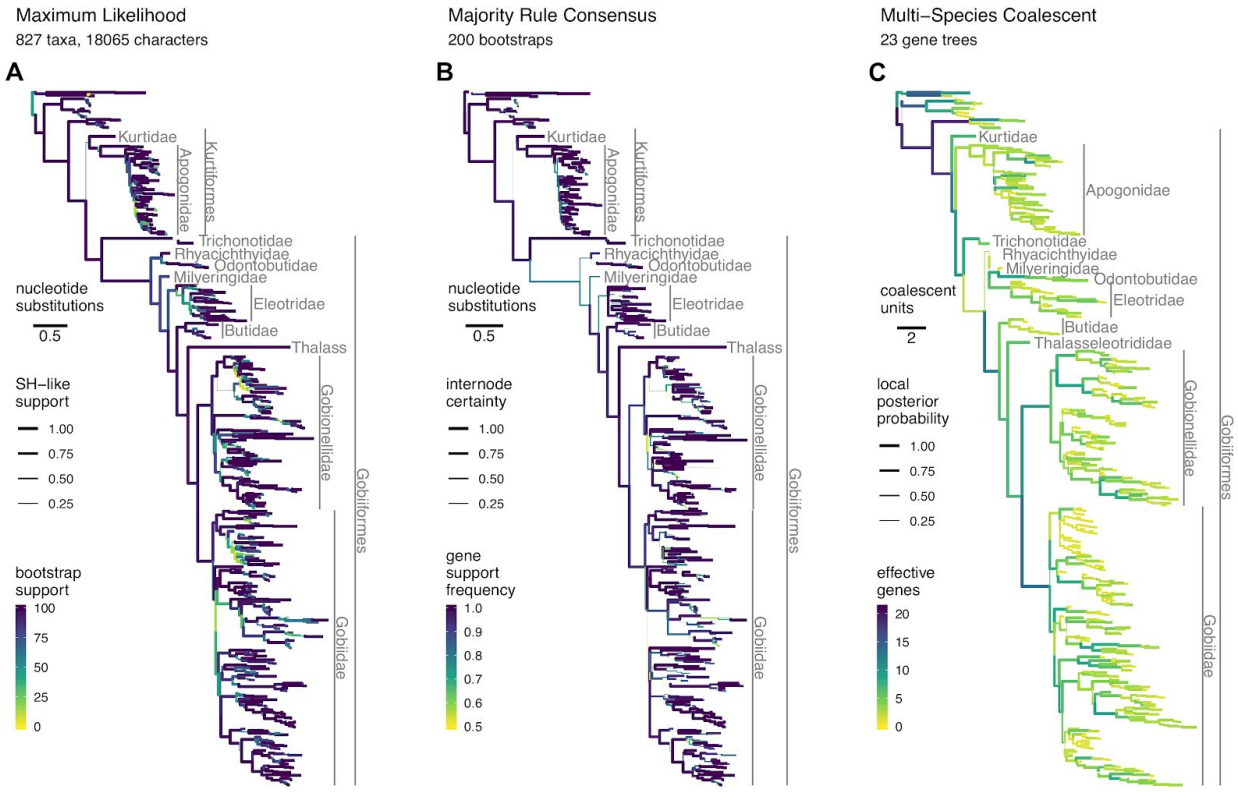


Fig. 5. Phylogenetic hypotheses of Gobiaria from a 23 locus supermatrix and three tree construction methods. Terminal nodes were pruned to display branch support.

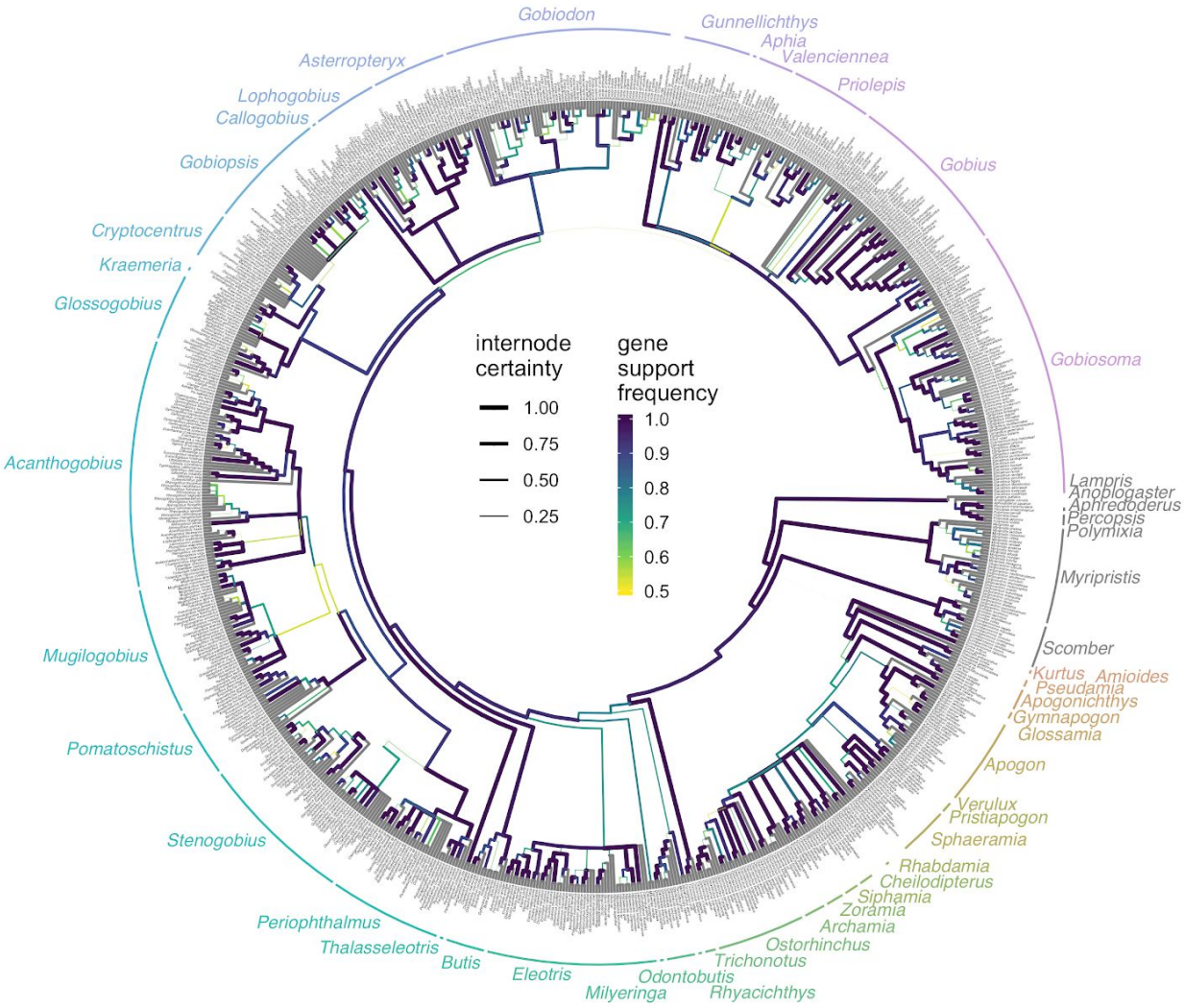


Fig. 6. Majority rule consensus hypothesis of gobiarian phylogeny constructed from 200 bootstrap ML inferences.

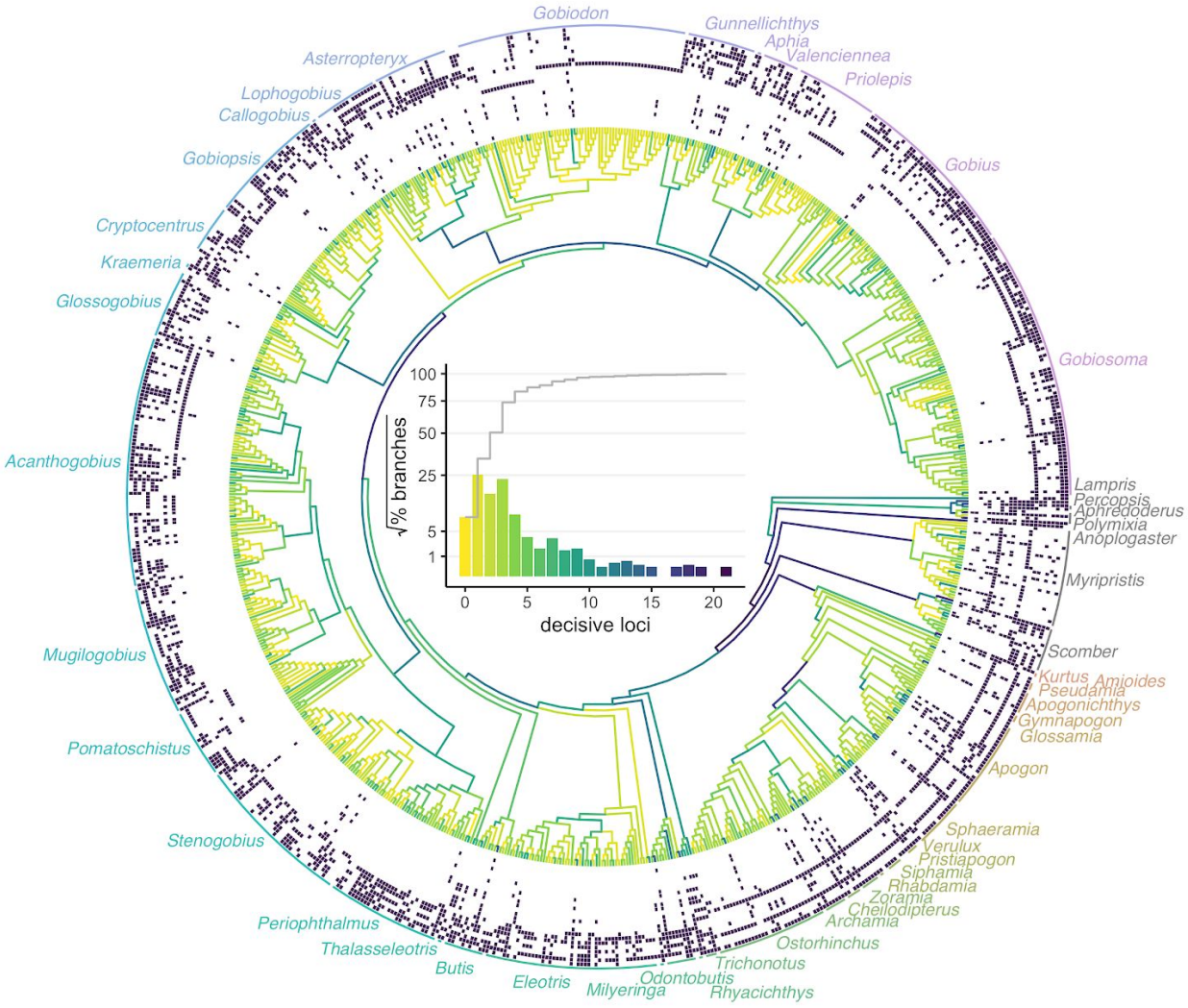


Fig. 7. Depth of locus coverage across the ML tree. The number of loci with decisive taxon sampling is shaded on internal branches. Terminal branches are shaded by number of loci present in the supermatrix. Inside the tree, the percent of internal branches at each number of decisive loci is plotted on a square root axis to display differences in small values, and the cumulative percent at each number of decisive loci is plotted with the gray step. Outside the tree, presence of sequence data at each locus is indicated by a filled cell, and missing sequences are indicated by white cells.

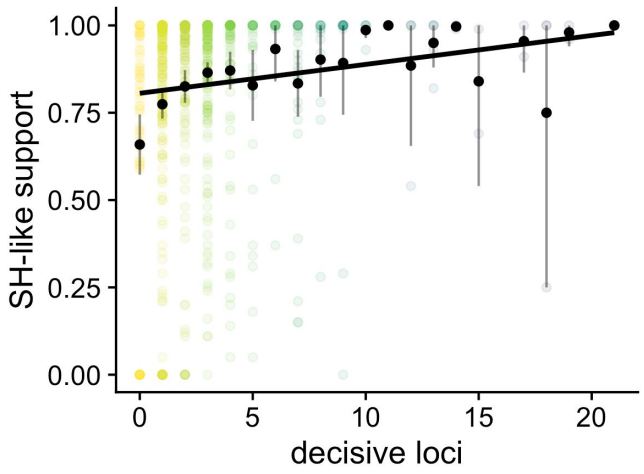


Fig. 8. Relationship of SH-like branch support and the number of loci with decisive taxon sampling. Shaded points represent individual branches of the ML tree. Black points represent mean SH-like branch support at the number of loci with decisive taxon sampling. Bars extend two standard errors above and below points, and represent approximate 95% confidence intervals for the means. Pearson's correlation coefficient (r = 0.561) was significantly different from zero (r^2 = 0.315, F = 8.29, df = 18, p = 0.01).

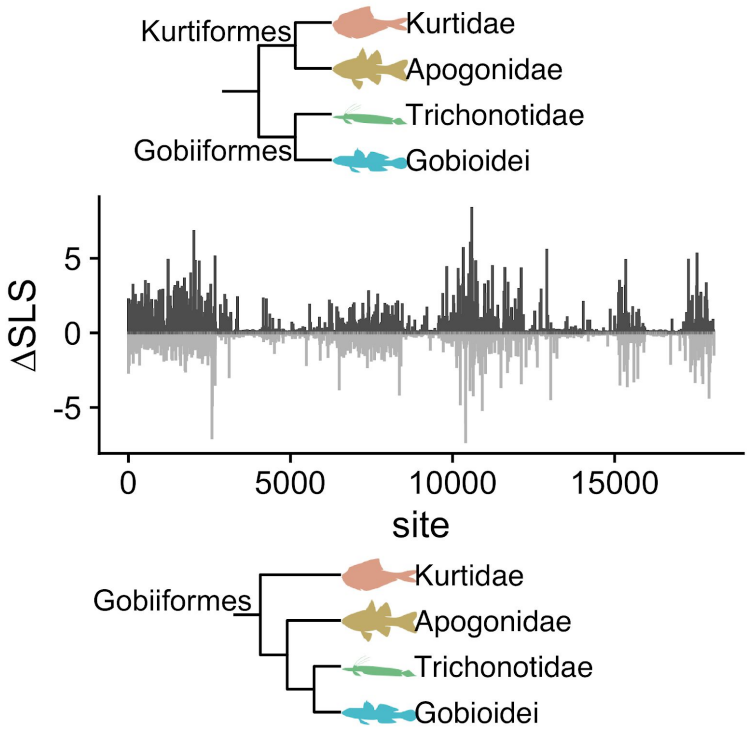


Fig. 9. Difference in site-wise log-likelihood scores (ΔSLS) between the best ML tree and the ML tree constrained to the minority root topology observed from bootstrapping the 18,065 site concatenated alignment. Positive ΔSLS values (dark shading) display phylogenetic signal supporting the best ML tree root topology that resolves nurseryfishes and cardinalfishes as sister taxa (Order Kurtiformes, Order Gobiiformes). Negative values (light shading) indicate phylogenetic signal supporting the minority root topology that resolves nurseryfishes and cardinalfishes as sequential sister taxa to the remainder of Gobiaria (Order Gobiiformes). Approximately 52 percent of individual ΔSLS supported the majority root topology (9,348 of 18,065 sites), while ΔSLS supported the minority root topology in the remaining 48 percent of sites (8,717 of 18,065).

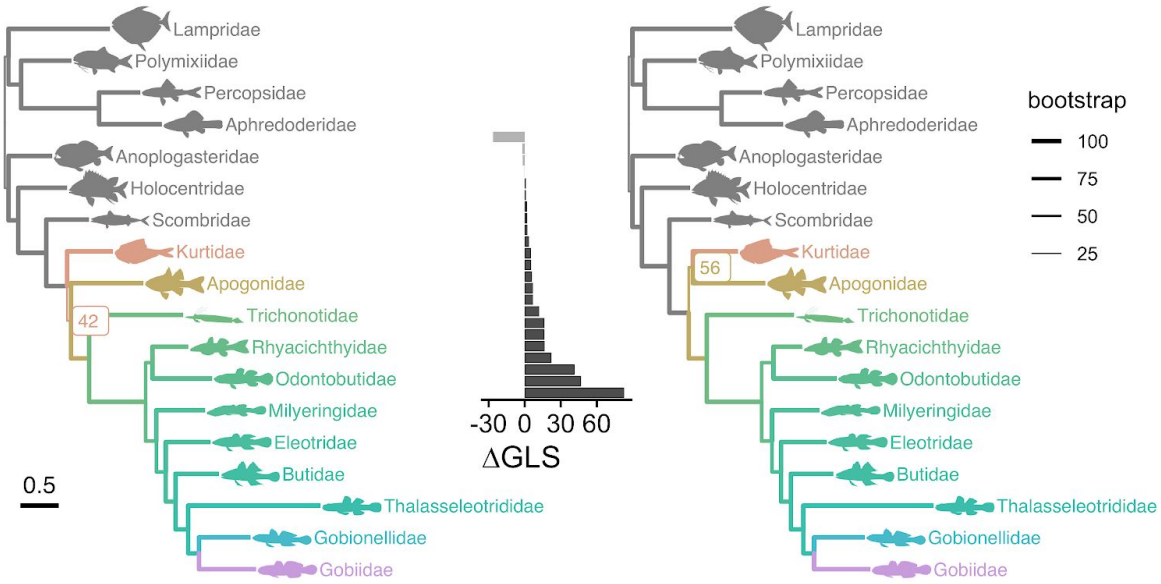


Fig. 10. Backbone trees of outgroups (grey) and gobiarian families inferred with constrained (left) and unconstrained (right) ML searches. Branch weight is used to display support from 200 bootstraps, and branches subtending alternative root topologies are labeled. Gene-wise log-likelihood scores (ΔGLS; center) measure the phylogenetic signal in support of conflicting root hypotheses.

Tab. 1. Alignment information of 23 loci used in this study.

Product	Locus	Reference	Sites	Taxa	Overlap threshold:	
					Residue	Sequence
cytochrome c oxidase subunit I	*co1*	Normark et al. (1991)	1545	282	0.65	85
cytochrome b	*cytb*	Kocher et al. (1989)	1140	394	0.50	95
ectodermal-neural cortex 1-like protein	*enc1*	Li et al. (2007)	759	129	0.05	95
Fic domain protein	*ficd*	Li et al. (2011)	675	28	0.05	90
glycosyltransferase	*glyt*	Li et al. (2007)	762	55	0.05	95
Homeo box C6a	*hoxc6a*	Betancur-R et al. (2013)	126	31	0.05	95
leucine-rich repeat and WD repeat-containing protein	*kiaa1239*	Li et al. (2011)	702	52	0.05	95
cardiac muscle myosin heavy chain 6 alpha	*myh6*	Li et al. (2007)	714	96	0.15	90
NADH dehydrogenase subunit 1	*nd1*	Thacker (2003)	969	256	0.95	95
NADH dehydrogenase subunit 2	*nd2*	Thacker (2003)	1026	236	0.90	90
pannexin 2	*panx2*	Broughton et al. (2013)	666	56	0.10	95
pleiomorphic adenoma protein-like 2 protein	*plagl2*	Li et al. (2007)	543	61	0.25	90
patched domain-containing 4	*ptr*	Li et al. (2007)	531	203	0.05	95
recombination activating gene 1	*rag1*	López et al. (2004)	1278	401	0.20	95
rhodopsin	*rhod*	Chen et al. (2003)	711	216	0.15	90
receptor-interacting serine-threonine kinase 4	*ripk4*	Li et al. (2011)	639	53	0.05	95
S7 ribosomal protein	*s7*	Chow and Hazama (1998)	687	79	NA	NA
SH3 and PX3 domain-containing 3-like protein	*sh3px3*	Li et al. (2007)	636	59	0.05	95
si:dkey-174m14.3	*sidkey*	Li et al. (2010)	978	24	0.05	90
brain super conserved receptor	*sreb2*	Li et al. (2007)	846	264	0.05	95
Sushi/von Willebrand factor type A/EGF/pentraxin domain-containing 1	*svep1*	Betancur-R et al. (2013)	693	17	0.05	90
T-box brain	*tbr1*	Li et al. (2007)	621	36	0.05	95
zic family member 1	*zic1*	Li et al. (2007)	819	285	0.10	95

1.9 References

Aberer, Andre J., Kassian Kobert, and Alexandros Stamatakis. 2014. “ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era.” *Molecular Biology and Evolution* 31 (10): 2553–56.

