Minnows and molecules: resolving the broad and fine-scale evolutionary patterns of Cypriniformes

by

Carla Cristina Stout

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 7, 2017

Keywords: fish, phylogenomics, population genetics, Leuciscidae, sequence capture

Approved by

Jonathan W. Armbruster, Chair, Professor of Biological Sciences and Curator of Fishes
Jason E. Bond, Professor and Department Chair of Biological Sciences
Scott R. Santos, Professor of Biological Sciences
Eric Peatman, Associate Professor of Fisheries, Aquaculture, and Aquatic Sciences

Abstract

Cypriniformes (minnows, carps, loaches, and suckers) is the largest group of freshwater fishes in the world. Despite much attention, previous attempts to elucidate relationships using molecular and morphological characters have been incongruent. The goal of this dissertation is to provide robust support for relationships at various taxonomic levels within Cypriniformes. For the entire order, an anchored hybrid enrichment approach was used to resolve relationships. This resulted in a phylogeny that is largely congruent with previous multilocus phylogenies, but has much stronger support. For members of Leuciscidae, the relationships established using anchored hybrid enrichment were used to estimate divergence times in an attempt to make inferences about their biogeographic history. The predominant lineage of the leuciscids in North America were determined to have entered North America through Beringia ~37 million years ago while the ancestor of the Golden Shiner (Notemigonus crysoleucas) entered ~20–6 million years ago, likely from Europe. Within Leuciscidae, the shiner clade represents genera with much historical taxonomic turbidity. Targeted sequence capture was used to establish relationships in order to inform taxonomic revisions for the clade. Presented is a revised, genus-level taxonomy for the group. Finally, for Notropis longirostris (now Miniellus longirostris), genetic analyses using mtDNA found four distinct, unconnected haplotype networks across its southeastern USA range with high genetic divergence, despite a lack of morphological differentiation.

Acknowledgments

I would like to thank my advisor, Jonathan W. Armbruster, for his intellectual and financial support over the course of this research. I am thankful for the feedback and instruction provided by my committee members Jason Bond, Scott Santos, and Eric Peatman, both in courses and in research. I am indebted to those who collected and provided animal tissues, including Richard Mayden and Susana Schönhuth at St. Louis University, David C. Werneke at Auburn Museum of Natural History, Kyle Piller at Southeastern Louisiana University, and Morgan Raley, without whom these projects would have never come to fruition.

I am grateful for the friendships and interactions with the many graduate students I have had the opportunity to meet while at Auburn, especially Shobnom Ferdous, C. Keith Ray, Milton Tan, Malorie Hayes, Edward D. Burress, Pamela B. Hart, Rebecca L. Godwin, Nicole L. Garrison, Kellie Bourguignon, and Charles Stephen. I would also like to thank the biological sciences department for providing teaching opportunities, both in labs and lectures.

I could not have maintained balance in my life without the support of my parents, sisters, and nephews. They, along with Brandon King and Attila Joo provided continued encouragement that has made an impact in the pursuit of my academic goals.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AE	Anchored enrichment
ASC	Alabama Supercomputer Center
AU	approximately-unbiased
BBM	Bayesian binary Markov Chain Monte Carlo
bp	base pairs
bs	bootstrap
DEC	Dispersal-extinction cladogenesis
COI	cytochrome oxidase subunit I
CC-P	Creek Chub-Plagopterin
CT	Concatenated Tree
CVA	Canonical Variates Analysis
GPA	General Procrustes Analysis
HPD	Highest Posterior Density
MCC	maximum clade credibility
ML	Maximum Likelihood
mya	million years ago
OPM	Open Posterior Myodome

PCA	Principal Components Analysis
ST	Species Tree
WNA	Western North America

UNIFYING THEME FOR DISSERTATION RESEARCH

Cypriniformes (minnows, carps, loaches, and suckers) constitutes a widely diverse clade of freshwater fishes found in Asia, Europe, Africa, and North America. There are approximately 4300 described species, and with such diversity and distribution, the order presents itself as a prime study group to investigate evolutionary patterns at various taxonomic levels. The size of the order and rapid diversification events within subclades have unfortunately led to difficulty in establishing relationships among the taxa, thus hampering subsequent studies of evolutionary processes. The goals of this dissertation are to first produce strong phylogenetic hypotheses at various taxonomic levels, not only for taxonomic clarity, but also for subsequent investigation into biologically relevant patterns. The first chapter utilizes anchored hybrid enrichment methods developed by Alan and Emily Moriarty Lemmon to produce a strongly supported phylogeny for the entire order. The second chapter then uses this phylogeny to examine more closely the biogeography and diversification dates of a family within Cypriniformes, Leuciscidae. The third chapter delves even deeper to examine relationships among *Notropis* and related shiners within Leuciscidae, this time using an exon capture method developed by the NSF-funded FishLife project headed by Guillermo Orti, Gavin Naylor, Ricardo Betancur, and Carole Baldwin. The fourth chapter uses genetic and geometric morphometric techniques to examine patterns among the populations of *Miniellus longirostris* (formerly *Notropis longirostris*). This dissertation illustrates the importance of employing different genetic and morphological strategies to address questions about evolutionary relationships and processes, spanning the entire order down to the

populations of a single species, and highlights the utility of Cypriniformes for answering an almost infinite number of biological questions.

CHAPTER ONE

RESOLVING CYPRINIFORMES RELATIONSHIPS USING AN ANCHORED ENRICHMENT APPROACH

NOTE: This chapter was a collaborative effort among myself, Milton Tan, Alan Lemmon, Emily Lemmon, and Jonathan W. Armbruster. This study was published in BMC Evolutionary Biology, November 2016. Portions regarding Danionidae are excluded here and can be found in Milton Tan's dissertation and in the published version.

INTRODUCTION

Cypriniformes (minnows, carps, loaches, and suckers) is the largest group of freshwater fishes in the world. Diversity ranges from some of the smallest vertebrates in the world (*Paedocypris*, 7.9 mm standard length) to members of *Tor* (almost 3 m standard length) (Mayden and W. J. Chen, 2010). The number of valid species is currently estimated at around 4300 (Eschmeyer and Fong, 2016) with as many as 2500 still awaiting description (Mayden et al., 2009). To place the Cypriniformes into perspective, about one third of freshwater fish species is a cypriniform and about 6% of all vertebrate species is a cypriniform (Eschmeyer and Fong, 2016). Species of Cypriniformes are distributed in freshwater habitats across Asia, Europe, Africa, and North America (Saitoh et al., 2011). Example representatives include the zebrafish (*Danio rerio*), a model organism used in genomic and developmental biology, important aquaculture species like the common carp (*Cyprinus carpio*), major invasive species to North America such as *Hypophthalmichthys* (silver carp), and many popular aquarium species (rasboras and barbs).

For taxonomic clarity, this study follows the proposition by Mayden and Chen (2010) that

elevates subfamilies within Cyprinidae to the family level based on consistent support of major clades. Superfamilies are elevated to the suborder level to be consistent with the recognition of suborders as the taxonomic level above family and below order in the classification of bony fishes (Betancur-R et al., 2013; 2014). Other taxonomic assignments follow designations established by Kottelat (2013) Tang et al. (2013a), van der Laan et al. (2014), and L. Yang et al. (2015a). Because of the great diversity within Cypriniformes, most phylogenetic studies have focused on smaller groups within the order (for example Bufalino and Mayden, 2010a; Mayden et al., 2007; Schönhuth and Mayden, 2010; K. L. Tang et al., 2011). Approaches used to resolve relationships at these levels have typically included standard methods using PCR to amplify targeted mitochondrial and/or nuclear genes (Bufalino and Mayden, 2010a; Doosey et al., 2010; Mayden et al., 2007; Pramuk et al., 2007; Schönhuth and Mayden, 2010; Slechtová et al., 2008; K. L. Tang et al., 2011; 2010; Q. Tang et al., 2005). These approaches have had varied success at elucidating relationships at these taxonomic levels, but deeper, all-inclusive studies have resulted in conflicting phylogenies. These major differences in findings even include two publications in the same volume (Mayden and W. J. Chen, 2010; K. L. Tang et al., 2010) whose results are incongruent. Morphological studies have also been at odds with the molecular hypotheses, particularly concerning placement of the paedomorphic taxa (*Danionella*, *Paedocypris*, and *Sundadanio*) (Britz and Conway, 2011a; 2009; Britz et al., 2014; Mayden and W. J. Chen, 2010). The results of analyses to date mean that this radiation of organisms that is nearly the size of the Mammalia and that is the predominant freshwater order of fishes has an unsettled taxonomy and phylogeny despite the fact that it has been very highly studied. With the vertebrate developmental model (zebrafish) being part of the Cypriniformes, we are currently lacking a

basic understanding of the evolutionary context of its characteristics, and it is clear that new approaches to the phylogenetics of this very important group of fishes must be employed.

To date, the only nuclear genomic scale study (Tao et al., 2010) consisted of 100 genes and was limited to only thirteen individuals, most of which belong to Xenocyprididae within Cyprinoidei. The large number of taxa in Cypriniformes has forced researches to either focus on a small subset of representatives with an increasing number of molecular loci, or focus on large taxonomic representation with relatively fewer numbers of markers.

Evaluating tree topologies from previous large-scale studies has led to moderate consensus supporting monophyly for some clades within the order, including families of loaches (e.g. Botiidae, Cobitidae, Balitoridae, Nemacheilidae), Catostomidae (suckers), Cyprinidae, Xenocyprididae, Gobionidae, Leuciscidae, and Acheilognathidae (W. J. Chen and Mayden, 2009; Cunha et al., 2002; Fang et al., 2009; Gaubert et al., 2009; H. Liu and Y. Chen, 2003; Mayden and W. J. Chen, 2010; Mayden et al., 2008; Saitoh et al., 2006; K. L. Tang et al., 2010; Thai et al., 2007; X. Wang et al., 2012; 2007). Despite support for monophyly of many families, clear establishment of the relationships among them still remains elusive. Other families, most notably Danionidae, have been more problematic, with paedomorphic genera like *Paedocypris* and *Sundadanio* changing placement across trees employing both morphological and varying molecular data (Britz et al., 2014; Britz and Conway, 2009; Fang et al., 2009; Mayden and W. J. Chen, 2010; Rüber et al., 2007; K. L. Tang et al., 2010).

If analyses result in incongruent relationships due to conflict or weak phylogenetic signal among individual genes, the next approach to establishing robust resolution would be to incorporate high-throughput sequencing data that can increase the signal to noise ratio and

reduce stochastic error. New methods have been established that have been specifically tailored for use in systematics (Faircloth et al., 2012; A. R. Lemmon et al., 2012; A. R. Lemmon and E. M. Lemmon, 2012) and that address problems typical of transcriptome approaches for phylogenomics. These problems include tissue preservation, orthology assessment, missing data, and resolution capabilities across various taxonomic levels (Faircloth et al., 2012; A. R. Lemmon et al., 2012; E. M. Lemmon and A. R. Lemmon, 2013). All of these factors make anchored hybrid enrichment an attractive option for addressing the phylogenetic uncertainties still present within Cypriniformes. This study represents the largest dataset developed for Cypriniformes, both in taxonomic representation and genetic data, ameliorating many of the problems associated with resolving the relationships among and within families of this order. Not until these relationships are resolved can researchers begin to take advantage of the size, diversity, and distribution of Cypriniformes to gain insight into various biological facets, such as biogeography, timing of diversifications, morphological and ecological evolution, and comparative genomics.