Aberer, Andre J., Denis Krompaß, and Alexandros Stamatakis. 2011. “RogueNaRok: An Efficient and Exact Algorithm for Rogue Taxon Identification.” *Heidelberg Institute for Theoretical Studies: Exelixis-RRDR-2011--10*. http://sco.h-its.org/exelixis/pubs/Exelixis-RRDR-2011-10.pdf.

Agorreta, Ainhoa, Diego San Mauro, Ulrich Schliewen, James L. Van Tassell, Marcelo Kovačić, Rafael Zardoya, and Lukas Rüber. 2013. “Molecular Phylogenetics of Gobioidei and Phylogenetic Placement of European Gobies.” *Molecular Phylogenetics and Evolution* 69 (3): 619–33.

Akihito, A. Iwata, T. Kobayashi, K. Ikeo, T. Imanishi, H. Ono, Y. Umehara, et al. 2000. “Evolutionary Aspects of Gobioid Fishes Based upon a Phylogenetic Analysis of Mitochondrial Cytochrome B Genes.” *Gene* 259 (1-2): 5–15.

Alda, Fernando, Victor A. Tagliacollo, Maxwell J. Bernt, Brandon T. Waltz, William B. Ludt, Brant C. Faircloth, Michael E. Alfaro, James S. Albert, and Prosanta Chakrabarty. 2019. “Resolving Deep Nodes in an Ancient Radiation of Neotropical Fishes in the Presence of Conflicting Signals from Incomplete Lineage Sorting.” *Systematic Biology* 68 (4): 573–93.

Alfaro, Michael E., Brant C. Faircloth, Richard C. Harrington, Laurie Sorenson, Matt Friedman, Christine E. Thacker, Carl H. Oliveros, David Černý, and Thomas J. Near. 2018. “Explosive Diversification of Marine Fishes at the Cretaceous–Palaeogene Boundary.” *Nature Ecology & Evolution*, March. https://doi.org/10.1038/s41559-018-0494-6.

Betancur-R., Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andrés

López, Chenhong Li, Nancy I. Holcroft, et al. 2013. “The Tree of Life and a New Classification of Bony FishesCall for SubmissionsAVAToL CollectionBlog PostTwitter.” *PLoS Currents*. https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, Ricardo, Edward O. Wiley, Gloria Arratia, Arturo Acero, Nicolas Bailly, Masaki Miya, Guillaume Lecointre, and Guillermo Ortí. 2017. “Phylogenetic Classification of Bony Fishes.” *BMC Evolutionary Biology* 17 (1): 162.

Birdsong, Ray S., Edward O. Murdy, and Frank L. Pezold. 1988. “A Study of the Vertebral Column and Median Fin Osteology in Gobioid Fishes with Comments on Gobioid Relationships.” *Bulletin of Marine Science* 42 (2): 174–214.

Boettiger, C., D. T. Lang, and P. C. Wainwright. 2012. “Rfishbase: Exploring, Manipulating and Visualizing FishBase Data from R.” *Journal of Fish Biology* 81 (6): 2030–39.

Borstein, Samuel R., and Brian C. O’Meara. 2017. “AnnotationBustR: An R Package to Extract Subsequences from GenBank Annotations.” *PeerJ PrePrints*. http://search.proquest.com/openview/c7707966a20c3d64431a9b11c1e7d8ea/1?pq-origsite=gscholar&cbl=2045933.

Broughton, Richard E., Ricardo Betancur-R, Chenhong Li, Gloria Arratia, and Guillermo Ortí. 2013. “Multi-Locus Phylogenetic Analysis Reveals the Pattern and Tempo of Bony Fish Evolution.” *PLoS Currents* 5 (April). https://doi.org/10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. “trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses.” *Bioinformatics* 25 (15): 1972–73.

Chakrabarty, Prosanta, Matthew P. Davis, and John S. Sparks. 2012. “The First Record of a Trans-Oceanic Sister-Group Relationship between Obligate Vertebrate Troglobites.” *PloS*

One 7 (8): e44083.

Chen, Wei-Jen, Céline Bonillo, and Guillaume Lecointre. 2003. “Repeatability of Clades as a Criterion of Reliability: A Case Study for Molecular Phylogeny of Acanthomorpha (Teleostei) with Larger Number of Taxa.” *Molecular Phylogenetics and Evolution* 26 (2): 262–88.

Chow, S., and K. Hazama. 1998. “Universal PCR Primers for S7 Ribosomal Protein Gene Introns in Fish.” *Molecular Ecology* 7 (9): 1255–56.

Depczynski, Martial, and David R. Bellwood. 2005. “Shortest Recorded Vertebrate Lifespan Found in a Coral Reef Fish.” *Current Biology: CB* 15 (8): R288–89.

Eschmeyer, W. N., R. Fricke, and R. Van der Laan. 2016. “Catalog of Fishes: Genera, Species, References.” *Electronic Version Accessed* 19. https://www.researchgate.net/profile/Richard_Van_Der_Laan2/publication/303633861_Catalog_of_Fishes_Genera_Species_References/links/574ed78708ae789584d80783/Catalog-of-Fishes-Genera-Species-References.pdf.

Froese, R., and D. Pauly. 2017. “FishBase 2017, Version (march, 2017).” *World Wide Web Electronic Publication Home Page at: Http://www Fishbase Org*.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.” *Systematic Biology* 59 (3): 307–21.

Hinchliff, Cody E., and Eric H. Roalson. 2013. “Using Supermatrices for Phylogenetic Inquiry: An Example Using the Sedges.” *Systematic Biology* 62 (2): 205–19.

Hinchliff, Cody E., and Stephen Andrew Smith. 2014. “Some Limitations of Public Sequence Data for Phylogenetic Inference (in Plants).” *PloS One* 9 (7): e98986.

Hosner, Peter A., Brant C. Faircloth, Travis C. Glenn, Edward L. Braun, and Rebecca T.

Kimball. 2016. “Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes).” *Molecular Biology and Evolution* 33 (4): 1110–25.

Hughes, Lily C., Guillermo Ortí, Yu Huang, Ying Sun, Carole C. Baldwin, Andrew W. Thompson, Dahiana Arcila, et al. 2018. “Comprehensive Phylogeny of Ray-Finned Fishes (Actinopterygii) Based on Transcriptomic and Genomic Data.” *Proceedings of the National Academy of Sciences of the United States of America*, May. https://doi.org/10.1073/pnas.1719358115.

Izquierdo-Carrasco, Fernando, Stephen A. Smith, and Alexandros Stamatakis. 2011. “Algorithms, Data Structures, and Numerics for Likelihood-Based Phylogenetic Inference of Huge Trees.” *BMC Bioinformatics* 12 (December): 470.

Jaafar, Zeehan, and Edward O. Murdy. 2017. *Fishes Out of Water: Biology and Ecology of Mudskippers*. CRC Press.

Johnson, David G., and Edward B. Brothers. 1993. “*Schindleria*: A Paedomorphic Goby (Teleostei: Aobioidei).” *Bulletin of Marine Science* 52 (1): 441–71.

Johnson, G. David. 1993. “Percomorph Phylogeny-Progress and Problems.” https://repository.si.edu/bitstream/handle/10088/9758/vz_93Percomorph.pdf.

Katoh, Kazutaka, and Martin C. Frith. 2012. “Adding Unaligned Sequences into an Existing Alignment Using MAFFT and LAST.” *Bioinformatics* 28 (23): 3144–46.

Katoh, Kazutaka, Kei-Ichi Kuma, Hiroyuki Toh, and Takashi Miyata. 2005. “MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment.” *Nucleic Acids Research* 33 (2): 511–18.

Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and*

Evolution 30 (4): 772–80.

Kearse, Matthew, Richard Moir, Amy Wilson, Steven Stones-Havas, Matthew Cheung, Shane Sturrock, Simon Buxton, et al. 2012. “Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data.” *Bioinformatics* 28 (12): 1647–49.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. “Dynamics of Mitochondrial DNA Evolution in Animals: Amplification and Sequencing with Conserved Primers.” *Proceedings of the National Academy of Sciences of the United States of America* 86 (16): 6196–6200.

Kottelat, M., and R. Pethiyagoda. 1989. “*Schismatogobius deraniyagalai*, a New Goby from Sri Lanka: Description and Field Observations (Osteichthyes, Gobiidae).” *Spixiana* 12: 315–20.

Kuang, Ting, Luke Tornabene, Jingyan Li, Jiamei Jiang, Prosanta Chakrabarty, John S. Sparks, Gavin J. P. Naylor, and Chenhong Li. 2018. “Phylogenomic Analysis on the Exceptionally Diverse Fish Clade Gobioidei (Actinopterygii: Gobiiformes) and Data-Filtering Based on Molecular Clocklikeness.” *Molecular Phylogenetics and Evolution* 128 (November): 192–202.

Kück, Patrick, and Torsten H. Struck. 2014. “BaCoCa--A Heuristic Software Tool for the Parallel Assessment of Sequence Biases in Hundreds of Gene and Taxon Partitions.” *Molecular Phylogenetics and Evolution* 70: 94–98.

Lanfear, Robert, Brett Calcott, David Kainer, Christoph Mayer, and Alexandros Stamatakis. 2014. “Selecting Optimal Partitioning Schemes for Phylogenomic Datasets.” *BMC Evolutionary Biology* 14 (April): 82.

Larson, H. K. 2001. “A Revision of the Gobiid Fish Genus *Mugilogobius* (Teleostei: Gobioidei), and Its Systematic Placement.”

https://s3.amazonaws.com/academia.edu.documents/49705952/Larson_2001_Mugilogobius_revision.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1555957959&Signature=fhfqfPIVGGR0G9zXMnqNzTRiXT0%3D&response-content-disposition=inline%3B%20filename%3DA_revision_of_the_gobiid_fish_genus_Mugi.pdf.

Li, Chenhong, Guillermo Ortí, Gong Zhang, and Guoqing Lu. 2007. “A Practical Approach to Phylogenomics: The Phylogeny of Ray-Finned Fish (Actinopterygii) as a Case Study.” *BMC Evolutionary Biology* 7 (March): 44.

Li, Chenhong, Guillermo Ortí, and Jinliang Zhao. 2010. “The Phylogenetic Placement of Sinipercid Fishes (‘Perciformes’) Revealed by 11 Nuclear Loci.” *Molecular Phylogenetics and Evolution* 56 (3): 1096–1104.

Li, Chenhong, Betancur-R Ricardo, Wm Leo Smith, and Guillermo Ortí. 2011. “Monophyly and Interrelationships of Snook and Barramundi (Centropomidae *sensu* Greenwood) and Five New Markers for Fish Phylogenetics.” *Molecular Phylogenetics and Evolution* 60 (3): 463–71.

Li, Hongjie, You He, Jiamei Jiang, Zhizhi Liu, and Chenhong Li. 2017. “Molecular Systematics and Phylogenetic Analysis of the Asian Endemic Freshwater Sleepers (Gobiiformes: Odontobutidae).” *Molecular Phylogenetics and Evolution* 121 (December): 1–11.

Linnaeus, C. v. 1758. “Systema Naturae, Edition 10 (Holmiae).”

López, J. Andrés, Wei-Jen Chen, and Guillermo Ortí. 2004. “Esociform Phylogeny.” *Copeia* 2004 (3): 449–64.

Mabuchi, Kohji, Thomas H. Fraser, Hayeun Song, Yoichiro Azuma, and Mutsumi Nishida. 2014. “Revision of the Systematics of the Cardinalfishes (Percomorpha: Apogonidae) Based on Molecular Analyses and Comparative Reevaluation of Morphological Characters.” *Zootaxa* 3846 (2): 151–203.

Mabuchi, Kohji, Noboru Okuda, and Mutsumi Nishida. 2006. “Molecular Phylogeny and Stripe Pattern Evolution in the Cardinalfish Genus *Apogon*.” *Molecular Phylogenetics and Evolution* 38 (1): 90–99.

Maeda, Ken, Toshifumi Saeki, Chuya Shinzato, Ryo Koyanagi, and Nori Satoh. 2018. “Review of *Schismatogobius* (Gobiidae) from Japan, with the Description of a New Species.” *Ichthyological Research* 65 (1): 56–77.

Miller, Peter J. 1992. “The Sperm Duct Gland: A Visceral Synapomorphy for Gobioid Fishes.” *Copeia* 1992 (1): 253–56.

Near, Thomas J., Alex Dornburg, Ron I. Eytan, Benjamin P. Keck, W. Leo Smith, Kristen L. Kuhn, Jon A. Moore, et al. 2013. “Phylogeny and Tempo of Diversification in the Superradiation of Spiny-Rayed Fishes.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (31): 12738–43.

Neilson, Matthew E., and Carol A. Stepien. 2009. “Escape from the Ponto-Caspian: Evolution and Biogeography of an Endemic Goby Species Flock (Benthophilinae: Gobiidae: Teleostei).” *Molecular Phylogenetics and Evolution* 52 (1): 84–102.

Nelson, Joseph S., Terry C. Grande, and Mark V. H. Wilson. 2016. “Phylum Chordata.” In *Fishes of the World*, 13–526. John Wiley & Sons, Inc.

Normark, Benjamin B., Amy R. Mc Cune, and Richard G. Harrison. 1991. “Phylogenetic Relationships of Neopterygian Fishes, Inferred from Mitochondrial DNA Sequences ’.” *Mol. Eiol. Evol* 8 (6): 819–34.

Pagès, H., P. Aboyoun, R. Gentleman, and S. DebRoy. 2017. “Biostrings: Efficient Manipulation of Biological Strings.” *R Package Version* 2 (0).

Paradis, Emmanuel, Simon Blomberg, Ben Bolker, Joseph Brown, Julien Claude, Hoa Sien Cuong, Richard Desper, and Gilles Didier. 2017. “Package ‘ape.’” *Analyses of*

Phylogenetics and Evolution, Version, 2–4.

Parker, Elyse, Alex Dornburg, Omar Domínguez-Domínguez, and Kyle R. Piller. 2019. "Assessing Phylogenetic Information to Reveal Uncertainty in Historical Data: An Example Using Goodeinae (Teleostei: Cyprinodontiformes: Goodeidae)." *Molecular Phylogenetics and Evolution* 134 (May): 282–90.

Pattengale, Nicholas D., Andre J. Aberer, Krister M. Swenson, Alexandros Stamatakis, and Bernard M. E. Moret. 2011. "Uncovering Hidden Phylogenetic Consensus in Large Data Sets." *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM* 8 (4): 902–11.

Pattengale, Nicholas D., Masoud Alipour, Olaf R. P. Bininda-Emonds, Bernard M. E. Moret, and Alexandros Stamatakis. 2010. "How Many Bootstrap Replicates Are Necessary?" *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 17 (3): 337–54.

Pyron, R. Alexander, Frank T. Burbrink, Guarino R. Colli, Adrian Nieto Montes de Oca, Laurie J. Vitt, Caitlin A. Kuczynski, and John J. Wiens. 2011. "The Phylogeny of Advanced Snakes (Colubroidea), with Discovery of a New Subfamily and Comparison of Support Methods for Likelihood Trees." *Molecular Phylogenetics and Evolution* 58 (2): 329–42.

Rabosky, Daniel L., Jonathan Chang, Pascal O. Title, Peter F. Cowman, Lauren Sallan, Matt Friedman, Kristin Kaschner, et al. 2018. "An Inverse Latitudinal Gradient in Speciation Rate for Marine Fishes." *Nature*, July. https://doi.org/10.1038/s41586-018-0273-1.

Salichos, Leonidas, Alexandros Stamatakis, and Antonis Rokas. 2014. "Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees." *Molecular Biology and Evolution* 31 (5): 1261–71.

Sayyari, Erfan, and Siavash Mirarab. 2016. "Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies." *Molecular Biology and Evolution* 33 (7):

1654–68.

Sayyari, Erfan, and Siavash Mirarab. 2018. “Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies.” *Genes* 9 (3). https://doi.org/10.3390/genes9030132.

Schöfl, Gerhard. 2016. “Reutils: Talk to the NCBI EUtils.” https://CRAN.R-project.org/package=reutils.

Schwery, Orlando, and Brian C. O’Meara. 2016. “MonoPhy: A Simple R Package to Find and Visualize Monophyly Issues.” *PeerJ Computer Science* 2 (March): e56.

Shen, Xing-Xing, Leonidas Salichos, and Antonis Rokas. 2016. “A Genome-Scale Investigation of How Sequence, Function, and Tree-Based Gene Properties Influence Phylogenetic Inference.” *Genome Biology and Evolution* 8 (8): 2565–80.

Shimodaira, H., and M. Hasegawa. 1999. “Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference.” *Molecular Biology and Evolution*. https://www.researchgate.net/profile/Hidetoshi_Shimodaira/publication/31240073_Multiple_Comparisons_of_Log-Likelihoods_with_Applications_to_Phylogenetic_Inference/links/02bfe50d08c88888d7000000/Multiple-Comparisons-of-Log-Likelihoods-with-Applications-to-Phylogenetic-Inference.pdf.

Shimodaira, H., and M. Hasegawa. 2001. “CONSEL: For Assessing the Confidence of Phylogenetic Tree Selection.” *Bioinformatics* 17 (12): 1246–47.

Shimodaira, Hidetoshi. 2002. “An Approximately Unbiased Test of Phylogenetic Tree Selection.” *Systematic Biology* 51 (3): 492–508.

Smith, Stephen A., Jeremy M. Beaulieu, and Michael J. Donoghue. 2009. “Mega-Phylogeny Approach for Comparative Biology: An Alternative to Supertree and Supermatrix Approaches.” *BMC Evolutionary Biology* 9 (February): 37.

Smith, Stephen A., and Casey W. Dunn. 2008. “Phyutility: A Phyloinformatics Tool for Trees,

Alignments and Molecular Data." *Bioinformatics* 24 (5): 715–16.

Smith, William Leo, and Ward C. Wheeler. 2006a. "Venom Evolution Widespread in Fishes: A Phylogenetic Road Map for the Bioprospecting of Piscine Venoms." *The Journal of Heredity* 97 (3): 206–17.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13.

Stamatakis, Alexandros, Paul Hoover, and Jacques Rougemont. 2008. "A Rapid Bootstrap Algorithm for the RAxML Web Servers." *Systematic Biology* 57 (5): 758–71.

Steel, Mike, and Michael J. Sanderson. 2010. "Characterizing Phylogenetically Decisive Taxon Coverage." *Applied Mathematics Letters* 23 (1): 82–86.

Thacker, Christine E. 2003. "Molecular Phylogeny of the Gobioid Fishes (Teleostei: Perciformes: Gobioidei)." *Molecular Phylogenetics and Evolution* 26 (3): 354–68.

Thacker, Christine E. 2009. "Phylogeny of Gobioidei and Placement within Acanthomorpha, with a New Classification and Investigation of Diversification and Character Evolution." *Copeia* 2009 (1): 93–104.

Thacker, Christine E. 2013. "Phylogenetic Placement of the European Sand Gobies in Gobionellidae and Characterization of Gobionellid Lineages (Gobiiformes: Gobioidei)." *Zootaxa* 3619: 369–82.

Thacker, Christine E. 2014. "Species and Shape Diversification Are Inversely Correlated among Gobies and Cardinalfishes (Teleostei: Gobiiformes)." *Organisms, Diversity & Evolution* 14 (4): 419–36.