METHODS

Taxon selection and tissue preparation

The 172 taxa selected for this study (Table 1.1) represent almost all families within the order. Families not represented in this study are: Psilorhynchidae (26 species), Barbuccidae (two species), Tincidae (13 species), Serpenticobitidae (three species), Ellopostomidae (two species) and Leptobarbidae (five species). Species were chosen based on tissue availability and because of their incorporation in recent studies that will allow for direct comparisons (Bufalino and Mayden, 2010a; W. J. Chen et al., 2009; W. J. Chen and Mayden, 2009; He et al., 2004; Mayden

et al., 2007; Saitoh et al., 2006; K. L. Tang et al., 2011). Type genera for each of the families were included if available. Exceptions include Botiidae, Balitoridae, Gastromyzontidae, and Xenocyprididae, but in these cases other representatives were chosen based on their supported inclusion within their respected families according to previous studies (Kottelat, 2013; K. L. Tang et al., 2013b). Three outgroup taxa were chosen to represent the three other ostariophysan orders: Siluriformes, Gymnotiformes, and Characiformes. Whole genomic DNA was prepared using the Omegabiotek E.Z.N.A. animal tissue extraction kit (product #D3396-02) and verified for quality and quantity using gel electrophoresis and nanodrop, respectively.

Locus selection and probe design

Although the Anchored Hybrid Enrichment kit developed for vertebrates by Lemmon et al. (2012) contains a fish reference (*Danio*) and has been utilized in teleosts with moderate success (Eytan et al., 2015), we desired an enrichment tool more efficient and appropriate for phylogenomics in teleosts. Because of the complex nature of teleost genome evolution, which involved multiple whole-genome duplications and lineage-specific gene losses (Glasauer and Neuhauss, 2014), it is impractical to identify a set of loci that are truly single-copy across all of Teleostei. Previous studies claiming to have identified single-copy loci in teleosts (e.g. Li et al., 2007) likely only identified loci that were single-copy in the species they considered; evaluation of those loci in additional teleost lineages suggests that these loci are not universally single-copy (see below). Consequently, we aimed to target loci containing up to four gene copies in each of three diverse lineages of teleosts: zebrafish, platyfish, and cichlids.

Candidate target regions for Teleostei were derived by combining the 394 Vertebrate

Anchor (v2) loci of Prum et al. (2015) and the 135 loci identified as Fugu-*Danio* single-copy orthologs by Li et al. (2007). For the vertebrate anchor loci, teleost orthologs were obtained for *Danio rerio* (danRer7) using the human (hg19) coordinates and the USCS genome browser batch-coordinate (liftover) tool (Kent et al., 2002). For the Fugu-*Danio* orthologs, orthologous human (hg19) and chicken (galGal3) coordinates were obtained using the USCS liftover tool and the *Danio* coordinates identified by Li et al. (2007). Once the coordinates for *Danio*, *Homo*, and *Gallus* were obtained for all 529 candidate target regions, sequences corresponding to those regions [plus sufficient flanking region to obtain up to 3000 base pairs (bp) total] were extracted from the genomes and aligned by locus using MAFFT (Katoh et al., 2002), v7.023b with "–genafpair" and "–maxiterate 1000" flags. The alignments were then used to generate a *Danio*-specific reference database containing spaced 20-mers. The *Danio* reference was then used to identify homologous regions in the genomes of zebrafish (Cypriniformes: Cyprinidae: *Danio rerio*; danRer7), platyfish (Cyprinodontiformes: Poeciliidae: *Xiphophorus maculatus* (Schartl et al., 2013), and cichlid (Perciformes, Cichlidae: *Maylandia zebra*; (Loh et al., 2008)).

As expected, we obtained multiple homologs for many of the candidate loci (only 64 loci were single-copy in all three species). Consequently, only 277 loci had fewer than five homologs per species and were considered further. We aligned with MAFFT (Katoh et al., 2002), v7.023b with "–genafpair" and "–maxiterate 1000" flags) all homolog sequences (up to 12 per locus) for each of the 277 candidates together with the homologous human probe region sequence from the Vertebrate Anchor (v2) design. Alignments were then manually inspected for misplaced and grossly misaligned sequences, which were removed. Finally, alignments were trimmed to include regions best suited for Anchored Hybrid Enrichment (conserved, low-gap, high taxon

representation), taking care that the chosen region contained the human probe region. A total of 260 loci were retained.

Finally, in order to ensure efficient enrichment, we checked for high-copy regions (e.g. microsatellites and transposable elements) in each of the three teleost references as follows. First, a database was constructed for each species using all 15-mers found in the trimmed alignments for that species. We also added to the database all 15-mers that were 1 bp removed from the observed 15-mers. The genome for the species was then exhaustively scanned for the presence of these 15-mers and matches were tallied at the alignment positions at which the 15-mer was found. Alignment regions containing > 100,000 counts in any of the three species were masked to prevent probe tiling across these regions. Probes of 120 bp were tiled uniformly at 5.5× tiling density.

Data collection

Multilocus sequence data were collected at the Center for Anchored Phylogenomics at Florida State University (www.anchoredphylogeny.com) following Lemmon et al. (Eytan et al., 2015) with some adjustments. Each genomic DNA sample was sonicated to a fragment size of ~175–300 bp using a Covaris E220 Focused-ultrasonicator with Covaris microTUBES. Library preparation and indexing followed Meyer and Kircher (2010). Indexed libraries were pooled at equal quantities (12 pools of 16 samples each), and the library pools were enriched using a custom Agilent Custom SureSelect kit (Agilent Technologies), with probes designed as described above. The 12 enriched library pools were pooled with equal quantities for sequencing on four PE150 Illumina HiSeq2000 lanes with eight bp indexing. Sequencing was performed at Florida

State University in the College of Medicine Translational Science Laboratory.

Data analysis

Reads were quality filtered using Illumina's Casava software with the chastity filter set to high. In order to increase read length and accuracy overlapping reads were then merged following Rokyta et al. (2012). Non-overlapping read pairs were kept separate but still used in the assembly. All reads were then assembled into contigs following Prum et al. (2015) using mapping references derived from the zebrafish, platyfish, and cichlid sequences used for probe design. This assembler produces separate contigs for gene copies differing by more than 5% sequence divergence. To reduce errors caused by low-level indexing errors during sequencing, contigs were then filtered by removing those derived from fewer than 50 reads.

Sets of homologs were produced by grouping by target locus (across individuals) and the filtered consensus sequences. Orthology was then determined for each target locus as follows: First, a pairwise distance measure was computed for pairs of homologs, with distance being computed as the percentage of 20-mers observed in the two sequences that were found in both sequences. A neighbor-joining clustering algorithm was then used to cluster the consensus sequences in to orthologous sets, with at most one sequence per species in each orthologous set [see Prum et al. (2015) for details]. In order to minimize the effects of missing data, clusters containing fewer than 130 (72%) of the species were removed from downstream processing.

Sequences in each orthologous set were aligned using MAFFT v7.023b (Katoh et al., 2002) with

"–genafpair" and "–maxiterate 1000" flags. In order to remove poorly aligned regions raw

alignments were then trimmed and masked following Prum et al. (2015), with the following adjustments: sites with > 50% similarity were identified as good, 20 bp regions containing < 14 good sites were masked, and sites with fewer than 30 unmasked bases were removed from the alignment.

For all phylogenetic analyses, sequences from the gymnotiform, siluriform, and characiform species were used as the outgroup. For the concatenated dataset, the alignment was partitioned by locus and the phylogeny estimated using RAxML using GTR + Γ model with 500 bootstrap replicates. For the species tree analysis, a maximum likelihood phylogeny was estimated with 100 bootstrap replicates for each of the separate loci using RAxML with GTR + Γ model assumed. We then used the RAxML bootstrap trees to estimate a species tree using STAR (L. Liu et al., 2009) with default parameters using STRAW (Shaw et al., 2013). ASTRAL-II (v4.10.2) (Mirarab and Warnow, 2015) was also used for species tree inference using the gene trees and their 100 bootstrap replicates. We performed 100 replicates of multi-locus bootstrapping.

To test our analyses against previous morphological hypotheses, we re-examined the datasets in Conway (Conway, 2011) and Britz et al. (Britz et al., 2014) by running 1000 replicates of a heuristic search in PAUP* (Swofford, 2002). We traced the characters in Mesquite v.3.04 (W. P. Maddison and D. R. Maddison, 2015). We also performed Bayesian analyses on these morphological datasets under the Mk + Γ model in mrBayes 3.2 (Ronquist et al., 2012), which has been demonstrated to perform better than parsimony due to rate heterogeneity in character evolution (Wright and Hillis, 2014). Estimating rate heterogeneity can be biased by sampling only variable or parsimony-informative characters, so we analyzed the data with

correction for parsimony-informative characters for the Conway (2011) dataset and variable characters for the Britz et al. (2014) datasets (one character in these datasets was not parsimony-informative). For each dataset, we ran MCMC with two runs of four chains for 1,000,000 generations, sampling every 1,000. We assessed convergence using Tracer v1.5 (Rambaut and Drummond, 2009).

RESULTS

A total of 315,288 base pairs (bp) spanning 219 loci were obtained for use in estimating the phylogenetic relationships. Average locus length was 1011 bp with a range of 134–2119 bp (Figure 1.1) The total number of informative characters was 295,252 bp with only 3.48% missing data (Dryad accession link: doi:10.5061/dryad.b3d03; raw reads available on NCBI SRA (Bioproject PRJNA345212). Our results show promise for the ability of this method to provide robust support for relationships, with 97% of nodes resolved at 100% bootstrap support. Findings include resolution of major clades supported by previous work (e.g. families within Cyprinoidei — see Figure 1.2), but relationships among these clades differ. Major results include paraphyly of Cobitoidei, with Gyrinocheilidae sister to the rest of Cypriniformes, followed by Catostomidae sister to the remaining ingroup (see below). We find support for Mayden and W. J. Chen's (2010) recognition of Paedocyprididae and Sundadanionidae since neither is recovered within Danionidae. Leuciscidae are sister to Tanichthyidae, Acheilognathidae are sister to Gobionidae, and these two clades are sister to each other [(Acheilognathidae + Gobionidae) + (Tanichthyidae + Leuciscidae)]. Xenocyprididae falls sister to these four families.

Concatenated tree vs. species tree

We find only a few major differences between our maximum likelihood concatenated tree (CT; Figure 1.3) and the species trees (ST; Figures 1.4 and 1.5). These include support for monophyly of Cobitoidei in the ST but not in the CT, and a different placement for the Danionidae between the two trees. Other minor differences are found among a few shallow sister relationships that had lower support values in both trees. Other studies have shown that concatenation methods may perform better over coalescent species tree methods, especially at deeper nodes, and our discussion of clades will focus on the CT tree (Gatesy and Springer, 2014; Prum et al., 2015; Tonini et al., 2015).