Thacker, Christine E. 2015. "Biogeography of Goby Lineages (Gobiiformes: Gobioidei): Origin, Invasions and Extinction throughout the Cenozoic." *Journal of Biogeography* 42 (9): 1615–25.

Thacker, Christine E. 2017. “Patterns of Divergence in Fish Species Separated by the Isthmus of Panama.” *BMC Evolutionary Biology* 17 (1): 111.

Thacker, Christine E., and Michael A. Hardman. 2005. “Molecular Phylogeny of Basal Gobioid Fishes: Rhyacichthyidae, Odontobutidae, Xenisthmidae, Eleotridae (Teleostei: Perciformes: Gobioidei).” *Molecular Phylogenetics and Evolution* 37 (3): 858–71.

Thacker, Christine E., and Dawn M. Roje. 2009. “Phylogeny of Cardinalfishes (Teleostei: Gobiiformes: Apogonidae) and the Evolution of Visceral Bioluminescence.” *Molecular Phylogenetics and Evolution* 52 (3): 735–45.

Thacker, Christine E., and Dawn M. Roje. 2011. “Phylogeny of Gobiidae and Identification of Gobiid Lineages.” *Systematics and Biodiversity* 9 (4): 329–47.

Thacker, Christine E., Takashi P. Satoh, Eri Katayama, Richard C. Harrington, Ron I. Eytan, and Thomas J. Near. 2015. “Molecular Phylogeny of Percomorpha Resolves *Trichonotus* as the Sister Lineage to Gobioidei (Teleostei: Gobiiformes) and Confirms the Polyphyly of Trachinoidei.” *Molecular Phylogenetics and Evolution* 93 (December): 172–79.

Thacker, Christine E., Andrew R. Thompson, and Dawn M. Roje. 2011. “Phylogeny and Evolution of Indo-Pacific Shrimp-Associated Gobies (Gobiiformes: Gobiidae).” *Molecular Phylogenetics and Evolution* 59 (1): 168–76.

Thomson, Robert C., and H. Bradley Shaffer. 2010. “Sparse Supermatrices for Phylogenetic Inference: Taxonomy, Alignment, Rogue Taxa, and the Phylogeny of Living Turtles.” *Systematic Biology* 59 (1): 42–58.

Tornabene, L., and B. Deis. 2017. “Evaluating the Phylogenetic Position of the Goby Genus *Kelloggella* (Teleostei: Gobiidae), with Notes on Osteology of the Genus and Description of a New Species from ….” *Zoological Journal of the Linnean Society*. https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlx060/47538

59.

Tornabene, Luke, Yongjiu Chen, and Frank Pezold. 2013. “Gobies Are Deeply Divided: Phylogenetic Evidence from Nuclear DNA (Teleostei: Gobioidei: Gobiidae).” *Systematics and Biodiversity* 11 (3): 345–61.

Tsubaki, Remi, and Makoto Kato. 2009. “Intertidal Slope of Coral Sand Beach as a Unique Habitat for Fish: Meiobenthic Diet of the Transparent Sand Dart, *Kraemeria cunicularia* (Gobiidae).” *Marine Biology* 156 (9): 1739–49.

Wang, H. Y., M. P. Tsai, J. Dean, and S. C. Lee. 2001. “Molecular Phylogeny of Gobioid Fishes (Perciformes: Gobioidei) Based on Mitochondrial 12S rRNA Sequences.” *Molecular Phylogenetics and Evolution* 20 (3): 390–408.

Watson, William, and H. J. Walker. 2004. “The World’s Smallest Vertebrate, *Schindleria brevipinguis*, a New Paedomorphic Species in the Family Schindleriidae (Perciformes: Gobioidei).” *RECORDS-AUSTRALIAN MUSEUM* 56 (2): 139–42.

Wickham, Hadley. 2017. “Tidyverse: Easily Install and Load the ‘Tidyverse.’” https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, James Hester, and Jeroen Ooms. 2017. “xml2: Parse XML.” R package version. https://CRAN.R-project.org/package=xml2.

Wiens, John J. 2006. “Missing Data and the Design of Phylogenetic Analyses.” *Journal of Biomedical Informatics* 39 (1): 34–42.

Wiens, John J., James W. Fetzner, Christopher L. Parkinson, and Tod W. Reeder. 2005. “Hylid Frog Phylogeny and Sampling Strategies for Speciose Clades.” *Systematic Biology* 54 (5): 778–807.

Wilke, Claus O. 2016. “Cowplot: Streamlined Plot Theme and Plot Annotations for ‘ggplot2.’” *CRAN Repos*.

Winterbottom, Richard. 1993. “Search for the Gobioid Sister Group (Actinopterygii: Percomorpha).” *Bulletin of Marine Science* 52 (1): 395–414.

Yu, Guangchuang, David K. Smith, Huachen Zhu, Yi Guan, and Tommy Tsan-Yuk Lam. 2017. “Ggtree : An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data.” Edited by Greg McInerny. *Methods in Ecology and Evolution / British Ecological Society* 8 (1): 28–36.

Zhang, Chao, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. 2018. “ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees.” *BMC Bioinformatics* 19 (Suppl 6): 153.

CHAPTER 2

Phylogenomic Analysis Resolves Contentious Relationships and Divergence Times of a Large Radiation of Small Fishes

2.1 Abstract

Gobiarian fishes, exemplified by gobies, sleepers and cardinalfishes, have successfully radiated across coastal marine and aquatic habitats worldwide, yet the biological traits deemed responsible for generating their great diversity, such as small body size, ecomorphological specialization, and rapid rates of evolution, have also mired resolution of their phylogeny. The current classification of Gobiaria is based on molecular phylogenetics, and while broad relationships among major groups have largely been settled by independent investigations, the placement of all higher taxa have not been verified with multilocus data. The root topology recovered from different molecular datasets is contentious, with multilocus and phylogenomic studies resolving nursery and cardinalfishes either in reciprocal or sequential sister clades to other gobiarians, and this deep systematic controversy questions the current two-order classification. Here we use comprehensive sampling of higher taxa and phylogenomic analysis to resolve systematic relationships of gobiarians and estimate a timescale for their diversification under the relaxed molecular clock. Our results reveal that uncertainty at the root can be attributed to incomplete lineage sorting, and indicate other biological factors such as paralogy are not causing the variable phylogenetic signals that confound resolution. We revisit the phylogenetic placement of collared wrigglers (*Xenisthmus spp.*), a group first classified in their own family (Xenisthmidae) but synonymized with sleepers (Eleotridae) based on a mitochondrial hypothesis. Our phylogenomic analyses resolve collared wrigglers outside of the sleepers,

either as sister to gudgeons (Butidae), or sister to ocean sleepers + gobies (Thalasseleotrididae + Gobionellidae + Gobiidae). We date origination of Gobiaria in the youngest age of the Early Cretaceous (104 Ma), find major clades of gobies, sleepers, and cardinalfishes arise in the early Eocene (~50 Ma), and place diversification of goby lineages in the Oligocene and Miocene. In summary, our results support the current two-order classification placing nursery and cardinalfishes in Kurtiformes and the remaining fishes in Gobiiformes, confirm the clade-based phylogenetic classification of eleven gobiarian families, and provide evidence for recognizing collared wrigglers in a twelfth family, Xenisthmidae.

2.2 Introduction

From coral reefs to mudflats, through coastal aquifers and mountain streams, more than 2,400 known species of gobiarian fishes comprise a diverse radiation inhabiting all continents except Antarctica. Composed mainly of gobies, gudgeons, sleepers, and cardinalfishes, gobiarians form one of the six major subclades of percomorph fishes (Alfaro et al. 2018), and are distinct for their extraordinarily small body sizes and plethora of ecological diversity.

Patterns of miniaturization, ecomorphological specialization, and secondary loss of ancestral characters have confounded efforts to resolve gobarian relationships using morphological characters alone (Birdsong et al. 1988; Nelson et al. 2016). The current higher-level classification of Gobiaria, summarized as a consensus tree in Figure 1A, is based on a series of molecular studies published over the last 10 years (Thacker 2009; Betancur-R. et al. 2013; Thacker et al. 2015; Betancur-R et al. 2017).

Major revisions to the sleepers (Eleotridae) and gobies (Gobiidae), and establishment of the clade-based classification were initially based on mitochondrial data (Thacker 2003; Thacker 2009), although consistent relationships have been recovered among the clades tested with multilocus (Thacker et al. 2015; Betancur-R et al. 2017) and phylogenomic data (Alfaro et al. 2018; Kuang et al. 2018) (Fig. 1A). The placement of collared wrigglers (*Xenisthmus spp.*), which resolved with sleepers in (Thacker 2003) mitochondrial phylogeny but retain their own family in taxonomic classification (Xenisthmidae; Eschmeyer et al. 2016), has not been verified with multilocus data. Furthermore, recent phylogenomic studies have not included all 11 families currently recognized in the clade-based classification of Gobiaria (Fig. 1A), or completely sampled all of the 19 subfamillial goby lineages (Thacker and Roje 2011; Agorreta et al. 2013; Thacker 2013).

At the origin of Gobiaria, conflicting results from multilocus and phylogenomic studies have cast doubt on the root topology used to classify Gobiaria into two orders. One topology consistently recovered from multilocus studies (T1; Fig. 1B) resolves nursery + cardinalfishes (=Kurtiformes) as an early-diverging sister clade to sanddivers + gobioids (=Gobiiformes) (Near et al. 2013; Thacker et al. 2015; Betancur-R et al. 2017; Rabosky et al. 2018). Another topology resolved from phylogenomic studies (T2; Fig. 1C) places nurseryfishes as the early-diverging sister clade to cardinalfishes + gobiiforms (Alfaro et al. 2018; Kuang et al. 2018). Neither of the two phylogenomic studies have sampled taxa from an early-diverging cardinalfish subfamily (e.g., Pseudamiinae), or included both nurseryfish species *Kurtus gulliveri* and *K. indicus*, which may have affected inferences if incomplete taxon sampling reduced phylogenetic signal or resulted in tree estimation errors (e.g., via long branch attraction).

The great species richness of Gobiaria has motivated research regarding their evolutionary rates and seemingly exceptional biodiversity. The earliest molecular phylogenetics study of gobioids concluded their lineage diversification was “explosive” based on the shape of a Neighbor-Joining tree estimated from a single mitochondrial gene (Akihito et al. 2000), but macroevolutionary conclusions from later work with more sophisticated methods have been mixed. One multilocus study using broad taxon sampling and Bayesian relaxed molecular clock divergence times estimated under the independent rates (IR) model dated origination of major clades in the late Oligocene and early Miocene, and found gobies to be exceptionally diverse among acanthomorphs (Near et al. 2013). However, two other multilocus studies using larger, more densely-sampled trees and autocorrelated rates (AR) model divergence times fit with maximum-penalized likelihood estimated clade dates that were up to 3X older, placing origination of major groups in the late Cretaceous and Paleocene, and estimating rates of gobiarian diversification that were not different from background rates (Rabosky et al. 2013,

2018). While these three example studies used different taxon sampling (e.g., broad versus dense) and modelling approaches (e.g., maximum likelihood versus Bayesian), their macroevolutionary conclusions were based on a similar underlying diversification model (Alfaro et al. 2009), which leads to the question as to whether differences in taxon sampling and the rate model selected for divergence time estimation were confounding variables that led to different conclusions. Nearly one in twenty described vertebrate species is a gobiarian, and the first step towards understanding the enormous diversity of gobies and their allies is to identify a timescale for clade origination and diversification of lineages.

To resolve questionable clade relationships and estimate a timescale for lineage diversification, we generated new phylogenomic data for over 1,000 UCE loci from 67 gobiarians, and augmented our matrix with UCE data from 18 ingroup and nine outgroup taxa from previously published targeted and whole-genome sequencing studies. We expected that if contentious relationships were due to varied, incomplete taxon and character sampling among studies, then our comprehensive sampling of higher taxa and phylogenomic-scale data would recover a topology robust to different tree inference and construction methods. Alternatively, we expected that if contentious relationships were due to biological scenarios that confused phylogenetic signals, such as deep coalescence, hybridization, or paralogy, or analytical factors that resulted in gene tree estimation error, such as long branch attraction or substitution model misspecification, then we would recover more than one topology from different tree inference and construction methods, because phylogenomic analysis would sum variable phylogenetic signals together in concatenated alignments and reveal gene tree discordance in coalescent species trees. Lastly, we estimated a comprehensive timescale for gobiarian lineage diversification under Bayesian relaxed molecular clock models. Our study is the most comprehensive assessment of higher-level gobiarian relationships yet, clarifying the systematics

and divergence times among all clades and lineages, and establishing an evolutionary foundation for understanding the great diversity of these tiny fishes.

2.3 Material and Methods

2.3.1 Sample Collection and Sequencing

We obtained tissue samples from museum specimens and field collections (Table 1). We performed field collections of new samples using SCUBA at Dongsha Atoll National Park and Kenting National Park, Taiwan. We captured fish with hand nets, euthanized them with clove oil, and preserved them whole in absolute ethanol. All field collections were permitted by domestic (UCLA ARC # 2015-040-01) and foreign institutions (NSYSU IACUC # 10409; TNP # 1046780767).

To extract genomic DNA we used the Qiagen DNeasy Blood and Tissue Kit (Qiagen Inc., Germantown MD, USA), then quantified extracts with a Qubit 2.0 Fluorometer, and sheared 1,000 ng of extract into 300-500 bp fragments using a Diagenode BioRuptor. We then prepared dual-indexed libraries using the Kapa Biosystems (Wilmington MA, USA) kit following a modified Illumina procedure (Rohland and Reich 2012; Faircloth et al. 2015; Glenn et al. 2019), and carried out target enrichment with a MYBaits UCE Capture Kit custom 1,000-locus probe set designed to resolve acanthomorphs (Alfaro et al. 2018), or a more universal, 500-locus probe set designed to resolve teleosts (Faircloth et al. 2013). We washed enrichments bound to streptavidin-coated beads (MyOne C1, Life Technologies, Inc.), quantified enriched libraries using Kapa Biosystems Kit, and sequenced samples on an Illumina NextSeq PE150 platform.

2.3.2 Assembly and Alignment

Reads were demultiplexed at the sequencing facility or with BBTools (Bushnell 2018). We cleaned reads of adapter contamination with the Illumiprocessor wrapper for Trimmomatic (Faircloth 2013; Bolger et al. 2014) and assembled reads into contigs with Trinity (Grabherr et al. 2011). To filter low-coverage contigs, we mapped reads to the Trinity-assembled contigs

using BWA-MEM (Li and Durbin 2009), and trimmed the contigs to a minimum of 5X read coverage with phylUCE (Faircloth 2016).

We downloaded UCE-enriched contigs of 14 ingroup and nine outgroup taxa generated in a previous phylogenomic study of acanthomorph fishes (Alfaro et al. 2018), and whole-genome contigs of four ingroup taxa used in low-coverage genomics studies (You et al. 2014; Malmstrøm et al. 2016) (Tab. 1). We isolated UCE-containing contigs from whole-genome contigs and combined them with newly-sequenced and previous contigs, matched UCE-enriched contigs to the 1,000-locus probe set, and extracted fasta files of UCE sequences with phylUCE.

We aligned UCE sequences with MAFFT under the local pairwise option (--localpair --maxiterate 1000) (Katoh and Standley 2013), and trimmed them with trimAl (-automated1) (Capella-Gutiérrez et al. 2009). We constructed two matrices for downstream analyses: a full 94-taxon matrix filtered to minimum 50% taxon coverage across loci and minimum 50% locus coverage across taxa, and an 18-taxon backbone matrix with 100% coverage across loci and taxa, which included 4 outgroups and 14 ingroup taxa, comprised of two *Kurtus*, one each of Pseudamiinae and Apogoninae, and a single representative from each remaining gobiarian family. Lastly we concatenated alignments of each matrix with phyutility (Smith and Dunn 2008).

2.3.3 Phylogenetic Analysis

We used maximum likelihood and Bayesian analysis under GTR+Γ models to infer phylogenies on locus-partitioned, concatenated alignments from the full 94-taxon and 18-taxon backbone matrices, maximum likelihood to infer gene trees of both matrices, and Bayesian analysis to infer gene trees of the backbone matrix (Tavaré 1986; Yang 1994). We then used the maximum likelihood gene trees from both matrices as input for species tree construction

under the multispecies coalescent model (MSC) (Maddison 1997), and the posterior distributions of Bayesian gene trees in the backbone matrix for Bayesian concordance analysis (BCA) (Ané et al. 2007). Lastly, we inferred maximum likelihood phylogenies from both matrices constrained to recover the conflicting root topology, and dissected phylogenetic signal in support of competing branches by individual alignment sites and loci.

We performed maximum likelihood searches with RAxML (Stamatakis 2014), calling the -f J option to optimize inferences with nearest-neighbor interchange (NNI) moves, and scored branches with SH-like support (Guindon et al. 2010). Then we inferred 100 bootstraps on the locus-partitioned, concatenated alignments (-x), and scored branches of the concatenated maximum likelihood trees with bootstrap support (-f b) and internode certainty (-C -f i) (Salichos et al. 2014).

We performed Bayesian phylogenetic inference with the multi-threaded, MPI-hybrid version of ExaBayes (Aberer et al. 2014). For locus-partitioned concatenated analyses, we used random starting trees, four parallel runs of 500,000 generations each, and sampled MCMC runs every 1,000 generations. To determine topological convergence of parallel runs we used an average standard deviation of split frequencies threshold of 1%. We discarded the initial 25% of generations as burn-in, combined the remaining samples from parallel runs, and verified the effective sample sizes (ESS) of all model parameters were above 200. Preliminary Bayesian analysis on the locus-partitioned, full 94-taxon matrix indicated ESS were small for substitution model parameters of several partitions of the concatenated alignment. We observed preliminary ESS would not increase with additional MCMC generations, so we added a single heated chain to each run on the full 94-taxon matrix to improve MCMC sampling efficiency. ExaBayes runs were configured similarly for gene trees except we used parsimony starting trees, two parallel runs of 1,000,000 generations each, and no heated chains.

To construct multispecies coalescent (MSC) species trees from maximum likelihood gene trees of the full and backbone matrices, we used ASTRAL-III (Zhang et al. 2018). ASTRAL-III summarizes gene trees into a species tree by finding the tree within a restricted search space of topologies that maximizes the number of quartets shared with gene trees. The multispecies coalescent (MSC) model assumes all gene tree discordance is caused by incomplete lineage sorting (ILS), therefore to summarize gene trees into a species tree under the MSC model, ASTRAL-III assumes gene trees were estimated without error. Trimmed alignments of UCE loci are typically short (average 305 sites for actinopterygians in Faircloth et al. 2013), and because previous studies have found locus alignment length to be the strongest predictor of phylogenetic signal (Shen et al. 2016), we attempted to mitigate error among the maximum likelihood gene trees by collapsing branches with very low SH-like support. Zhang et al. (2018) found accuracy of ASTRAL-III was improved by contracting branches with bootstrap scores below 10, but that accuracy was reduced by contracting branches with bootstrap scores above 10. We used an SH-like branch support threshold of 0.05 and contracted poorly-supported branches using NewickUtilities (Junier and Zdobnov 2010).