Reanalysis of Cobitoidei morphological datasets

The most robust morphological phylogenies putatively supporting a monophyletic Cobitoidei is that of Conway (2011); however, when we reanalyzed the characters using parsimony in PAUP* (Swofford, 2002), we achieved different results. We ran the analysis according to Conway (2011) with the exception that we ran 1000 replicates of a heuristic search; it appears Conway (2011) only ran a single replicate of a heuristic search, and that search settled on a tree island of 14 most parsimonious trees. We found one additional tree island with an additional 56 trees, which was found nearly as often as the 14-tree island (515 times vs. 485). The strict consensus of the 70 trees showed a polytomy at the base of the Cypriniformes with the gyrinocheilids, catostomids, loaches, and cyprinoids. The analyses in Britz et al. (2014) did use 10 replicates of the heuristic search and are more accurate (we found more trees for their Morphological Dataset 3), and always found a monophyletic Cobitoidei, but this was weakly

supported. Conway (2011) lists seven characters supporting Cobitoidei, but our analysis showed that two of these (characters 32:1 and 99:1) were not listed as changed along the branch leading to the Cobitoidea and only one (character 19:1) is actually present in all families of cobitoids. All the remaining derived character states are absent in one of the three lineages (gyrinocheilids, catostomids, or loaches) meaning morphological support for a monophyletic group containing these three clades is poor. Support was stronger for a sister group relationship between gyrinocheilids plus catostomids [seven characters in Conway (2011), six in our analysis]; however, we found seven characters supporting loaches plus cyprinoids (characters 7:0, 18:0, 46:1, 76:0, 83:2, 100:0, and 111:2) and seven characters supporting catostomids plus loaches plus cyprinoids (characters 11:0, 31:1, 36:1, 53:1, 68:1, 69:1, and 77:1) indicating roughly equal morphological support for the two hypotheses. Considerable homoplasy is found in most of the characters under all arrangements; however, characters 53, 83, and 77 provide unambiguous support for the relationships presented in this study.

In addition, the Bayesian analysis of the morphological characters resulted in only poor support [<.95 posterior probability, following Alfaro & Holder (2006)] for monophyly of the Cobitoidei. In the analysis of the Conway (2011) dataset, the catostmoids, gyrinocheilids, loaches, and cyprinoids form an unresolved polytomy in the consensus tree; this differs from the support present in Conway (2011) for this node (.5–.9 pp). In the analyses of the Britz et al. (2014) datasets, support ranged from .57 to.63 posterior probability across datasets, indicating low levels of support.

DISCUSSION

We have presented the first order-wide, phylogenomic analysis of the Cypriniformes, and we demonstrate the utility of anchored enrichment at assessing the relationships of fishes from deep to more recent divergences. Our analyses demonstrate conflict in the relationships of the Cobitoidei, the placement of *Paedocypris* as sister to all other cyprinoids, and a validation of the previously well-supported monophyly of many major cypriniform families. Although the wide variety of different hypotheses for the cypriniforms has been called the "Cypriniformes tree of confusion" (Britz and Conway, 2011a; 2011b), the anchored enrichment phylogenomic tree that we present provides the most robust phylogenetic analysis to date, supporting many of the previous hypotheses of relationships and providing new ideas that will require further scrutiny.

Non-monophyly of Cobitoidei

The most surprising result of the study is the nonmonophyly of Cobitoidei in the concatenation analysis (Figure 1.6). Cobitoids are largely believed to be monophyletic, however, many different placements of the taxa have been found. The Gyrinocheilidae (three species), Catostomidae (83 species), and loaches (Botiidae, 56 species; Balitoridae, 229 species; Cobitidae, ~198 species; Nemacheilidae, 658 species; Vaillantellidae, three species; and Gastromyzontidae, 137 species) represent successive sister groups to the Cyprinoidei in our concatenated analyses. Species tree analysis did find a monophyletic Cobitoidei; however, recent research has found that species tree analyses may not be as accurate at deeper levels of the phylogeny (Gatesy and Springer, 2014; Prum et al., 2015; Tonini et al., 2015). Considering these studies, the depth of the nodes leading to members of Cobitoidei, and the results of the reanalysis of morphological data that had previously supported monophyly of the group, we are compelled

to follow the relationships presented in the concatenation analysis until further exploration regarding the discrepancies between concatenation versus species trees is conducted and consensus by the scientific community is reached.

Phylogenetic reanalysis of available morphological characters does not provide strong evidence for a monophyletic Cobitoidei, and morphological characters provide at least equally strong support for the relationships presented here. We restrict Cobitoidei to the loaches, and erect new suborders for the Gyrinocheilidae (Gyrinocheiloidei) and the Catostomidae (Catostomoidei).

Cyprinidae

Among the Labeoninae (Figure 1.7), we find support for many of the tribes (discussed as subtribes in L. Yang et al. (2012). These tribes, based on analysis of four nuclear and five mitochondrial genes, are: Labeonini, Garrini, “Osteochilini”, and “Semilabeonini” (quotation marks denote a lack of formal description). Labeonini was resolved as monophyletic as in L. Yang et al. (2012). We also obtained *Gibelion* nested within *Labeo*, and non-monophyly of *Cirrhinus*. Although Kottelat (2013) recognized *Gymnostomus* as the valid generic name for *Henicorhynchus siamensis*, we find a pattern similar to L. Yang et al. (2012) where this species is within the “Osteochilini” species group instead of with other members of *Gymnostomus* in Labeonini. *Placocheilus cryptonemus* was resolved as belonging to “Semilabeonini” in L. Yang et al. (2012) but *Placocheilus dulongensis* in our Anchored Enrichment (AE) tree is resolved within Garrini. Lothongkham et al. (2014) established *Placocheilus* as a synonym of *Garra*, but members of this group need further study to determine which species should be synonymized

with *Garra* (e.g. *P. dulongensis*). Because of the particular placement of *Placocheilus dulongensis* within Garrini (compared to other members of *Placocheilus* in "Semilabeonini"), our analyses did not include a representative of the "Semilabeonini" species group, but the relationships among the tribes of Labeoninae presented in this study are consistent with L. Yang et al. (2012).

For the remaining members of Cyprinidae, we find resolution for clades similar to those by L. Yang et al. (2015a) although none of the AE relationships among these clades are consistent with their results. For example, we resolve Labeoninae as sister to remaining members of Cyprinidae as opposed to Probarbinae as presented in L. Yang et al. (2015a). Of particular interest is *Chagunius chagunio*, which L. Yang et al. (2015a) placed in the Smiliogastrinae. We obtain it as sister to a clade comprised of Spinibarbinae, Acrossocheilinae, Schizopygopsinae, Schizothoracinae, Torinae and Barbinae, with other Smiliogastrinae species more closely related to "Poropuntiinae" than to *Chagunius*. Lei Yang et al. (2015a) had 0.80 posterior probability support for their placement based on mitogenome data, but less than 0.50 in their nuclear analysis (RAG1). Lei Yang et al. (2015) found numerous inter-clade hybridization events leading to allopolyploidy, which greatly complicates phylogenetic analysis within the Cyprinidae. We leave *Chagunius* as *incertae sedis* within Cyprinidae.