We input the posterior distributions of Bayesian gene trees from the backbone matrix for Bayesian concordance analysis (BCA) with BUCKy (Larget et al. 2010). BUCKy estimates the proportion of genes in a sample (and genome) with concordant phylogeny for a clade without making assumptions regarding the cause of discordance. After identified by Bayesian concordance analysis, BUCKy summarizes clades with the highest concordance factors (CF) in the posterior distribution of Bayesian gene trees into a primary concordance tree by greedy consensus. BUCKy also constructs a population tree under the multispecies coalescent model, assuming all discordance caused by incomplete lineage sorting, by greedy consensus of quartets with the highest concordance factors. Topological differences between the coalescent

population and the primary concordance trees, and overlapping credible intervals (CIs) of concordance factors for conflicting clades (observed with CF > 0.05), are evidence that incomplete lineage sorting is the only cause of gene tree discordance. We used mbsum to prepare posterior distributions of Bayesian gene trees (25% burn-in discarded) for input to BUCKy. To choose the prior for the number of topological clusters of gene trees (α), we binned the maximum likelihood gene trees by topology in R with distory (Chakerian and Holmes 2013), and input the number of unique topologies to obtain an appropriate α value using the BUCKy-provided script *prior_standalone.R*. All maximum likelihood gene trees in the 18-taxon backbone matrix were unique, and we determined an appropriate prior of α = 10,000. We used four independent, coupled-chain BUCKy runs of 11,000,000 generations each, sampled MCMC runs every 10 generations, and discarded the initial 10% of generations as burn-in.

We inferred maximum likelihood phylogenies from both matrices constrained to recover the conflicting root topology, and evaluated phylogenetic signal in support of competing root branches by each site and gene of the locus-partitioned, concatenated alignments. We called the -g option in RAxML to perform locus-partitioned, maximum likelihood tree searches constrained to the alternative topology for both the full 94-taxon and the backbone 18-taxon datasets. Then we input the conflicting maximum likelihood trees into RAxML and computed per-site log likelihood scores with the -f G option on locus-partitioned, concatenated alignments. We finished by summarizing site-wise log likelihood scores (SLS) into gene-wise log likelihood scores (GLS), and plotting ΔSLS and ΔGLS by alignment position in R with tidyverse.

2.3.4 Divergence Dating

We used 15 fossils to constrain calibration intervals for 12 nodes and the acanthomorph root (Fig. 2; Tab. 2). We used the algorithm of Hedman (2010) to compute the upper 95%

highest probability density (HPD) bound for the range of fossil taxon occurrence dates based on fossil preservation rates and a series of stratigraphically-consistent outgroups. The latest geologic age described from the fossil collection locality was used as a “hard” minimum constraint for node calibrations, and the upper 95% HPD bound for the ages of fossil taxa was used as a “soft” maximum constraint for nodes (i.e., with a 5% probability of node ages exceeding the maximum constraint bounds).

Bayesian relaxed clock divergence time estimation is computationally intensive, so to improve efficiency we reduced the number of taxa and characters in our full 94-taxon matrix, and partitioned the pruned alignment by substitution rates. We reduced the matrix to two taxa per lineage or clade, and trimmed alignment positions containing gaps with trimAl (-nogaps). When more than two taxa were sampled from a lineage we retained the earliest-diverging taxon and the taxon with the best data coverage. Extra taxa were pruned from the locus-partitioned, concatenated Bayesian topology with NewickUtilities. The pruned dataset contained 80 taxa and 138,008 sites. To approximate the substitution rates of loci for alignment partitioning, we computed pairwise maximum likelihood distances for each locus under HKY+Γ models with RAxML (-f x --HKY85) (Hasegawa et al. 1985), partitioned loci by quartiles of mean pairwise distances in R with tidyverse (Wickham 2017), and concatenated rate-partitioned alignments with phyutility. Alignment partition lengths ranged from 38,638 sites for the first quartile partition (slowest-evolving loci), to 29,933 sites for the fourth quartile partition (fastest-evolving loci).

We estimated node ages on the reduced Bayesian topology under relaxed clock models with MCMCTree (Yang 2007). We first used baseml to calculate the gradient and Hessian on the rate-partitioned alignment under HKY+Γ substitution and relaxed molecular clock models (Yang 2007; dos Reis et al. 2017). We then used MCMCTree to sample posterior distributions of node ages and relaxed clock model parameters using an approximate computation of the

likelihood (dos Reis and Yang 2011). To generate uniform age priors for nodes in the tree without fossil calibrations, birth-death process parameters were set to $\lambda = \mu = 1$ and sampling $\rho = 0.1$ (BDparas = 1 1 0.1). We set Γ priors for the substitution model transition/transversion rate parameter $\kappa = 3$ (kappa_gamma = 6 2) and rate heterogeneity shape parameter $\alpha = 1$ (alpha_gamma = 1 1). We set Dirichlet-Γ priors for the relaxed clock models rate parameter $r = 0.05$ (rgene_gamma = 2 40 1) and rate heterogeneity parameter $\sigma^2 = 1$ (sigma2_gamma = 2 2 1). We ran MCMCTree independently two times of 2,000,000 generations each, sampled MCMC runs every 100 generations, and discarded the first 25% of generations as burn-in. We checked for convergence of posterior node age estimates from independent runs with ruled scatterplots, and then combined MCMC samples with LogCombiner (Rambaut and Drummond 2014). We used the combined MCMC output to estimate the effective sample sizes (ESS) of parameters in R with coda (Plummer et al. 2006), and verified ESS > 200 before computing 95% highest probability density (HPD) credibility intervals for parameters with bayestestR (Makowski et al. 2019).

Bayesian divergence times can be affected by the choice of relaxed molecular clock model used in estimating evolutionary rate parameters (dos Reis et al. 2018). We conducted Bayesian model selection to choose the best-fitting relaxed molecular clock model for explaining evolutionary rate variation in our dataset. The posterior distributions sampled with the approximate likelihood method cannot be used for computing posterior model probabilities (dos Reis and Yang 2011), and because sampling of marginal likelihoods with the exact method is computationally-prohibitive on phylogenomic data, we carried out Bayesian model selection on a subset of eight randomly-chosen UCE loci. We used stepping stone sampling of marginal likelihoods to compute posterior probabilities of the strict clock (SC), independent rates (IR), and autocorrelated rates (AR) models (Xie et al. 2011). UCE loci with relatively moderate sampling

of taxa (i.e., 30 < n < 60 with unique UCE sequences) and alignment lengths of at least 150 sites were used. To prepare for marginal likelihood sampling, we pruned the Bayesian topology to match taxon occupancy of individual loci with NewickUtilities, removed all fossil constraints, and fixed the root constraint to 143 Ma. We then generated MCMCTree control files replicated at eight β values ranging from 0 to 0.513 in R with mcmc3r (dos Reis et al. 2018). Priors for molecular clock models were set the same as above, except we used birth-death process sampling ρ = 0 (BDparas = 1 1 0), and a Dirichlet-Γ prior for the rate heterogeneity parameter σ^2 = 0.1 (sigma2_gamma = 1 10 1). We used the exact likelihood option in MCMCTree to generate power-posteriors at each β value for 525,000 steps, sampled MCMC runs every 200 generations, and discarded the first 25% of generations as burn-in. We replicated the analysis for strict and relaxed molecular clock models, and computed Bayes Factors and posterior model probabilities (assuming equal prior model probabilities) in R with mcmc3r.

Lastly, we generated infinite sites plots to assess uncertainty in posterior node age estimates attributed to the sequence data under each relaxed clock model (Inoue et al. 2010). To prepare the infinite site plots we calculated 95% highest probability density (HPD) credibility interval widths (w) of the mean posterior divergence times (t), fit simple linear regression models without intercept terms, and produced plots in R with tidyverse.

2.4 Results

2.4.1 Sample Collection and Sequencing

Sixty seven samples were newly sequenced for this study (Tab. 1). The majority of samples (*n* = 61) were enriched for 1,000 UCE loci (Alfaro et al. 2018), and a smaller number of samples (*n* = 6) were enriched for 500 UCE loci (Faircloth et al. 2013). The number of paired reads per sample ranged from 30,858 for *Perccottus* to 22,866,896 for *Guavina*, with an overall average of 5,427,222 paired reads per sample.

2.4.2 Assembly and Alignment

The average proportion of paired reads remaining after trimming adapters and low-quality bases with Illumiprocessor was 92.7%, which ranged from 82.6% for *Gnatholepis* to 98.4% for *Koumansetta*. The number of Trinity-assembled contigs per sample ranged from 988 for *Perccottus* to 196,583 for *Grahamichthys*, with an overall average of 26,975 contigs per sample. After aligning reads to contigs with BWA-MEM and filtering with a 5X coverage threshold, the mean coverage for contigs was 24.5X, and ranged from 7.9X for *Perccottus* to 43.8X for *Xenisthmus polyzonatus*. Combining contigs of newly-sequenced samples with contigs of 27 previously published samples yielded an incomplete matrix of 1,194 UCE loci. The number of UCE loci per sample in the incomplete matrix ranged from 361 for *Giuris*, to 1,076 for *Odontobutis*, with an overall average of 849 UCE loci per sample (Tab. 1).

The full 94-taxon matrix contained 704 loci and was 85% complete. The mean length of trimmed alignments was 461 sites, and the concatenated alignment length was 325,561 sites. The 100% complete, 18-taxon backbone matrix contained 120 loci, and the concatenated alignment length was 65,732 sites.

2.4.3 Phylogenetic Analysis

Two ingroup topologies were recovered from all tree inference and construction methods. One topology (T1) resolved nursery + cardinalfishes (=Kurtiformes) in an early-diverging clade sister to remaining gobiarians (=Gobiiformes), and the other topology (T2) resolved nurseryfishes in the early-diverging sister clade. The T1 topology inferred on locus-partitioned, concatenated alignments was robust to both maximum likelihood and Bayesian methods of inference, and to full 94-taxon (85% complete) and backbone 18-taxon (100% complete) matrices (Fig. 3; Fig. 5A). Similarly, the same multispecies coalescent species tree T2 topology was estimated from summaries of 704 contracted maximum likelihood gene trees with incomplete taxon occupancy, and 120 contracted maximum likelihood gene trees with no missing taxa (Fig. 4; Fig. 5B). However, both topologies were recovered from Bayesian concordance analysis of the posterior distributions of Bayesian gene trees in the backbone matrix (Fig. 5C; Fig. 5D). The root topology of the Bayesian concordance analysis population tree matched the T1 topology of concatenated maximum likelihood and Bayesian inferences (Fig. 5A; Fig. 5C), and the root topology of the primary concordance tree matched the T2 topology of the multispecies coalescent species tree (Fig. 5B; Fig. 5D).

The majority (92%) of T1 branches from concatenated maximum likelihood analysis of the full 94-taxon matrix were recovered with full support for all measures (e.g., SH-like: SHL = 1.00; Bootstrap: BS = 100; Internode Certainty: IC = 1.00), and 98% of T1 branches had full Bayesian posterior probability (PP = 1.00; Fig. 3). Similar patterns were recovered from concatenated analyses of the 18-taxon backbone matrix, where 82% of T1 branches were fully supported by maximum likelihood measures, and 94% of T1 branches were fully supported by Bayesian posterior probability (Fig. 5A). Relatively fewer branches were fully supported in the multispecies coalescent species trees, with 76% and 86% of branches supported with 100%

local posterior probability (LPP = 1.00) in the 94-taxon and 18-taxon species trees, respectively (Fig. 4; Fig. 5B).

The distribution of phylogenetic signal in locus partitioned, concatenated alignments favored the T1 root topology in both full and backbone matrices (Fig. 6). For the full matrix, the log-likelihood score of the unconstrained T1 tree was 15.3 units greater than the constrained T2 tree score, and for the 18-taxon backbone matrix, the log-likelihood score of the unconstrained T1 tree was 27.0 units greater than the constrained T2 tree. Dissection of log-likelihood scores by sites and genes revealed relatively more phylogenetic signal supporting the T1 topology. The T1 topology was favored by 57.1% of sites in the full 94-taxon concatenated alignment, whereas the T2 topology was favored by the remaining 42.9% of 325,561 sites. The distribution of gene-wise log-likelihood scores was close, but favored the T1 topology in 354 loci (50.3%), whereas the T2 topology was favored in 350 loci (49.7%; Fig. 6A). For the 18-taxon backbone alignment, the T1 tree was supported by 61.7% of aligned sites whereas the T2 was supported by the remaining 38.3% of 65,732 sites, and 67 loci (55.8%) favored the T1 topology whereas 53 loci (44.2%) favored the T2 topology (Fig. 6B).

2.4.4 Divergence Dating

Bayesian posterior model probabilities indicated that independent rates (IR) was the best-fitting relaxed molecular clock model for explaining evolutionary rate variation of the gene trees we tested (Tab. 3). Of the eight randomly-chosen UCE loci examined, the independent rates (IR) model was selected in six loci, and the autocorrelated rates (AR) and strict clock (SC) models were each selected in a single locus (Tab. 3). We present clade age results estimated under the independent rates (IR) model, summarize patterns of goby lineage divergence times, and compare the IR timescale to the one estimated under the autocorrelated rates (AR) model.

The posterior mean time (*t*) estimate for the origination of Gobiaria was *t* = 103.7 Ma (95% CI: 93.1-114.8) (Fig. 7). Origination estimates for kurtiform and gobiiform clades were just before (*t* = 101.8 Ma; 95% CI: 91.1-113.1) and after (*t* = 99.1 Ma; 95% CI: 89.0-110.0) the earliest age of the Late Cretaceous, respectively (Fig. 7). Within the kurtiform clade, the crown age of nurseryfishes (Kurtidae) was estimated in the Eocene (*t* = 40.0 Ma; 95% CI: 19.6-63.4), and the crown age of cardinalfishes (Apogonidae) was estimated just before the Cretaceous-Paleogene boundary (*t* = 67.4 Ma; 95% CI: 54.6-83.0).

In the gobiiform clade, origination estimated for the crown of loach gobies + freshwater sleepers (Rhyacichthyidae + Odontobutidae) was just before the Eocene-Oligocene boundary (*t* = 34.8 Ma; 95% CI: 21.5-49.9), and crown freshwater sleepers were placed in the Oligocene (*t* = 27.5 Ma; 95% CI: 14.5-41.5) (Fig. 7). Origination times of blind cave gudgeons (Milyeringidae) (*t* = 53.0 Ma; 95% CI: 35.2-72.3), sleepers (Eleotridae) (*t* = 50.1 Ma; 95% CI: 42.0-59.4), and gudgeons (Butidae) (*t* = 43.7 Ma; 95% CI: 32.3-54.3) were all Eocene, whereas collared wrigglers (Xenisthmidae) (*t* = 11.6 Ma; 95% CI: 5.9-18.1) and ocean sleepers (*t* = 14.8 Ma; 95% CI: 7.6-23.4) were both set in the mid-Miocene (Fig. 7). Origination of the two largest clades Gobionellidae (*t* = 48.9 Ma; 95% CI: 42.2-55.7) and Gobiidae (*t* = 49.2 Ma; 95% CI: 42.8-55.3) was placed in the Eocene (Fig. 7).

Origination of all goby lineages was estimated in the Oligocene and Miocene (Fig. 7). Four gobionellid goby lineages were placed in the Oligocene, and in Gobiidae, four goby lineages were placed in the Oligocene, and seven were placed in the Miocene. Node ages were not available for the *Pomatoschistus* lineage of Gobionellidae, or the *Kraemeria*, *Callogobius*, and *Aphia* lineages of gobiidae, because these lineages were only represented by a single branch (Fig. 7).

Gobiarian root age estimates were robust to different relaxed clock models (Fig. 7; Fig. 8; Fig. 9). The origination estimate of Gobiaria was *t* = 103.7 Ma (95% CI: 93.1-114.8) under the independent rates (IR) model, and *t* = 104.5 Ma (95% CI: 98.5-110.4) under the autocorrelated rates (AR) model. Node ages of kurtiform and gobiiform clades were also robust to evolutionary rate models, with estimates under competing models differing by 0.5 Ma to 1.5 Ma (Fig. 7; Fig. 8).

Divergence times within orders were complementary under different relaxed molecular clock models (Fig. 7; Fig. 8; Fig. 9). Age estimates of kurtiform clades were younger under the independent rates (IR) model, with differences ranging from +5.8 Ma for Apogoninae to +37.8 Ma for nurseryfishes (Kurtidae) under the autocorrelated (AR) model (Fig. 9). Within Gobiiformes, age estimates were older for most nodes under the independent rates (IR) model, with exception to five nodes: loach gobies + freshwater sleepers (Rhyacichthyidae + Odontobutidae; -27.4 Ma), freshwater sleepers (Odontobutidae; -25.5 Ma), blind cave gudgeons (Milyeringidae; -9.0 Ma), gudgeons (Butidae; -10.7 Ma and -7.5 Ma), and the *Gobiopsis* lineage of Gobiidae (-0.5 Ma) (Fig. 9).

The linear fit to infinite sites plots was moderate under both evolutionary rate models (Fig. 10). Posterior mean times (*t*) explained 70.3% of the variation in posterior CI widths (*w*) under the independent rates (IR) model (Fig. 10A), whereas 60.8% of the variation in *w* was explained by *t* under the autocorrelated rates (AR) model (Fig. 10B). Regression coefficients for *t* were β = 0.256 (95% CI: 0.218-0.293) under independent rates (IR) and β = 0.170 (95% CI: 0.139-0.200) under autocorrelated rates (AR) models (Fig. 10).

2.5 Discussion

We used a phylogenomic dataset of ultraconserved element loci and comprehensive sampling of higher taxa to interrogate uncertain relationships and estimate a timescale for gobiarian diversification. Our study revealed that contentious relationships at the root of Gobiaria can be attributed to the presence of relatively homogeneous, discordant phylogenetic signals persisting from ancient deep coalescence. Overall, we found more evidence for the currently-accepted, reciprocal sister clade topology placing nursery + cardinalfishes (=Kurtiformes) in a clade sister to Gobiiformes. We resolved collared wrigglers (*Xenisthmus spp.*) in a clade sister to gudgeons (Butidae), in contrast to earlier gobioid revisions based on mitochondrial DNA evidence, and found uncertain resolution among some sleepers (Eleotridae). Bayesian relaxed molecular clock model selection carried out on a subset of gene trees was variable but favored the independent rates (IR) model, dating the origination of gobiarian clades in the early Eocene, and setting diversification of goby lineages in the Oligocene and Miocene.

2.5.1 Root Topology

The two different root topologies recovered among all methods and matrices corresponded to the competing phylogenies of previous studies we set out to resolve. Topology T1, placing nursery + cardinalfishes sister to gobiiforms, was recovered from locus-partitioned, concatenated maximum likelihood and Bayesian inferences, and the Bayesian concordance analysis population tree. Conflicting topology T2, placing nurseryfishes sister to cardinalfishes + gobiiforms, was recovered in the primary concordance tree, the multispecies coalescent species tree, and was also observed as the minority bootstrap consensus topology with 26% support (e.g., observed T1 topology BS = 74). To determine which topology reflects the phylogeny of Gobiaria, we evaluated differences in relative branch support and phylogenetic signal for

competing topologies and found slightly better support for the T1 topology, placing nursery + cardinalfishes in a clade sister to the gobiiforms.

If clades conflicting with the Bayesian concordance analysis primary concordance tree have overlapping CIs, then incomplete lineage sorting (ILS) can be assumed to be the only cause of gene tree discordance, and if CIs are not overlapping then hybridization, horizontal gene transfer, or gene duplication can be implicated (Ané et al. 2007). The Bayesian concordance analysis primary concordance tree placement of nurseryfishes as the earliest-diverging clade was supported by CF = 0.182 (95% CI: 0.125-0.233 (Fig. 5D), yet nearly identical support was estimated for the conflicting clade placing nursery + cardinalfishes in a clade sister to gobiiforms (CF = 0.181; 95% CI: 0.133-0.233), as well as another conflicting arrangement placing sanddivers (Trichonotidae) as the earliest-diverging lineage (CF = 0.159; 95% CI: 0.108-0.208). The overlapping concordance factor credible intervals of conflicting clades at the gobiarian root indicate incomplete lineage sorting (ILS) to be the only cause of discordance, and exclude other biological explanations such as hybridization and paralogy. The Bayesian concordance analysis population tree is constructed under the multispecies coalescent model and reflects the true phylogeny when incomplete lineage sorting (ILS) is assumed to be the only cause of gene tree discordance (Larget et al. 2010). Thus, our results indicate the T1 topology observed in concatenated maximum likelihood/Bayesian and (multispecies coalescent) population trees reflects the phylogeny of Gobiaria.

Dissection of phylogenetic signals in support of conflicting root topologies revealed a relatively homogeneous distribution of sites and genes favoring alternative trees (Fig. 6). Previous research has shown that contentious relationships in phylogenomic studies can be attributed to just one or two genes with extremely high phylogenetic signals (Shen et al. 2017). We did not observe any extreme patterns in site-wise or gene-wise log-likelihood scores

favoring either conflicting root topology (Fig. 6). Therefore, we attribute previous recovery of alternative root topologies among concatenated tree inferences to sampling error. In contrast, the recovery of the conflicting T2 topology in our multispecies coalescent species trees can be attributed to gene tree estimation error, which is expected given our inference on short ultra-conserved element (UCE) sequence alignments (e.g., average length <500 sites).

2.5.2 Clade Relationships

We recovered the expected cardinalfish relationships with complete branch support, placing *Pseudamia* (Pseudamiinae) as an early-diverging sister lineage to Apogoninae, and resolving two major subclades within Apogoninae (Fig. 3; Fig. 4). One of the subclades groups Western Atlantic *Phaeoptyx*, *Astrapogon*, *Paroncheilus*, and *Apogon*, consistent with the Apogonini tribe of Mabuchi et al. (2014). The remaining Indo-West Pacific genera *Pristiapogon*, *Taeniamia*, *Cheilodipterus*, *Ostorhinchus*, and *Fibramia* were grouped in another subclade. We unexpectedly resolved *Ostorhinchus* paraphyletic with respect to *Fibramia* with 100% branch support (Fig. 3; Fig. 4). The diverse (~100 species) genus *Ostorhinchus* was not monophyletic in Mabuchi et al. (2014), but the two species we analyzed were grouped together with the majority of congeners in that hypothesis. Our results highlight the need for a comprehensive revision of *Ostorhinchus*.

All of our trees revealed a deep division at the origin of the sleepers (Eleotridae; Fig. 3; Fig. 4). Consistent with previous studies that used mitochondrial data (Thacker and Hardman 2005; Thacker 2017), we resolved a subclade (I) composed of *Ratsirakia*, *Giurus*, *Mogurnda*, and *Tateurndina*, and another subclade (II) grouping *Eleotris, Erotelis*, *Philypnodon*, *Gobiomorphus*, *Hypseleotris*, *Dormitator*, *Guavina*, *Microphilypnus*, *Hemieleotris*, and *Gobiomorus*. The taxa sampled in subclade I are generally brightly colored, reside in aquatic to

slightly-brackish environments, and are distributed throughout tropical and subtropical regions of Africa, Asia, and Oceania. The distribution of taxa in subclade II overlaps with subclade I but extends to New Zealand and the New World tropics, with some taxa inhabiting marine environments. The uncertainty along the backbone of subclade II corresponds to diversification of lineages currently inhabiting Australia and Central America (Fig. 3; Fig. 4), and may reflect an ancient period of rapid lineage diversification. Nevertheless, several well-supported generic lineages are apparent within the two subclades of Eleotridae (Fig. 3; Fig. 4). Within subclade I, we placed *Ratsirakia* as sister to *Mogurnda*, *Giurus,* + *Tateurndina*. We identified five additional groups in subclade II: *Eleotris* (including *Erotelis*), *Gobiomorphus* + *Philypnodon*, *Hypseleotris*, *Dormitator* + *Guavina*, and *Gobiomorus, Microphilypnus* + *Hemieleotris*. Some of these genera have been previously aligned with morphology. Birdsong et al. (1988) placed *Guavina* in the *Dormitator* group, *Erotelis* in the *Eleotris* group, *Philypnodon* in the *Gobiomorphus* group (which also contained *Mogurnda*), and *Hypseleotris* in its own group (which also included *Hemieleotris*). Further investigation with comprehensive sampling of genera is warranted to fully describe the phylogenetic structure and ecological characteristics of sleeper lineages.

We resolved collared wrigglers (*Xenisthmus spp.*) in a clade sister to gudgeons (Butidae) in concatenated and multispecies coalescent species trees (Fig. 3; Fig. 4; Fig. 5A; Fig. 5B), and as an early-diverging sister lineage to ocean sleepers + gobies (Thalasseleotrididae + Gobionellidae + Gobiidae) in Bayesian concordance analysis population and primary concordance trees (Fig. 5C; Fig. 5D). The recovered topology was perfectly supported in concatenated and coalescent analyses of the full 94-taxon matrix, and strongly supported in analyses of the 18-taxon backbone matrix (PP = 1; SHL = 0.99; BS = 97; IC = 0.81). The branch was recovered with less support in the multispecies coalescent species tree (LPP = 0.78), and the Bayesian concordance analysis population (CF = 0.44) and primary concordance trees (CF

= 0.32). The phylogenetic placement of collared wrigglers was previously hypothesized based on fragmentary mitochondrial data, which grouped them with eleotrid sleepers (Thacker 2003; Thacker 2009). However, the previous grouping of collared wrigglers was unusual because xenisthmids are strictly marine, reef-associated, and distributed in the tropical Indo-West Pacific, which contrasts with the mostly aquatic or estuarine, globally-distributed sleepers. The concordance factors (CF) for competing topologies involving collared wrigglers and gudgeons were nearly identical, with the conflicting sister clade CF = 0.32 (95% CI: 0.26-0.38) and observed sequential clade CF = 0.31 (95% CI: 0.26-0.38). Agorreta and Rüber (2012) recovered collared wrigglers sister to gudgeons in their reanalysis of Akihito et al. (2000), and Hoese and Gill (1993) noted the extrascapulae morphological character was present in *Xenisthmus* and most of Butidae, but absent in sleepers (Eleotridae). Regardless of the relationship between collared wrigglers and gudgeons, our results overwhelmingly support recognition of Xenisthmidae in the phylogenetic classification of Gobiaria.

Consistent with Kuang et al.'s (2018) phylogenomic placement, we resolved *Butis* as the earliest-diverging gudgeon (Fig. 3; Fig. 4). However, contrary to phylogenetic hypotheses based on mitochondrial data (Thacker and Hardman 2005; Thacker 2017), we recovered a topology that places *Kribia* and *Ophiocara* as sister taxa, consistent with the morphological phylogenetic hypothesis of Gierl et al. (2013).

We recovered the expected relationships among gobies with strong branch support (Fig. 3; Fig. 4). Gobionellidae relationships had complete maximum likelihood and Bayesian analysis branch support and nearly-perfect local posterior probability (LPP) in the multispecies coalescent species tree, with a single branch subtending the division of *Pomatoschistus* and *Acanthogobius* + *Mugilogobius* lineages recovered with LPP = 0.99 (Fig. 4). Branch support was more variable within Gobiidae but strong overall, with the backbone completely supported in

concatenated analyses, except for the relationship of *Gobiosoma*, *Gobius*, *Priolepis*, and *Valenciennea* + *Aphia* lineages (Fig. 3). The backbone branches subtending these diverse goby lineages were short, and a poorly-supported rearranged topology was observed in the multispecies coalescent species tree (Fig. 4). The uncertain relationship of *Gobiosoma*, *Gobius*, *Priolepis*, and *Valenciennea* + *Aphia* lineages was also documented in the phylogenomic analysis of Kuang et al. (2018), although the *Priolepis* and *Aphia* lineages were not sampled in that study. We observed one other marginally-supported relationship in the multispecies coalescent tree backbone, on the branch subtending *Kraemeria* and the *Gobiopsis* + *Cryptocentrus* lineages, recovered with 75% local posterior probability (LPP = 0.75; Fig. 4). *Kraemeria* was previously identified as a rogue taxon by Agorreta et al. (2013).

2.5.3 Divergence Times

Our Bayesian relaxed molecular clock divergence time estimates dated the origination of Gobiaria in the youngest age of the Early Cretaceous (104 Ma; Fig. 7). Previous multilocus and phylogenomic analyses have estimated younger origination dates, ranging from 89 Ma (Near et al. 2013) to 102 Ma (Betancur-R et al. 2017), with an average age of 97 Ma drawn from recent estimates (Li et al. 2017; Alfaro et al. 2018; Hughes et al. 2018; Rabosky et al. 2018). Similarly, Malmstrøm et al. (2016) estimated the divergence of a single goby (*Aphia* lineage) and a scombrid outgroup (*Thunnus*) at 104 Ma with Bayesian analysis of 111 clock-like exons across broadly sampled teleost fishes.

Previous studies have estimated a wide range of divergence times for cardinalfishes, with origination dates of the main apogonine clade ranging from 10.6 Ma to 65.2 Ma for phylogenomic and multilocus estimates (Alfaro et al. 2018; Rabosky et al. 2018). Most of the previous estimates for cardinalfishes are pushed towards the younger range of ages, with an

average age of 29.4 Ma drawn across recent studies (Near et al. 2013; Betancur-R et al. 2017; Li et al. 2017; Hughes et al. 2018). Our Bayesian relaxed molecular clock estimate under the independent rates (IR) model dated the apogonines at 50 Ma, however we constrained the minimum age of that clade to 49.0 Ma based on the fossil †*Eosphaeramia margaritae* (Bannikov 2008). Placement of the minimum age constraint on apogonine cardinalfishes is supported by parsimony analysis of morphology that resolve †*Eosphaeramia margaritae* sister to *Sphaeramia*, who group with *Ostorhinchus* in molecular phylogeny (Mabuchi et al. 2014). Cardinalfishes were diverse in the Eocene, and a total of five species have been described from the Pescaria and Postale sites of Bolca lagerstätten (Bannikov and Fraser 2016). Extant cardinalfishes remain diverse today (>350 spp) and are considered one of the major coral reef fish families (Cowman and Bellwood 2011). We observed a broad mismatch in prior versus posterior distributions for cardinalfishes under the independent rates (IR) model (Fig. 7), and their divergence times were the most different between IR and autocorrelated rates (AR) models (Fig. 8; Fig. 9), which could be attributed to comparatively slow substitution rates in Apogoninae, apparent from relatively short branches recovered in the Bayesian analysis tree (Fig. 3). Further investigation into cardinalfish divergence times using additional taxon sampling is warranted to understand the mismatch of neontological and paleontological timescales.

We estimated the age of blind cave gudgeons (Milyeringidae) divergence at 53 Ma (95% CI: 35.2-72.3) under the independent rates (IR) model. Blind cave gudgeons have a Gondwanan distribution, with disjunct lineages inhabiting northwestern Australia and Madagascar (Chakrabarty et al. 2012). The landmasses that blind cave gudgeons currently inhabit were proximate until breakup in mid-Early Cretaceous (~120 Ma), which nearly reaches the 116 Ma maximum bound from a previous divergence time estimate of 77 Ma (95% CI: 44-116) (Chakrabarty et al. 2012). The estimate of Chakrabarty et al. (2012) was interrogated

by de Bruyn et al. (2013), who found one of their fossil constraints internally inconsistent, and re-estimated blind cave gudgeon age (excluding one fossil calibration) at 40 Ma (95% CI: 27-53). Curiously, the blind cave gudgeon posterior age distribution that we generated under the independent rates (IR) model was bimodal (Fig. 11), and the width (*w*) of the CI for that time was the second-largest *w* observed across all of our estimates (Fig. 10A). One of the modes of the blind cave gudgeon posterior age was at 64 Ma, and the other mode was at 49 Ma (Fig. 11), with the posterior mean time shifted towards the younger mode (*t* = 53; 95% CI: 35.2-72.3). Our posterior for blind cave gudgeon age under the autocorrelated rates (AR) model was normal (*t* = 61.9; 95% CI: 56.5-67.2), and was close to the older posterior mode generated under the independent rates (IR) model (Fig. 11). Further research on rates of molecular evolution of blind cave gudgeons is warranted given their extraordinary natural history and the uncertainty we observed in posterior divergence times.

The goodness-of-fit of infinite sites plots models saturation of evolutionary rates in molecular sequence data (Inoue et al. 2010). Goodness-of-fit for the infinite site plot was marginal under both rate models (IR: R^2 = 0.703; AR: R^2 = 0.608), and patterns of posterior CI width (*w*) in relation to posterior mean time (*t*) were heteroscedastic (Fig. 10). Posterior mean times (*t*) estimated under the independent rates (IR) model explained 10% more variation in posterior CI widths (*w*) than times estimated under the autocorrelated rates (AR) model explained, which may reflect the better fit of IR model to our phylogenomic data (Tab. 3). Infinite sites theory predicts a perfect correlation of $w \sim t$ when adding more sequence data does not improve accuracy of divergence times. The marginal fit of the linear model to our infinite sites plot (R^2 = 0.704) suggests that posterior estimates of gobiarian divergence times may be improved with additional sequence data (dos Reis et al. 2016).

The slope of infinite sites plot models the relationship of uncertainty in posterior divergence time estimates with respect to increasingly older node age estimates (Inoue et al. 2010). For our phylogenomic data, the slope fit to the infinite site plot under the independent rates (IR) clock model (β = 0.256; 95% CI: 0.218-0.293) indicates that for every 1.0 Ma of species divergence time, we are adding 0.256 Ma of uncertainty to the posterior estimate (Inoue et al. 2010). Interestingly, we found the slope fit under the autocorrelated rates (AR) model was significantly lower (β = 0.170; 95% CI: 0.139-0.200) than the slope fit under the independent rates (IR) model (β = 0.256; 95% CI: 0.218-0.293) based on non-overlapping confidence intervals for regression coefficients estimated in R (Team 2019). The lower slope fit under the autocorrelated rates (AR) model suggests that if sequence data reached saturation, relatively less uncertainty would be added to divergence times estimated under the AR model as compared to times estimated under the independent rates (IR) model. This difference may be characteristic of the autocorrelated rates (AR) model, where the rate drifts over time in Brownian motion. One of the criticisms of the autocorrelated rates (AR) model is that variance in rates increases with time and without bound, and at ancient timescales the mode of the rate distribution is pushed to zero (dos Reis et al. 2016).

2.6 Conclusion

Our phylogenomic analysis with comprehensive sampling of higher taxa supports the current two-order classification of Gobiaria. However, our resolution of collared wrigglers separate from sleepers supports the resurrection of Xenisthmidae. The deep division that we recovered in Eleotridae, and the identity and relationships of sleeper lineages warrants further study with increased taxon sampling.

The divergence timescale we present places origination of major gobiarian clades in the Eocene, and sets diversification of goby lineages in the Oligocene and Miocene. The recent divergence times we estimated for cardinalfishes are similar to other molecular studies, but are at odds with the fossil record of Apogonidae, and warrant further study. Lastly, the ancient vicariance of blind cave gudgeons is not supported by the divergence times we estimated under either relaxed clock model, but the uncertainty we observed in the posterior ages suggests additional taxon sampling may improve estimates.

2.7 Permits

Tyler McCraney was approved by UCLA (ARC # 2015-040-01), National Sun-Yat Sen University, (IACUC # 10409), and National Parks Taiwan (Permit # 1046780767).

2.8 Funding

Tyler McCraney was supported by NSF EAPSI, Society for the Study of Evolution Rosemary Grant, American Society of Ichthyologists and Herpetologists Raney Award, Center for Tropical Research Franklin Grant, and UCLA Ecology and Evolutionary Biology Department Travel and Research Awards.

2.9 Acknowledgements

We are grateful to Jimmy Zheng and Evan McCartney-Melstad for their technical advice in the laboratory, and to Paul Barber, Vanson Liu, and Keryea Soong for their support of fieldwork.

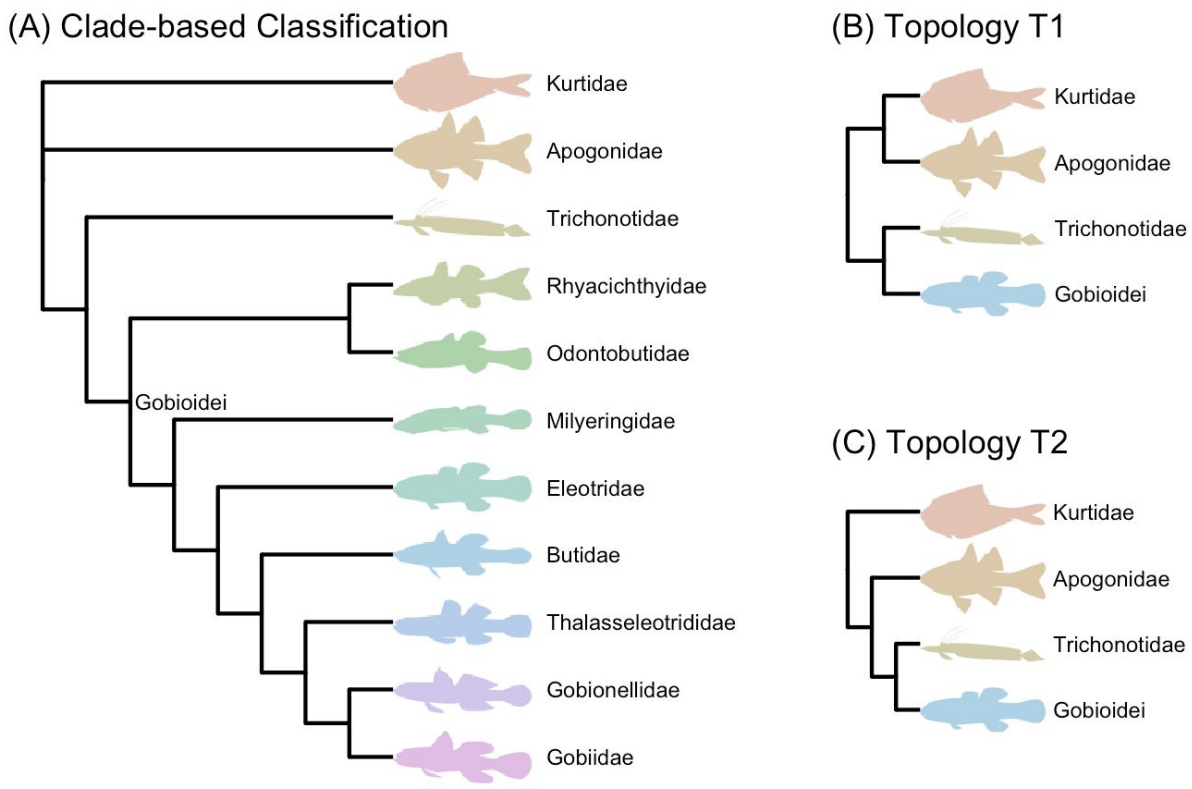


Fig. 1. Consensus hypothesis of gobiarian phylogeny. (A) Current clade-based phylogenetic classification of 11 families (Thacker 2009; Thacker et al. 2015). (B) Topology T1 is supported by multilocus data (Near et al. 2013; Betancur-R. et al. 2017) and resolves Kurtidae + Apogonidae (=Kurtiformes) as a sister clade to Trichonotidae + Gobioidei (=Gobiiformes). (C) Topology T2 is supported by phylogenomic data (Alfaro et al. 2018; Kuang et al. 2018) and resolves Kurtidae as a sister clade to Apogonidae + Gobiiformes.