Xenocyprididae, Acheilognathidae, Gobionidae, Tanichthyidae, and Leuciscidae

Placement of these families has varied across different studies (W. J. Chen et al., 2013; Mayden and W. J. Chen, 2010; Saitoh et al., 2011; K. L. Tang et al., 2013a; Tao et al., 2013) and here we obtain sister relationships between Acheilognathidae + Gobionidae and Tanichthyidae +

Leuciscidae, with Xenocyprididae sister to all four of these families (Figure 1.8). Within Xenocyprididae, relationships are similar to those found by Tao et al. (2010) for the five taxa common to both studies. This differs from relationships reported by He et al. (2004) and Wang et al. (2007), but the congruencies to Tao et al. (2010) are not surprising given that their data were also acquired on a phylogenomic scale (100 genes, 13 taxa). Kevin L. Tang et al. (2013a) used two nuclear and two mitochondrial markers to elucidate the relationships among Xenocyprididae (van der Laan et al., 2014; referred to as Oxygastrinae in their paper) and our results only differ for those relationships they obtained that were poorly supported. These include a different placement of the *Metzia* + *Hemmigrammocypris* clade and differing relationships among genera within a clade that includes *Hypophthalmichthys*, *Parabramis*, *Chanodichthys*, *Squaliobarbus*, *Ctenopharyngadon*, and *Elopichthys*. For Gobionidae, results in this study are highly congruent with previous molecular studies (Saitoh et al., 2011; K. L. Tang et al., 2011; J. Yang et al., 2006) that resolve the following clades and their relationships to each other: *Pseudogobio* group, *Gobio* group, *Sarcocheilichthys* group, and *Hemibarbus* group [see J. Yang et al. (2006) for group designations]. Leuciscidae has long been supported as monophyletic across many studies (Briolay et al., 1998; W. J. Chen and Mayden, 2009; Cunha et al., 2002; Gaubert et al., 2009; Mayden et al., 2009; 2008; Mayden and W. J. Chen, 2010; Saitoh et al., 2006; K. L. Tang et al., 2010; Thai et al., 2007; C. Wang et al., 2012; X. Wang et al., 2007) but relationships among the genera within have had differing results. Clades have been resolved in multiple studies and include: (1) far eastern phoxinins (Eurasian), (2) open posterior myodome (OPM), (3) creek chub – plagopterin (CC-P), (4) western North America (WNA), and (5) leuciscin (European) (Bufalino and Mayden, 2010a; 2010b; 2010c; Cavender and Coburn, 1992;

Cunha et al., 2002; Imoto et al., 2013; Rüber et al., 2007; Saitoh et al., 2011; Sakai et al., 2006; Sasaki et al., 2007; Strange and Mayden, 2009; Zhang et al., 2008). Our results also obtained the five major clades within Leuciscidae (Figure 1.8), but yield strongly supported novel relationships that change our understanding of the biogeographical patterns exhibited by this family. Similar to the previous studies, we find *Notemigonus* (North American) within the leuciscin (European) clade, but in sharp contrast to these studies, all other North American Leuciscidae are monophyletic. This study provides a framework to further investigate the timing and number of invasions of leuciscids to North America. The hypothesized rapid diversification of North American leuciscids has led to difficulty in resolving relationships within this clade, but our robust phylogeny exemplifies the potential for anchored enrichment and next-generation sequencing in elucidating the relationships within problematic clades. The biogeographical patterns of Leuciscidae is further discussed in chapter two.

The Cypriniformes is among the most important clades of freshwater fishes and among the most studied with phylogenetic inference. This great deal of work makes them a key group in understanding the various pit-falls of phylogenetic studies, and they exemplify the phylogenetic conflicts from the varying analyses of morphological, mitochondrial, and nuclear data. While many major clades of Cypriniformes have been long supported, relationships within and among them have proven difficult to resolve across the entire order. Varying markers and morphological data have given different results and have been difficult to apply across such a large and diverse group. With the development of phylogenomic techniques, researchers can now acquire a substantial amount of highly informative, quality data for resolving dynamic relationships, and we demonstrate the efficacy of the approach using the very complex cypriniforms. Robust

phylogenies are not only a prerequisite for a stable taxonomy, but are needed to address important evolutionary questions such as the timing of diversification, the geographic origins of clades, and the evolution of morphological and ecological novelty. For example, according to our results, Cypriniformes appear to have invaded North America at least twice and Africa several times from Eurasia, with these transcontinental migrations resulting in very diverse clades. With the robust phylogeny we present here, we provide a framework for studying the consequences of these transcontinental migrations and how clades can diversify from within established ecosystems. Such studies will have broad consequences in studies on the evolution of diversity. The great diversity of Cypriniformes and the inclusion of perhaps the most important vertebrate model organism (Zebra Danio) make Cypriniformes an ideal group for comparative analyses. Considerable insight into the functioning of genes within vertebrate organisms has been obtained from the analysis of the Zebra Danio including forced mutations that often result in unviable larvae. By comparing the genome of the Zebra Danio with close relatives, the role of mutations and gene expression can be determined. Comparative genomic studies within Cypriniformes have already benefited from the foundation and annotation of the Zebra Danio genome sequence to generate insights into the functional evolution of various adaptations including adaptation to harsh environments such as caves and high altitude streams (Meng et al., 2013; L. Yang et al., 2015b). With a robust phylogeny, we can get a much better understanding of the function of genes by treating relatives of the Zebra Danio as natural mutants screened by natural selection (Mayden and W. J. Chen, 2010). As the Cypriniformes continues to become a more genome-enabled clade, with several new genomes published in the last few years (Burns et al., 2015; Xu et al., 2014; J. Yang et al., 2016; L. Yang et al., 2015b), we expect our phylogeny to provide a

useful framework for comparative genomics.