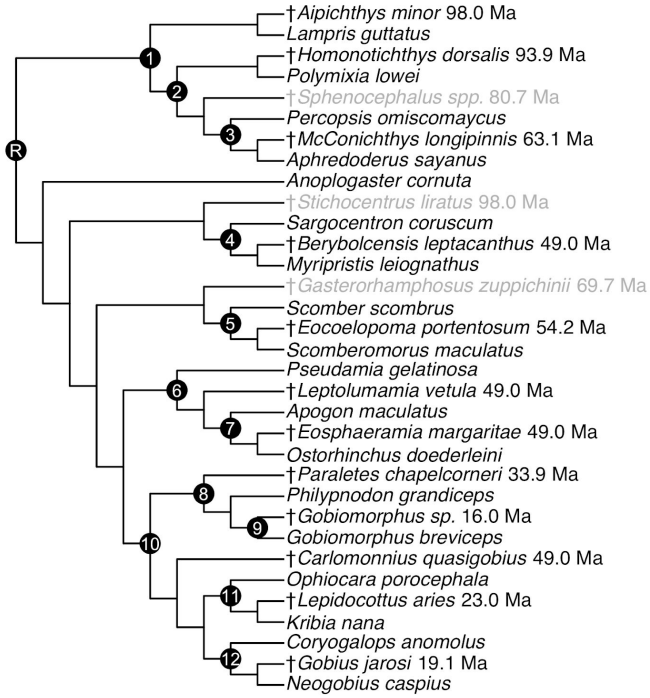
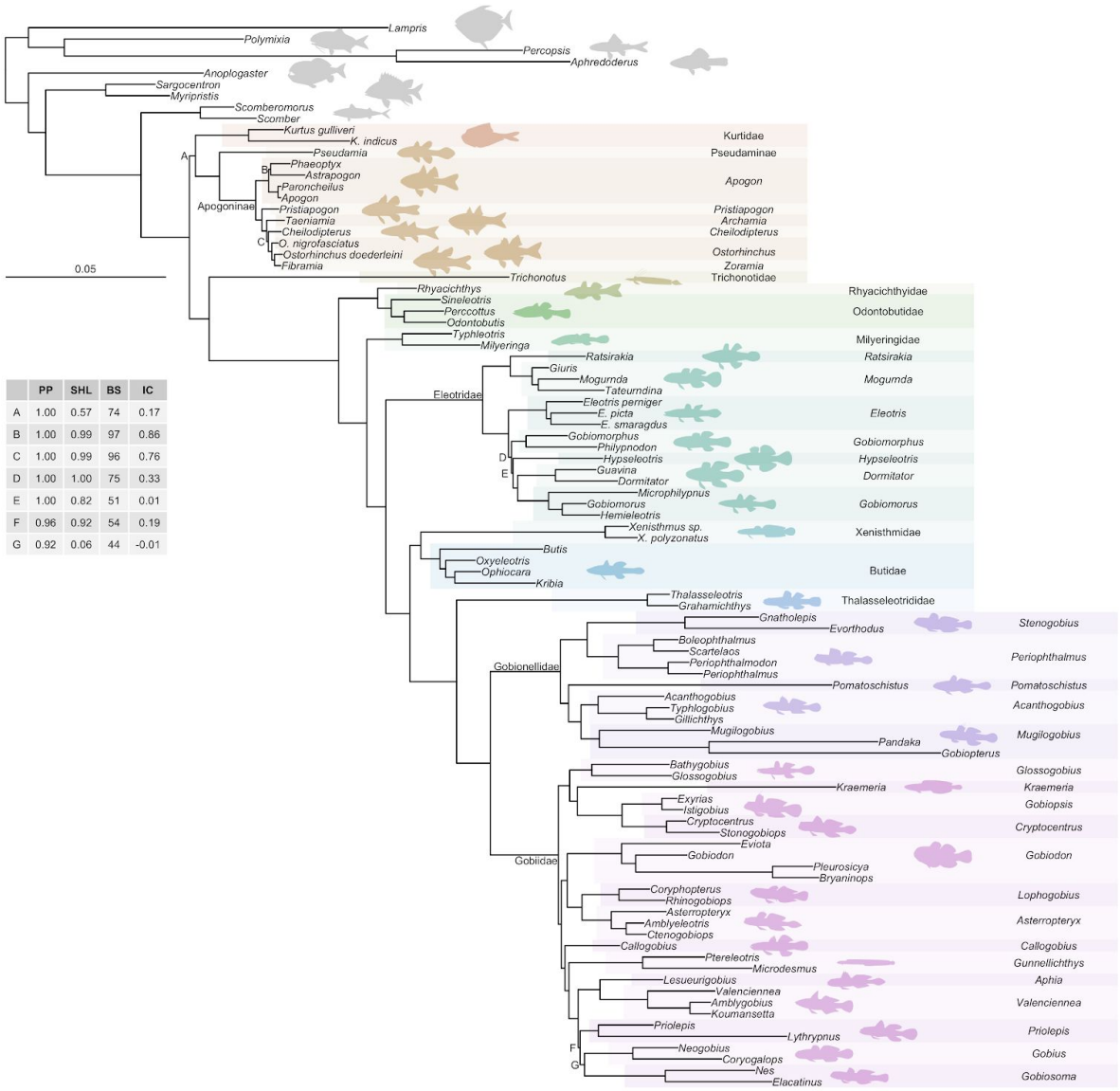


Fig. 2. Phylogenetic placement of fossil calibrations used to constrain priors for Bayesian divergence time estimation under relaxed molecular clocks. Fossil taxa in gray were only used for computing 95% HPD bounds for fossil age constraints, which are provided in Tab. 2. Node labels correspond to calibration points in Tab. 2, Fig. 7, and Fig. 8.



	PP	SHL	BS	IC
A	1.00	0.57	74	0.17
B	1.00	0.99	97	0.86
C	1.00	0.99	96	0.76
D	1.00	1.00	75	0.33
E	1.00	0.82	51	0.01
F	0.96	0.92	54	0.19
G	0.92	0.06	44	-0.01

Fig. 3. Bayesian inference of gobiarian phylogeny on a locus-partitioned, concatenated alignment of 704 ultraconserved element (UCE) loci. Branches scored with less than 100% posterior probability (PP), SH-like support (SHL), bootstrap support (BS), or internode certainty (IC) are labelled, with their support measures provided in the table.

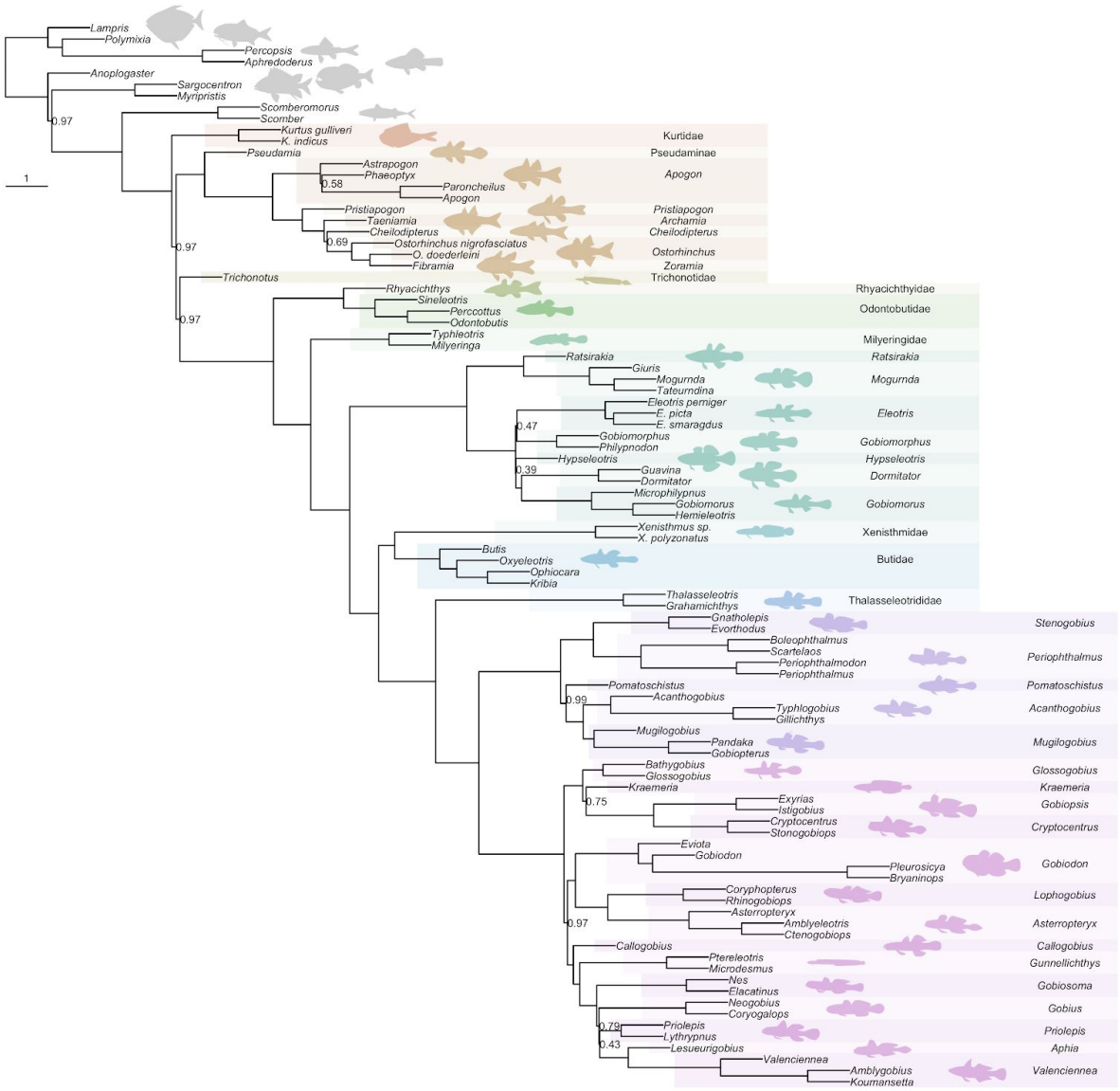


Fig. 4. Multispecies coalescent species tree constructed from 704 contracted maximum likelihood gene trees. Gene tree branches scored with less than 5% SH-like support were collapsed into polytomies before construction of the species tree. Internal branches are scaled by coalescent units and labelled with local posterior probabilities below 100% (i.e., LPP < 1.00).

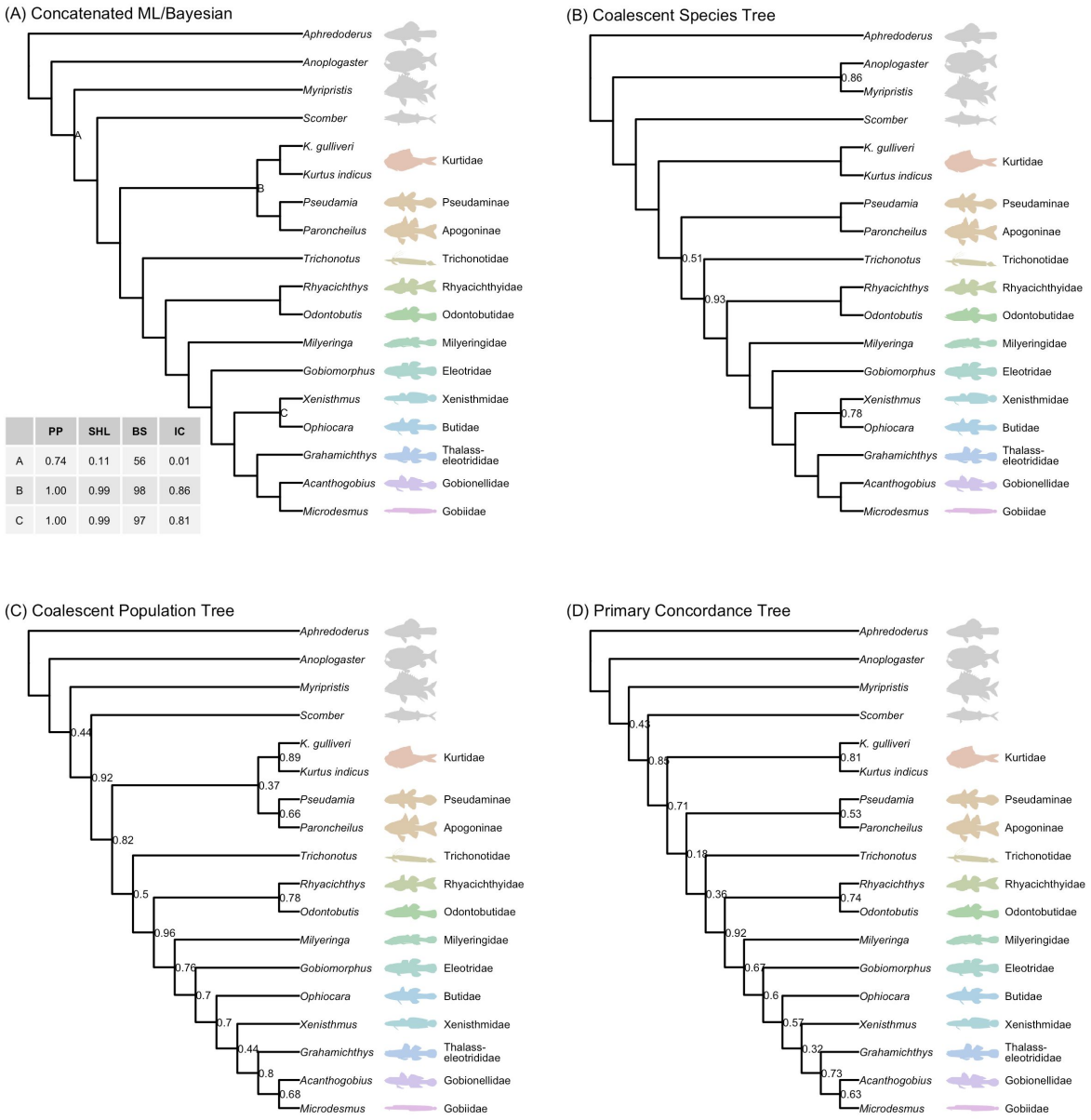


Fig. 5. Trees from the 100% complete backbone matrix of 120 UCE loci. (A) Concatenated maximum likelihood/Bayesian inference. Branches scored with less than 100% posterior probability (PP), SH-like support (SHL), bootstrap support (BS), or internode certainty (IC) are labelled. (B) Multispecies coalescent species tree. Branches are supported by 100% local posterior probability or labelled. (C) Bayesian concordance analysis (coalescent) population tree and (D) primary concordance tree. Branches are labelled with concordance factors.

Fig. 6. Distribution of phylogenetic signal in locus-partitioned, concatenated ultra-conserved element (UCE) sequence alignments. ΔGLS values (*y* axis) were computed by taking the difference in gene-wise log-likelihood scores (GLS) for topology T1 versus topology T2. (A) The phylogenetic signal of 354 UCE loci support topology T1 in the full 94-taxon matrix, and 350 UCEs support topology T2. (B) The phylogenetic signal of 67 UCE loci support topology T1 in the 18-taxon backbone matrix, and 53 UCEs support topology T2.

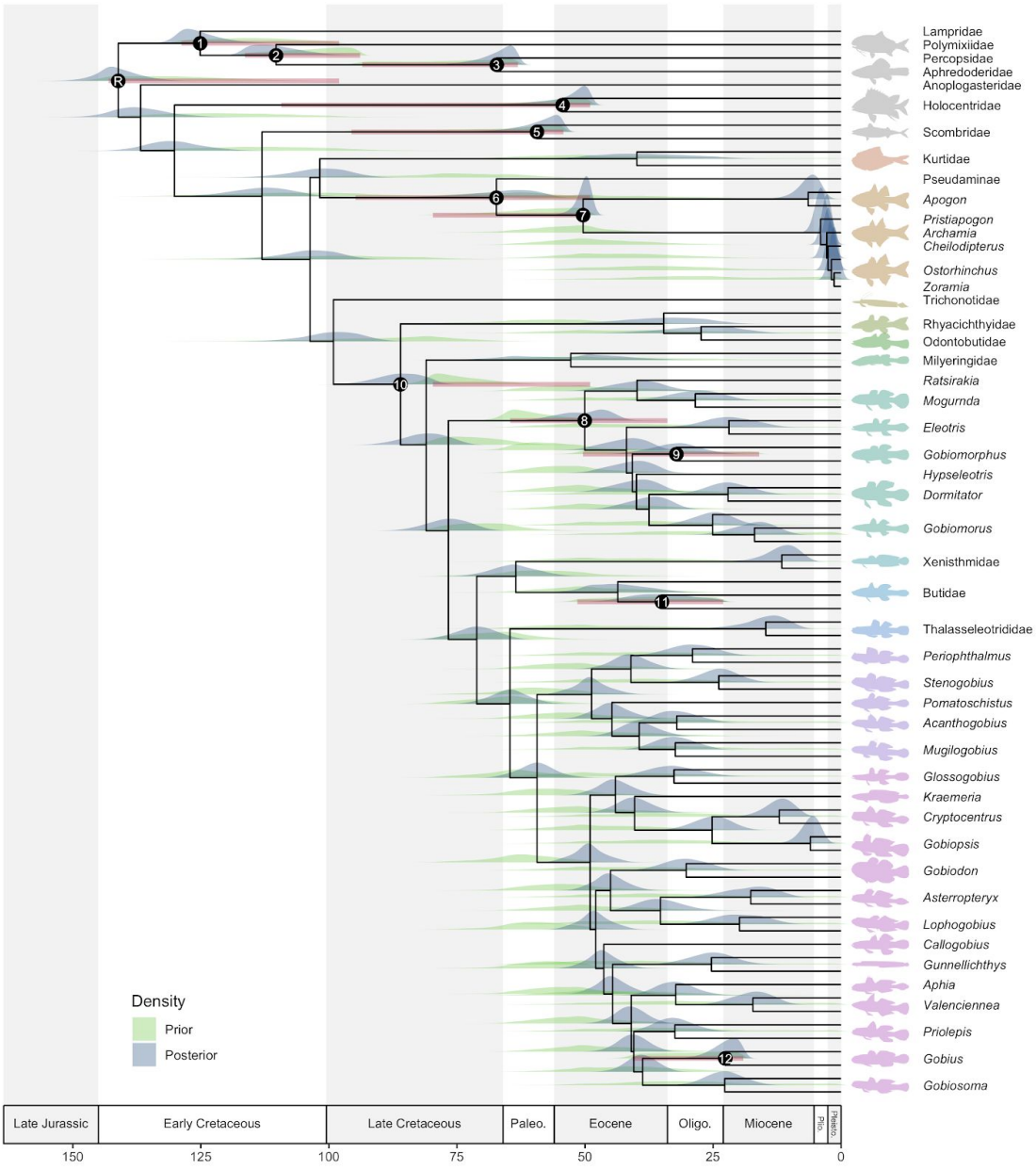


Fig. 7. Bayesian divergence time estimates of Gobiaria under the independent rates (IR) relaxed molecular clock model. Node labels and (red) bars indicate placement and calibration of fossil constraints, respectively.

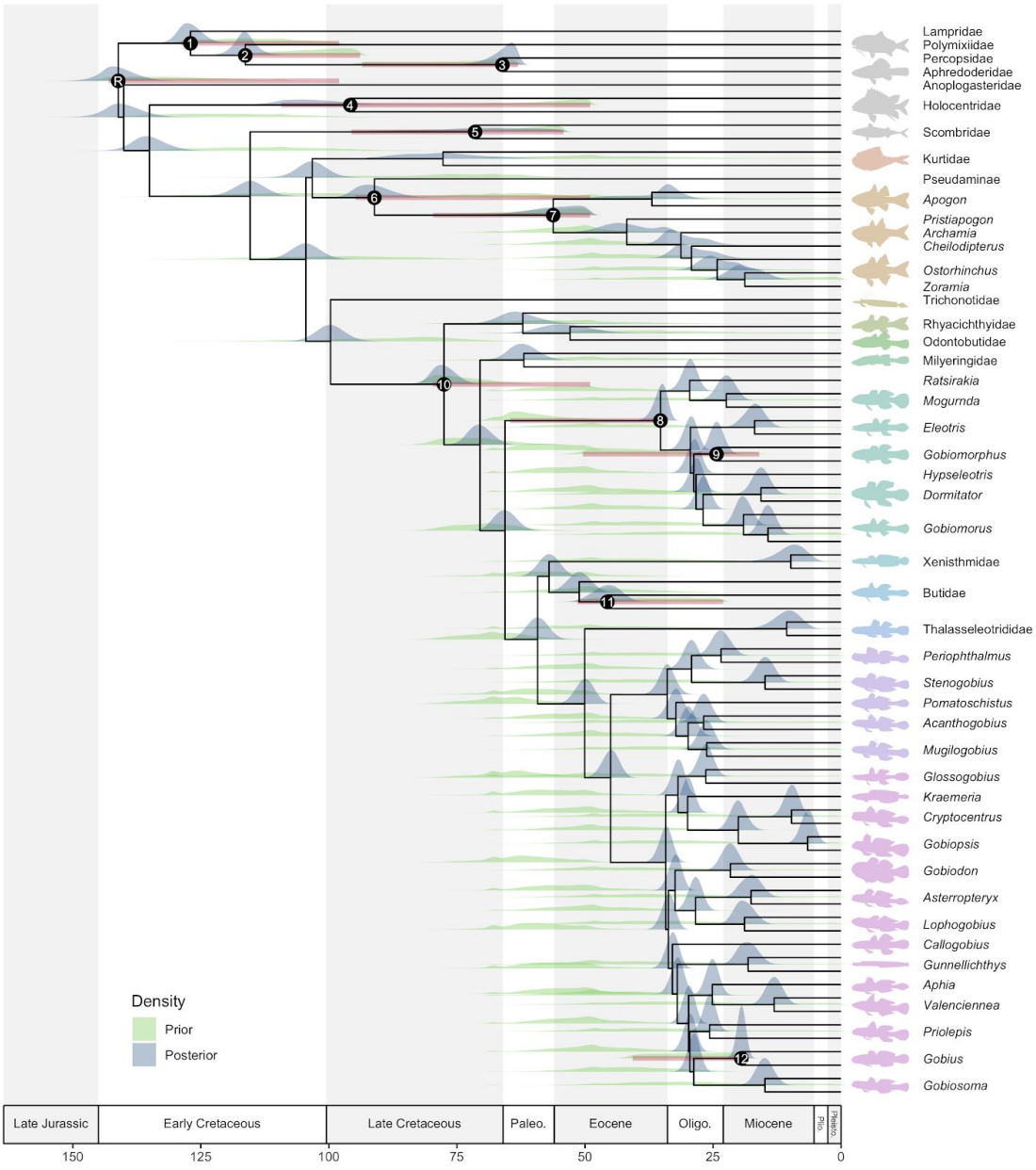


Fig. 8. Bayesian divergence time estimates of Gobiaria under the autocorrelated rates relaxed molecular clock model. Node labels and (red) bars indicate placement and calibration of fossil constraints, respectively.

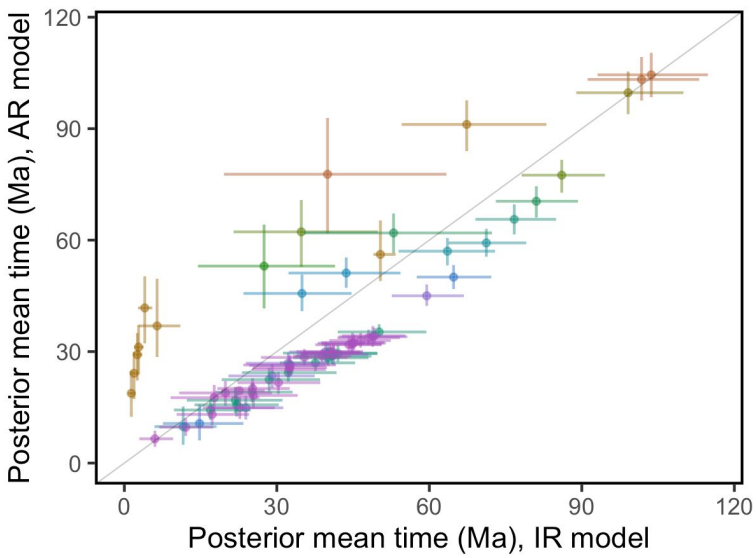


Fig. 9. Ingroup node ages estimated under independent rates (IR) and autocorrelated rates (AR) relaxed molecular clock models. Error bars represent the 95% HPD intervals.

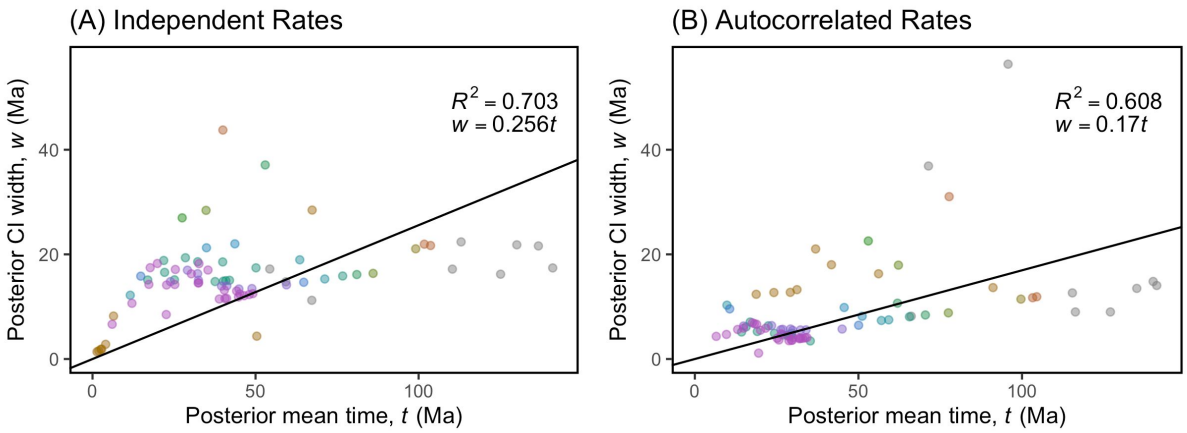


Fig. 10. Infinite sites plots. The credibility interval (CI) width is the difference in 95% HPD bounds. Regression lines were fit through the origin.

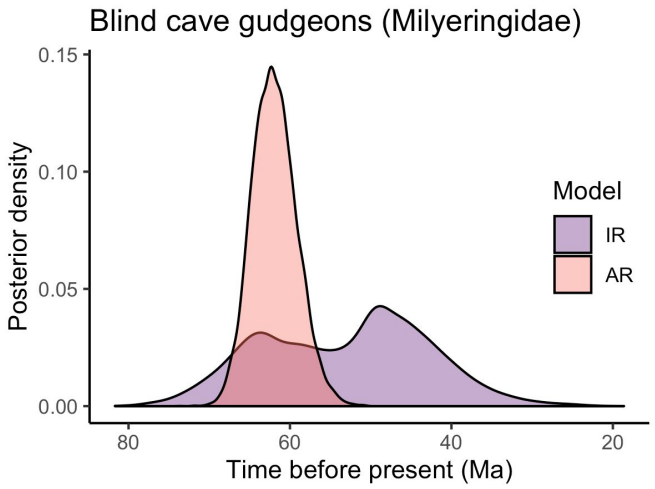


Fig. 11. Bayesian divergence time estimates of blind cave gudgeons (Milyeringidae) under the independent rates (IR) and autocorrelated rates (AR) relaxed molecular clock models.

Tab. 1. Sample information and locus counts from the incomplete matrix (ANSP: American Museum of Natural History; CAS: California Academy of Sciences; FMNH: Florida Museum of Natural History; LACM: Natural History Museum of Los Angeles County; MEA: Michael Alfaro personal collection; NCSM: North Carolina State Museum; SIO, S: Smithsonian Institute; VanTass/Rob: Robert Van Tassel personal collection; WTM: Tyler McCraney personal collection; YFTC: Yale Fish Tissue Collection).

Family	Species	Specimen code	Loci	Flowcell/Reference
	outgroups			
Lampridae	*Lampris guttatus*	SIO 04-195	997	Alfaro et al. 2018
Polymixiidae	*Polymixia lowei*	YFTC 25250	1015	Alfaro et al. 2018
Percopsidae	*Percopsis omiscomaycus*	YFTC 11164	970	Alfaro et al. 2018
Aphredoderidae	*Aphredoderus sayanus*	YFTC 21473	937	Alfaro et al. 2018
Anoplogasteridae	*Anoplogaster cornuta*	YFTC 24906	855	Alfaro et al. 2018
Holocentridae	*Myripristis leiognathus*	MEA 875	995	Alfaro et al. 2018
	Sargocentron coruscum	YFTC 23957	976	Alfaro et al. 2018
Scombridae	*Scomber scombrus*	YFTC 13855	1029	Alfaro et al. 2018
	Scomberomorus maculatus	YFTC 11500	886	Alfaro et al. 2018
	ingroup			
Kurtidae	*Kurtus gulliveri*	YFTC 18514	985	Alfaro et al. 2018
	Kurtus indicus	YFTC 20021	989	NA
Apogonidae	*Paroncheilus affinis*	MEA	879	C13YDACXX
	Fibramia lateralis	YFTC 12659	951	Alfaro et al. 2018
	Apogon maculatus	CAS AMA01	951	Alfaro et al. 2018
	Taeniamia biguttata	LACM 000975	1015	HTCJ7BGXX
	Astrapogon stellatus	YFTC 16204	1010	NA
	Cheilodipterus quinquelineatus	YFTC 12396	979	NA
	Ostorhinchus doederleini	LACM 000974	1016	HTCJ7BGXX
	Ostorhinchus nigrofasciatus	LACM 000970	1032	Alfaro et al. 2018
	Phaeoptyx pigmentaria	LACM 000493	1025	Alfaro et al. 2018
	Pristiapogon kallopterus	YFTC 23090	993	NA
	Pseudamia gelatinosa	S27	934	A44TA
Trichonotidae	*Trichonotus filamentosus*	YFTC 24183	930	NA
Rhyacichthyidae	*Rhyacichthys aspro*	YFTC 25765	988	NA
Odontobutidae	*Odontobutis obscura*	YFTC 11217	1076	Alfaro et al. 2018

	Perccottus glenii	S14; S14	654	A43MM; A44AD
	Sineleotris saccharae	Tuncat	893	HT5WJCCXY
Milyeringidae	*Milyeringa veritas*	ABTC 22891-1	892	HT5WJCCXY
	Typhleotris madagascariensis	FMNH 116498	854	HT5WJCCXY
Eleotridae	*Dormitator maculatus*	LACM 000017	934	HTCJ7BGXX
	Eleotris perniger	NCSM 054870	800	HT5WJCCXY
	Eleotris pisonis	LACM 000019	779	HT5WJCCXY
	Erotelis smaragdus	LACM 000042	965	Alfaro et al. 2018
	Giuris margaritacea	LACM 000073	361	CCBT1ANXX
	Gobiomorphus breviceps	LACM 000064	997	HTCJ7BGXX
	Gobiomorus dormitator	NCSM 068278	844	HT5WJCCXY
	Guavina micropus	VanTass/Rob	726	HT5WJCCXY
	Hemieleotris latifasciata	LACM 000018	750	HT5WJCCXY
	Hypseleotris compressa	LACM 000104	890	HTCJ7BGXX
	Microphilypnus ternetzi	ANSP 180643	823	HT5WJCCXY
	Mogurnda adspersa	LACM 000069	977	Alfaro et al. 2018
	Philypnodon grandiceps	U23	428	CCBT1ANXX
	Ratsirakia lengendrei	LACM 000007	845	HT5WJCCXY
	Tateurndina ocellicauda	LACM 000066	830	HT5WJCCXY
Xenisthmidae	*Xenisthmus polyzonatus*	WTM 021	654	HKVJNDSXX
	Xenisthmus sp.	LACM 000009	768	HT5WJCCXY
Butidae	*Butis butis*	LACM 000045	852	HT5WJCCXY
	Kribia nana	LACM 000077	832	HT5WJCCXY
	Ophiocara porocephala	LACM 000001	1010	Alfaro et al. 2018
	Oxyeleotris lineolata	LACM 000022	976	HTCJ7BGXX
Thalasseleotrididae	*Grahamichthys radiata*	YFTC 25851	803	NA
	Thalasseleotris iota	YFTC 25849	933	NA
Gobionellidae	*Acanthogobius flavimanus*	LACM 000295	971	HTCJ7BGXX
	Boleophthalmus pectinirostris	JACK00000000	794	You et al. 2014
	Evorthodus minutus	LACM 000265	757	Alfaro et al. 2018
	Gillichthys seta	LACM 000281	956	Alfaro et al. 2018
	Gnatholepis cauerensis	LACM 001122	735	C1LVJACXX
	Gobiopterus semivestitus	LACM 000090	697	HT5WJCCXY
	Mugilogobius rivulus	LACM 000262	765	HT5WJCCXY
	Pandaka lidwilli	LACM 000254	890	HTCJ7BGXX
	Periophthalmodon schlosseri	JACM00000000	799	You et al. 2014
	Periophthalmus barbarus	LACM 000273	908	Alfaro et al. 2018

	Pomatoschistus microps	LACM 000984	374	CCBT1ANXX
	Scartelaos histophorus	JACN00000000	800	You et al. 2014
	Typhlogobius californiensis	LACM 000025	936	HTCJ7BGXX
Gobiidae	*Amblyeleotris guttata*	LACM 001025	821	C1LVJACXX
	Amblygobius phaelena	LACM 000195	772	C1LVJACXX
	Asterropteryx ensifera	LACM 001050	948	HTCJ7BGXX
	Bathygobius soporator	LACM 000511	910	Alfaro et al. 2018
	Bryaninops yongei	LACM 000164	813	HT5WJCCXY
	Callogobius sclateri	LACM 000172	954	HTCJ7BGXX
	Coryogalops anomolus	LACM 000243	786	HT5WJCCXY
	Coryphopterus glaucofraenum	LACM 000578	799	C1LVJACXX
	Cryptocentrus cinctus	LACM 001086	929	HTCJ7BGXX
	Ctenogobiops mitodes	WTM 025	777	HKVJNDSXX
	Eviota sigillata	WTM 013	619	HKVJNDSXX
	Exyrias belissimus	LACM 000168	957	HTCJ7BGXX
	Glossogobius flavipinnis	LACM 000311	726	HT5WJCCXY
	Gobiodon histrio	LACM 000239	780	C1LVJACXX
	Elacatinus oceanops	LACM 000230	934	HTCJ7BGXX
	Istigobius rigilius	LACM 001078	763	C1LVJACXX
	Koumansetta rainfordi	S11	809	A44TA
	Kraemeria bryani	LACM 000097	378	CCBT1ANXX
	Lesueurigobius sanzi	OMNZ00000000	792	Malmstrøm et al. 2016
	Lythrypnus dalli	LACM 000190	845	HTCJ7BGXX
	Microdesmus longipinnis	LACM 000617	931	HTCJ7BGXX
	Neogobius melanostomus	LACM 000146	407	CCBT1ANXX
	Nes longus	LACM 000222	780	HT5WJCCXY
	Pleurosicya micheli	WTM 024	633	HKVJNDSXX
	Priolepis cincta	LACM 000904	982	HTCJ7BGXX
	Ptereleotris zebra	LACM 000594	951	HTCJ7BGXX
	Rhinogobiops nicholsii	LACM 001020	447	CCBT1ANXX
	Stonogobiops xanthorhinica	LACM 001049	751	HT5WJCCXY
	Valenciennea strigata	LACM 000898	928	Alfaro et al. 2018

Tab. 2. Fossil calibrations used for constraining priors for Bayesian estimation of relaxed molecular clock divergence times.

Species	Clade	Calibration	Constraint (Ma)	Reference
†*Aipichthys minor*	Acanthopterygii	R	98.0-143.0	Davesne et al. 2014; Delbarre et al. 2016
†*Aipichthys minor*	Paracanthopterygii	1	98.0-128.8	Davesne et al. 2014; Delbarre et al. 2016
†*Homonotichthys dorsalis*	Polymixiiformes + Percopsiformes	2	93.9-116.4	Patterson and White 1964
†*Sphenocephalus spp.*	Percopsiformes	NA	NA	Alfaro et al. 2018
†*McConichthys longipinnis*	Percopsiformes	3	63.1-93.5	Grande 1988; Murray and Wilson 1999
†*Stichocentrus liratus*	Holocentridae	NA	NA	Alfaro et al. 2018
†*Berybolcensis leptacanthus*	Holocentridae	4	49.0-109.3	Stewart 1984; Papazzoni et al. 2014
†*Gasterorhamphosus zuppichinii*	Scombridae	NA	NA	Alfaro et al. 2018
†*Eocoelopoma portentosum*	Scombridae	5	54.2-95.6	Monsch and Bannikov 2012
†*Leptolumamia vetula*	Apogonidae	6	49.0-94.8	Bannikov and Fraser 2016
†*Eosphaeramia margaritae*	Apogoninae	7	49.0-79.7	Bannikov 2008
†*Carlomonnius quasigobius*	Gobioidei	8	49.0-79.7	Bannikov and Carnevale 2016
†*Paralates chapelcorneri*	Eleotridae	9	33.9-64.6	Gierl and Reichenbacher 2017
†*Gobiomorphus sp.*	*Gobiomorphus*	10	16.0-50.4	McDowall et al. 2006
†*Lepidocottus aries*	Butidae	11	23.0-51.5	Gierl et al. 2013
†*Gobius jarosi*	*Gobius*	12	19.1-40.7	Reichenbacher et al. 2018

Tab. 3. Bayesian relaxed clock model selection for randomly selected loci (SC: strict clock; IR: independent rates; AR: autocorrelated rates; *m*L: marginal likelihood).

Partition	Locus	Taxa	Characters	Model	Log *m*L ±SE	Probability
1	uce-30	53	213	SC	-947.7 ±0.11	0.000
				IR	-941.8 ±0.36	0.000
				AR	**-932.2 ±0.39**	**1.000**
1	uce-920	39	178	SC	-866.7 ±0.05	0.000
				IR	**-827.1 ±0.28**	**0.997**
				AR	-833.1 ±0.39	0.003
2	uce-420	54	231	SC	-1272.6 ±0.05	0.000
				IR	**-1214.9 ±0.46**	**1.000**
				AR	-1225.9 ±0.24	0.000
2	uce-1191	47	207	**SC**	**-1106.3 ±0.16**	**1.000**
				IR	-1118.0 ±0.44	0.000
				AR	-1115.1 ±0.47	0.000
3	uce-392	45	204	SC	-1151.9 ±0.16	0.000
				IR	**-1135.9 ±1.03**	**1.000**
				AR	-1149.1 ±0.94	0.000
3	uce-570	49	184	SC	-1275.8 ±0.08	0.000
				IR	**-1224.4 ±0.87**	**1.000**
				AR	-1234.0 ±0.81	0.000
4	uce-253	56	225	SC	-1879.6 ±0.06	0.000
				IR	**-1792.8 ±0.83**	**1.000**
				AR	-1818.0 ±0.61	0.000
4	uce-957	53	240	SC	-2086.0 ±0.08	0.000
				IR	**-1867.0 ±0.94**	**1.000**
				AR	-1899.2 ±0.88	0.000

2.10 References

Aberer, Andre J., Kassian Kobert, and Alexandros Stamatakis. 2014. “ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era.” *Molecular Biology and Evolution* 31 (10): 2553–56.

Agorreta, Ainhoa, and Lukas Rüber. 2012. “A Standardized Reanalysis of Molecular Phylogenetic Hypotheses of Gobioidei.” *Systematics and Biodiversity* 10 (3): 375–90.

Agorreta, Ainhoa, Diego San Mauro, Ulrich Schliewen, James L. Van Tassell, Marcelo Kovačić, Rafael Zardoya, and Lukas Rüber. 2013. “Molecular Phylogenetics of Gobioidei and Phylogenetic Placement of European Gobies.” *Molecular Phylogenetics and Evolution* 69 (3): 619–33.

Akihito, A. Iwata, T. Kobayashi, K. Ikeo, T. Imanishi, H. Ono, Y. Umehara, et al. 2000. “Evolutionary Aspects of Gobioid Fishes Based upon a Phylogenetic Analysis of Mitochondrial Cytochrome B Genes.” *Gene* 259 (1-2): 5–15.

Alfaro, Michael E., Brant C. Faircloth, Richard C. Harrington, Laurie Sorenson, Matt Friedman, Christine E. Thacker, Carl H. Oliveros, David Černý, and Thomas J. Near. 2018. “Explosive Diversification of Marine Fishes at the Cretaceous–Palaeogene Boundary.” *Nature Ecology & Evolution*, March. https://doi.org/10.1038/s41559-018-0494-6.

Alfaro, Michael E., Francesco Santini, Chad Brock, Hugo Alamillo, Alex Dornburg, Daniel L. Rabosky, Giorgio Carnevale, and Luke J. Harmon. 2009. “Nine Exceptional Radiations plus High Turnover Explain Species Diversity in Jawed Vertebrates.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (32): 13410–14.

Ané, Cécile, Bret Larget, David A. Baum, Stacey D. Smith, and Antonis Rokas. 2007. “Bayesian Estimation of Concordance among Gene Trees.” *Molecular Biology and Evolution* 24 (2): 412–26.

Bannikov, Alexandre F. 2008. “Revision of Some Eocene Fishes from Bolca, Northern Italy, Previously Classified with the Apogonidae and Enoplosidae (Perciformes).” *Studi E Ricerche Sui Giacimenti Terziari Di Bolca, XII Miscellanea Paleontologica* 9: 65–76.

Bannikov, Alexandre F., and Giorgio Carnevale. 2016. “Carlomonnius Quasigobius Gen. et Sp. Nov.: The First Gobioid Fish from the Eocene of Monte Bolca, Italy.” *Bulletin of Geosciences* 91 (1): 13–22.

Bannikov, Alexandre F., and Thomas F. Fraser. 2016. “A New Genus and Species of Cardinalfi Sh (Percomorpha, Apogonidae) from the Eocene of Bolca, Northern Italy (Monte Postale Site).” 2016. https://www.researchgate.net/profile/A_Bannikov/publication/311913539_A_new_genus_and_species_of_cardinalfi_sh_Percomorpha_Apogonidae_from_the_Eocene_of_Bolca_northern_Italy_Monte_Postale_site/links/586243a808ae8fce490962de/A-new-genus-and-species-of-cardinalfi-sh-Percomorpha-Apogonidae-from-the-Eocene-of-Bolca-northern-Italy-Monte-Postale-site.pdf.

Betancur-R., Ricardo, Richard E. Broughton, Edward O. Wiley, Kent Carpenter, J. Andrés López, Chenhong Li, Nancy I. Holcroft, et al. April 18, 2013June 15, 2013February 4, 2017. “The Tree of Life and a New Classification of Bony FishesCall for SubmissionsAVAToL CollectionBlog PostTwitter.” *PLoS Currents*. https://doi.org/10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, Ricardo, Edward O. Wiley, Gloria Arratia, Arturo Acero, Nicolas Bailly, Masaki Miya, Guillaume Lecointre, and Guillermo Ortí. 2017. “Phylogenetic Classification of Bony Fishes.” *BMC Evolutionary Biology* 17 (1): 162.

Birdsong, Ray S., Edward O. Murdy, and Frank L. Pezold. 1988. “A Study of the Vertebral Column and Median Fin Osteology in Gobioid Fishes with Comments on Gobioid

Relationships." *Bulletin of Marine Science* 42 (2): 174–214.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

Bruyn, Mark de, Björn Stelbrink, Timothy J. Page, Matthew J. Phillips, David J. Lohman, Christian Albrecht, Robert Hall, et al. 2013. "Time and Space in Biogeography: Response to Parenti & Ebach (2013)." Edited by Liliana Katinas. *Journal of Biogeography* 40 (11): 2204–6.

Bushnell, B. 2018. "BBTools: A Suite of Fast, Multithreaded Bioinformatics Tools Designed for Analysis of DNA and RNA Sequence Data." *Joint Genome Institute. Https://jgi. Doe. Gov/data-and-Tools/bbtools*.

Capella-Gutiérrez, Salvador, José M. Silla-Martínez, and Toni Gabaldón. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25 (15): 1972–73.

Chakerian, John, and Susan Holmes. 2013. "DISTORY: Distance between Phylogenetic Histories." *R Package Version* 1 (2).

Chakrabarty, Prosanta, Matthew P. Davis, and John S. Sparks. 2012. "The First Record of a Trans-Oceanic Sister-Group Relationship between Obligate Vertebrate Troglobites." *PloS One* 7 (8): e44083.

Cowman, P. F., and D. R. Bellwood. 2011. "Coral Reefs as Drivers of Cladogenesis: Expanding Coral Reefs, Cryptic Extinction Events, and the Development of Biodiversity Hotspots." *Journal of Evolutionary Biology* 24 (12): 2543–62.

Davesne, Donald, Matt Friedman, Véronique Barriel, Guillaume Lecointre, Philippe Janvier, Cyril Gallut, and Olga Otero. 2014. "Early Fossils Illuminate Character Evolution and Interrelationships of Lampridiformes (Teleostei, Acanthomorpha)." *Zoological Journal of the*

Linnean Society 172 (2): 475–98.

Delbarre, Daniel J., Donald Davesne, and Matt Friedman. 2016. “Anatomy and Relationships of †Aipichthys Pretiosus and †‘Aipichthys’ Nuchalis (Acanthomorpha: Lampridomorpha), with a Review of Late Cretaceous Relatives of Oarfishes and Their Allies.” *Journal of Systematic Palaeontology* 14 (7): 545–67.

Eschmeyer, W. N., R. Fricke, and R. Van der Laan. 2016. “Catalog of Fishes: Genera, Species, References.” *Electronic Version Accessed* 19. https://www.researchgate.net/profile/Richard_Van_Der_Laan2/publication/303633861_Catalog_of_Fishes_Genera_Species_References/links/574ed78708ae789584d80783/Catalog-of-Fishes-Genera-Species-References.pdf.

Faircloth, B. C. 2013. “Illumiprocessor: A Trimmomatic Wrapper for Parallel Adapter and Quality Trimming.” *See Http://dx. Doi. org/10. 6079/J9ILL (accessed 4 November 2016)*.

Faircloth, Brant C. 2016. “PHYLUCE Is a Software Package for the Analysis of Conserved Genomic Loci.” *Bioinformatics* 32 (5): 786–88.

Faircloth, Brant C., Michael G. Branstetter, Noor D. White, and Seán G. Brady. 2015. “Target Enrichment of Ultraconserved Elements from Arthropods Provides a Genomic Perspective on Relationships among Hymenoptera.” *Molecular Ecology Resources* 15 (3): 489–501.

Faircloth, Brant C., Laurie Sorenson, Francesco Santini, and Michael E. Alfaro. 2013. “A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs).” *PloS One* 8 (6): e65923.

Gierl, Christoph, and Bettina Reichenbacher. 2017. “Revision of so-Called Pomatoschistus (Gobiiformes, Teleostei) from the Late Eocene and Early Oligocene.” *Palaeontologia Electronica* 20 (2): 1–17.

Gierl, Christoph, Bettina Reichenbacher, Jean Gaudant, Dirk Erpenbeck, and André Pharisat.

2013. “An Extraordinary Gobioid Fish Fossil from Southern France.” *PloS One* 8 (5): e64117.

Glenn, Travis C., Roger A. Nilsen, Troy J. Kieran, Jon G. Sanders, Natalia J. Bayona-Vásquez, John W. Finger, Todd W. Pierson, et al. 2019. “Adapterama I: Universal Stubs and Primers for 384 Unique Dual-Indexed or 147,456 Combinatorially-Indexed Illumina Libraries (iTru & iNext).” *PeerJ* 7 (October): e7755.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, et al. 2011. “Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome.” *Nature Biotechnology* 29 (7): 644–52.

Grande, Lance. 1988. “A Well Preserved Paracanthopterygian Fish (Teleostei) from Freshwater Lower Paleocene Deposits of Montana.” *Journal of Vertebrate Paleontology* 8 (2): 117–30.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.” *Systematic Biology* 59 (3): 307–21.

Hasegawa, M., H. Kishino, and T. Yano. 1985. “Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA.” *Journal of Molecular Evolution* 22 (2): 160–74.

Hedman, Matthew M. 2010. “Constraints on Clade Ages from Fossil Outgroups.” *Paleobiology* 36 (1): 16–31.

Hoese, Douglass F., and Anthony Gill. 1993. “Phylogenetic Relationships of Eleotridid Fishes (Perciformes: Gobioidei),” January. http://www.ingentaconnect.com/content/umrsmas/bullmar/1993/00000052/00000001/art00015.

Hughes, Lily C., Guillermo Ortí, Yu Huang, Ying Sun, Carole C. Baldwin, Andrew W. Thompson,

Dahiana Arcila, et al. 2018. “Comprehensive Phylogeny of Ray-Finned Fishes (Actinopterygii) Based on Transcriptomic and Genomic Data.” *Proceedings of the National Academy of Sciences of the United States of America*, May. https://doi.org/10.1073/pnas.1719358115.

Inoue, Jun, Philip C. J. Donoghue, and Ziheng Yang. 2010. “The Impact of the Representation of Fossil Calibrations on Bayesian Estimation of Species Divergence Times.” *Systematic Biology* 59 (1): 74–89.

Junier, Thomas, and Evgeny M. Zdobnov. 2010. “The Newick Utilities: High-Throughput Phylogenetic Tree Processing in the UNIX Shell.” *Bioinformatics* 26 (13): 1669–70.

Katoh, Kazutaka, and Daron M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80.

Kuang, Ting, Luke Tornabene, Jingyan Li, Jiamei Jiang, Prosanta Chakrabarty, John S. Sparks, Gavin J. P. Naylor, and Chenhong Li. 2018. “Phylogenomic Analysis on the Exceptionally Diverse Fish Clade Gobioidei (Actinopterygii: Gobiiformes) and Data-Filtering Based on Molecular Clocklikeness.” *Molecular Phylogenetics and Evolution* 128 (November): 192–202.

Larget, Bret R., Satish K. Kotha, Colin N. Dewey, and Cécile Ané. 2010. “BUCKy: Gene Tree/species Tree Reconciliation with Bayesian Concordance Analysis.” *Bioinformatics* 26 (22): 2910–11.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60.

Li, Hongjie, You He, Jiamei Jiang, Zhizhi Liu, and Chenhong Li. 2017. “Molecular Systematics and Phylogenetic Analysis of the Asian Endemic Freshwater Sleepers (Gobiiformes:

Odontobutidae).” *Molecular Phylogenetics and Evolution* 121 (December): 1–11.

Mabuchi, Kohji, Thomas H. Fraser, Hayeun Song, Yoichiro Azuma, and Mutsumi Nishida. 2014. “Revision of the Systematics of the Cardinalfishes (Percomorpha: Apogonidae) Based on Molecular Analyses and Comparative Reevaluation of Morphological Characters.” *Zootaxa* 3846 (2): 151–203.

Maddison, Wayne P. 1997. “Gene Trees in Species Trees.” *Systematic Biology* 46 (3): 523–36.

Makowski, Dominique, Mattan Ben-Shachar, and Daniel Lüdecke. 2019. “bayestestR: Describing Effects and Their Uncertainty, Existence and Significance within the Bayesian Framework.” *Journal of Open Source Software* 4 (40): 1541.

Malmstrøm, Martin, Michael Matschiner, Ole K. Tørresen, Bastiaan Star, Lars G. Snipen, Thomas F. Hansen, Helle T. Baalsrud, et al. 2016. “Evolution of the Immune System Influences Speciation Rates in Teleost Fishes.” *Nature Genetics* 48 (10): 1204–10.

McDowall, R. M., E. M. Kennedy, J. K. Lindqvist, D. E. Lee, B. V. Alloway, and M. R. Gregory. 2006. “Probable Gobiomorphus Fossils from the Miocene and Pleistocene of New Zealand (Teleostei: Eleotridae).” *Journal of the Royal Society of New Zealand* 36 (3): 97–109.

Monsch, Kenneth A., and Alexandre F. Bannikov. 2012. “New Taxonomic Synopses and Revision of the Scombroid Fishes (Scombroidei, Perciformes), Including Billfishes, from the Cenozoic of Territories of the Former USSR.” *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 102: 253–300.

Murray, Alison M., and Mark V. H. Wilson. 1999. “Contributions of Fossils to the Phylogenetic Relationships of the Percopsiform Fishes (Teleostei: Paracanthopterygii): Order Restored.” https://www.pfeil-verlag.de/wp-content/uploads/2015/05/2_48d20.pdf.

Near, Thomas J., Alex Dornburg, Ron I. Eytan, Benjamin P. Keck, W. Leo Smith, Kristen L. Kuhn, Jon A. Moore, et al. 2013. “Phylogeny and Tempo of Diversification in the

Superradiation of Spiny-Rayed Fishes." *Proceedings of the National Academy of Sciences of the United States of America* 110 (31): 12738–43.

Nelson, Joseph S., Terry C. Grande, and Mark V. H. Wilson. 2016. "Phylum Chordata." In *Fishes of the World*, 13–526. John Wiley & Sons, Inc.

Papazzoni, Cesare Andrea, Giorgio Carnevale, Eliana Fornaciari, Luca Giusberti, and Enrico Trevisani. 2014. "The Pesciara-Monte Postale Fossil-Lagerstätte: 1. Biostratigraphy, Sedimentology and Depositional Model." *The Bolca Fossil-Lagerstätte: A Window into the Eocene World: Rendiconti Della Società Paleontologica Italiana* 4: 29–36.

Patterson, Colin, and Errol Ivor White. 1964. "A Review of Mesozoic Acanthopterygian Fishes, with Special Reference to Those of the English Chalk." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 247 (739): 213–482.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6 (1): 7–11.

Rabosky, Daniel L., Jonathan Chang, Pascal O. Title, Peter F. Cowman, Lauren Sallan, Matt Friedman, Kristin Kaschner, et al. 2018. "An Inverse Latitudinal Gradient in Speciation Rate for Marine Fishes." *Nature*, July. https://doi.org/10.1038/s41586-018-0273-1.

Rabosky, Daniel L., Francesco Santini, Jonathan Eastman, Stephen A. Smith, Brian Sidlauskas, Jonathan Chang, and Michael E. Alfaro. 2013. "Rates of Speciation and Morphological Evolution Are Correlated across the Largest Vertebrate Radiation." *Nature Communications* 4: 1958.

Rambaut, A., and A. J. Drummond. 2014. "LogCombiner v2. 1.3." *Institute of Evolutionary Biology, University of Edinburgh, UK*.

Reichenbacher, Bettina, Růžena Gregorová, Katarína Holcová, Radek Šanda, Jasna Vukić, and Tomáš Přikryl. 2018. "Discovery of the Oldest Gobius (Teleostei, Gobiiformes) from a

Marine Ecosystem of Early Miocene Age." *Journal of Systematic Palaeontology* 16 (6): 493–513.

Reis, Mario dos, Philip C. J. Donoghue, and Ziheng Yang. 2016. "Bayesian Molecular Clock Dating of Species Divergences in the Genomics Era." *Nature Reviews. Genetics* 17 (2): 71–80.

Reis, Mario Dos, Gregg F. Gunnell, Jose Barba-Montoya, Alex Wilkins, Ziheng Yang, and Anne D. Yoder. 2018. "Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case." *Systematic Biology* 67 (4): 594–615.

Reis, Mario dos, and Ziheng Yang. 2011. "Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times." *Molecular Biology and Evolution* 28 (7): 2161–72.

Reis Sandra Álvarez-Carretero Ziheng Yang, Mario dos. 2017. "MCMCTree Tutorials." http://abacus.gene.ucl.ac.uk/software/MCMCtree.Tutorials.pdf.

Rohland, Nadin, and David Reich. 2012. "Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture." *Genome Research* 22 (5): 939–46.

Salichos, Leonidas, Alexandros Stamatakis, and Antonis Rokas. 2014. "Novel Information Theory-Based Measures for Quantifying Incongruence among Phylogenetic Trees." *Molecular Biology and Evolution* 31 (5): 1261–71.

Shen, Xing-Xing, Chris Todd Hittinger, and Antonis Rokas. 2017. "Contentious Relationships in Phylogenomic Studies Can Be Driven by a Handful of Genes." *Nature Ecology & Evolution* 1 (5): 126.

Shen, Xing-Xing, Leonidas Salichos, and Antonis Rokas. 2016. "A Genome-Scale Investigation of How Sequence, Function, and Tree-Based Gene Properties Influence Phylogenetic

Inference.” *Genome Biology and Evolution* 8 (8): 2565–80.

Smith, Stephen A., and Casey W. Dunn. 2008. “Phyutility: A Phyloinformatics Tool for Trees, Alignments and Molecular Data.” *Bioinformatics* 24 (5): 715–16.

Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–13.

Stewart, Joe D. 1984. “Taxonomy, Paleoecology, and Stratigraphy of the Halecostome-Inoceramid Associations of the North American Upper Cretaceous Epicontinental Seaways.” University of Kansas, Systematics and Ecology.

Tavaré, Simon. 1986. “Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences.” *Lectures on Mathematics in the Life Sciences* 17 (2): 57–86.

Team, R. Core. 2019. “A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.” *URL Https://www. R-Project. Org*.

Thacker, Christine E. 2003. “Molecular Phylogeny of the Gobioid Fishes (Teleostei: Perciformes: Gobioidei).” *Molecular Phylogenetics and Evolution* 26 (3): 354–68.

Thacker, Christinee. 2009. “Phylogeny of Gobioidei and Placement within Acanthomorpha, with a New Classification and Investigation of Diversification and Character Evolution.” *Copeia* 2009 (1): 93–104.

Thacker, Christine E. 2013. “Phylogenetic Placement of the European Sand Gobies in Gobionellidae and Characterization of Gobionellid Lineages (Gobiiformes: Gobioidei).” *Zootaxa* 3619: 369–82.

———. 2017. “Patterns of Divergence in Fish Species Separated by the Isthmus of Panama.” *BMC Evolutionary Biology* 17 (1): 111.

Thacker, Christine E., and Michael A. Hardman. 2005. “Molecular Phylogeny of Basal Gobioid Fishes: Rhyacichthyidae, Odontobutidae, Xenisthmidae, Eleotridae (Teleostei:

Perciformes: Gobioidei).” *Molecular Phylogenetics and Evolution* 37 (3): 858–71.

Thacker, Christine E., and Dawn M. Roje. 2011. “Phylogeny of Gobiidae and Identification of Gobiid Lineages.” *Systematics and Biodiversity* 9 (4): 329–47.

Thacker, Christine E., Takashi P. Satoh, Eri Katayama, Richard C. Harrington, Ron I. Eytan, and Thomas J. Near. 2015. “Molecular Phylogeny of Percomorpha Resolves Trichonotus as the Sister Lineage to Gobioidei (Teleostei: Gobiiformes) and Confirms the Polyphyly of Trachinoidei.” *Molecular Phylogenetics and Evolution* 93 (December): 172–79.

Wickham, Hadley. 2017. “Tidyverse: Easily Install and Load the ‘Tidyverse.’” https://CRAN.R-project.org/package=tidyverse.

Xie, Wangang, Paul O. Lewis, Yu Fan, Lynn Kuo, and Ming-Hui Chen. 2011. “Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection.” *Systematic Biology* 60 (2): 150–60.

Yang, Z. 1994. “Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods.” *Journal of Molecular Evolution* 39 (3): 306–14.

Yang, Ziheng. 2007. “PAML 4: Phylogenetic Analysis by Maximum Likelihood.” *Molecular Biology and Evolution* 24 (8): 1586–91.

You, Xinxin, Chao Bian, Qijie Zan, Xun Xu, Xin Liu, Jieming Chen, Jintu Wang, et al. 2014. “Mudskipper Genomes Provide Insights into the Terrestrial Adaptation of Amphibious Fishes.” *Nature Communications* 5 (December): 5594.

Zhang, Chao, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. 2018. “ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees.” *BMC Bioinformatics* 19 (Suppl 6): 153.