References

Alfaro, M.E., Holder, M.T., 2006. The Posterior and the Prior in Bayesian Phylogenetics. Annu. Rev. Ecol. Evol. Syst. 37, 19–42. doi:10.1146/annurev.ecolsys.37.091305.110021

Betancur-R, R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton, J.C., II, Zhang, F., Buser, T., Campbell, M.A., Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C., Hough, D.J., Lu, G., Grande, T., Arratia, G., Orti, G., 2013. The tree of life and a new classification of bony fishes. PLoS Curr 5. doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288

Betancur-R, R., Wiley, E.O., Bailly, N., Miya, M., Lecointre, G., Orti, G., 2014. Phylogenetic Classification of Bony Fishes — Version 3 [WWW Document]. sites.google.com. URL https://sites.google.com/site/guilleorti/home/classification (accessed 3.17.16).

Briolay, J., Galtier, N., Brito, R.M., Bouvet, Y., 1998. Molecular Phylogeny of Cyprinidae Inferred from cytochrome b DNA Sequences. Molecular Phylogenetics and Evolution 9, 100–108.

Britz, R., Conway, K.W., 2011a. The Cypriniformes Tree of Confusion. Zootaxa 2946, 73–78.

Britz, R., Conway, K.W., 2011b. Additions to “The Cypriniformes Tree of Confusion.” Zootaxa 2946, 142–142.

Britz, R., Conway, K.W., 2009. Osteology of *Paedocypris*, a miniature and highly developmentally truncated fish (Teleostei: Ostariophysi: Cyprinidae). J. Morphol. 270, 389–412. doi:10.1002/jmor.10698

Britz, R., Conway, K.W., Rüber, L., 2014. Miniatures, morphology and molecules: *Paedocypris* and its phylogenetic position (Teleostei, Cypriniformes). Zoological Journal of the Linnean Society 172, 556–615. doi:10.1111/zoj.12184

Bufalino, A.P., Mayden, R.L., 2010a. Molecular phylogenetics of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) based on RAG1 and S7 nuclear DNA sequence data. Molecular Phylogenetics and Evolution 55, 274–283. doi:10.1016/j.ympev.2009.12.017

Bufalino, A.P., Mayden, R.L., 2010b. Phylogenetic relationships of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) as inferred from S7 nuclear DNA sequences. Molecular Phylogenetics and Evolution 55, 143–152. doi:10.1016/j.ympev.2009.11.001

Bufalino, A.P., Mayden, R.L., 2010c. Phylogenetic evaluation of North American Leuciscidae (Actinopterygii: Cypriniformes: Cyprinoidea) as inferred from analyses of mitochondrial

and nuclear DNA sequences. Systematics and Biodiversity 8, 493–505. doi: 10.1080/14772000.2010.516029

Burns, F.R., Cogburn, A.L., Ankley, G.T., Villeneuve, D.L., Waits, E., Chang, Y.-J., Llaca, V., Deschamps, S.D., Jackson, R.E., Hoke, R.A., 2015. Sequencing and de novo draft assemblies of a fathead minnow (Pimephales promelas) reference genome. Environ Toxicol Chem 35, 212–217. doi:10.1002/etc.3186

Cavender, T.M., Coburn M.M., 1992. In: Mayden, R.L. ed. Systematics, historical ecology, and North American freswhater fishes. Palo Alto, CA: Stanford University Press, 293-327.

Chen, W.J., Lavoué, S., Mayden, R.L., 2013. Evolutionary Origin and Early Biogeography of Otophysan Fishes (Ostariophysi: Teleostei). Evolution 67, 2218–2239. doi:10.1111/evo. 12104

Chen, W.J., Lheknim, V., Mayden, R.L., 2009. Molecular phylogeny of the Cobitoidea (Teleostei: Cypriniformes) revisited: position of enigmatic loach *Ellopostoma* resolved with six nuclear genes. Journal of Fish Biology 75, 2197–2208. doi:10.1111/j. 1095-8649.2009.02398.x

Chen, W.J., Mayden, R.L., 2009. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: Further evidence from six nuclear genes. Molecular Phylogenetics and Evolution 52, 544–549. doi:10.1016/ j.ympev.2009.01.006

Conway, K.W., 2011. Osteology of the South Asian Genus *Psilorhynchus* McClelland, 1839 (Teleostei: Ostariophysi: Psilorhynchidae), with investigation of its phylogenetic relationships within the order Cypriniformes. Zoological Journal of the Linnean Society 163, 50–154. doi:10.1111/j.1096-3642.2011.00698.x

Cunha, C., Mesquita, N., Dowling, T.E., Gilles, A., Coelho, M.M., 2002. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences 61, 1–16. doi:10.1006/jfbi.2002.2105

Doosey, M.H., Bart, H.L., Jr, Saitoh, K., Miya, M., 2010. Phylogenetic relationships of catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. Molecular Phylogenetics and Evolution 54, 1028–1034. doi:10.1016/ j.ympev.2009.06.006

Eschmeyer, W.N., Fong, J.D., 2016. Catalog of Fishes [WWW Document]. researcharchive.calacademy.org. URL http://researcharchive.calacademy.org/research/ ichthyology/catalog/speciesbyfamily.asp (accessed 3.17.16).

Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Lemmon, E.M., Wainwright, P.C., Near, T.J., 2015. Are 100 enough? Inferring acanthomorphteleost phylogeny using Anchored HybridEnrichment. BMC Evolutionary Biology 15, 113. doi:10.1186/s12862-015-0415-0

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Systematic Biology. doi:10.1093/sysbio/sys004

Fang, F., Norén, M., Liao, T.Y., Källersjö, M., Kullander, S.O., 2009. Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). Zoologica Scripta 38, 237–256. doi:10.1111/j.1463-6409.2008.00373.x

Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution 80, 231–266. doi:10.1016/j.ympev.2014.08.013

Gaubert, P., Denys, G., Oberdorff, T., 2009. Genus-level supertree of Cyprinidae (Actinopterygii: Cypriniformes), partitioned qualitative clade support and test of macro-evolutionary scenarios. Biological Reviews 84, 653–689. doi:10.1111/j.1469-185X.2009.00091.x

Glasauer, S.M.K., Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics 289, 1045–1060. doi:10.1007/s00438-014-0889-2

He, S., Liu, H., Yiyu, C., Kuwahara, M., Nakajima, T., Zhong, Y., 2004. Molecular phylogenetic relationships of Eastern Asian Cyprinidae (Pisces: Cypriniformes) inferred from cytochrome b sequences. Science in China Series C 47, 130–138. doi:10.1360/03yc0034

Imoto, J.M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y.P., Miya, M., Nishida, M., Hanzawa, N., 2013. Phylogeny and biogeography of highly diverged freshwater fish species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. Gene 514, 112–124. doi:10.1016/j.gene.2012.10.019

Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30, 3059–3066.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Research 12, 996–1006. doi:10.1101/gr.229102

Kottelat, M., 2013. The fishes of the inland waters of Southeast Asia: a catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. The Raffles Bulletin of Zoology 27, 1–663.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Systematic Biology 61, 727–744. doi:10.1093/sysbio/sys049

Lemmon, A.R., Lemmon, E.M., 2012. High-throughput identification of informative nuclear loci for shallow-scale phylogenetics and phylogeography. Systematic Biology 61, 745–761. doi:10.1093/sysbio/sys051

Lemmon, E.M., Lemmon, A.R., 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. ….

Li, C., Orti, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evolutionary Biology 7, 44. doi:10.1186/1471-2148-7-44

Liu, H., Chen, Y., 2003. Phylogeny of the East Asian cyprinids inferred from sequences of the mitochondrial DNA control region. Can. J. Zool. 81, 1938–1946. doi:10.1139/z03-169

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating Species Phylogenies Using Coalescence Times among Sequences. Systematic Biology 58, 468–477. doi:10.1093/sysbio/syp031

Loh, Y.-H.E., Katz, L.S., Mims, M.C., Kocher, T.D., Yi, S.V., Streelman, J.T., 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. Genome Biol 9, R113. doi:10.1186/gb-2008-9-7-r113

Lothongkham, A., Arbsuwan, S., Musikasinthorn, P., 2014. *Garra waensis*, a new cyprinid fish (Actinopterygii: Cypriniformes) from the Nan River basin of the Chao Phraya River system, northern Thailand. Zootaxa 3790, 543–554. doi:10.11646/zootaxa.3790.4.3

Maddison, W.P., Maddison, D.R., 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04 http://mesquiteproject.org.

Mayden, R.L., Chen, W.J., 2010. The world“s smallest vertebrate species of the genus *Paedocypris*: a new family of freshwater fishes and the sister group to the world”s most diverse clade of freshwater fishes (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 57, 152–175. doi:10.1016/j.ympev.2010.04.008

Mayden, R.L., Chen, W.J., Bart, H.L., Doosey, M.H., Simons, A.M., Tang, K.L., Wood, R.M., Agnew, M.K., Yang, L., Hirt, M.V., Clements, M.D., Saitoh, K., Sado, T., Miya, M., Nishida, M., 2009. Reconstructing the phylogenetic relationships of the earth’s most diverse clade of freshwater fishes—order Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and the mitochondrial genome. Molecular Phylogenetics and Evolution 51, 500–514. doi:10.1016/j.ympev.2008.12.015

Mayden, R.L., Tang, K.L., Conway, K.W., 2007. Phylogenetic relationships of *Danio* within the order Cypriniformes: a framework for comparative and evolutionary studies of a model species. Journal of Experimental Biology 308B, 642–654.

Mayden, R.L., Tang, K.L., Wood, R.M., Chen, W.J., Agnew, M.K., Conway, K.W., Yang, L., Simons, A.M., Bart, H.L., Harris, P.M., 2008. Inferring the Tree of Life of the order Cypriniformes, the earth's most diverse clade of freshwater fishes: Implications of varied taxon and character sampling. Journal of Systematics and Evolution 424–438.

Meng, F., Braasch, I., Phillips, J.B., Lin, X., Titus, T., Zhang, C., Postlethwait, J.H., 2013. Evolution of the Eye Transcriptome under Constant Darkness in Sinocyclocheilus Cavefish. Molecular Biology and Evolution 30, 1527–1543. doi:10.1093/molbev/mst079

Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. Cold Spring Harbor Protocols 2010, pdb.prot5448–pdb.prot5448. doi:10.1101/pdb.prot5448

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–52. doi:10.1093/bioinformatics/btv234

Pramuk, J.B., Grose, M.J., Clarke, A.L., Greenbaum, E., Bonaccorso, E., Guayasamin, J.M., Smith-Pardo, A.H., Benz, B.W., Harris, B.R., Siegfreid, E., Reid, Y.R., Holcroft-Benson, N., Wiley, E.O., 2007. Phylogeny of finescale shiners of the genus *Lythrurus* (Cypriniformes: Cyprinidae) inferred from four mitochondrial genes. Molecular Phylogenetics and Evolution 42, 287–297. doi:10.1016/j.ympev.2006.06.008

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569–573. doi:10.1038/nature15697

Rambaut, A., Drummond, A.J., 2009. Tracer v.1.5 [WWW Document]. URL http://tree.bio.ed.ac.uk/software/tracer/ (accessed 3.18.16).

Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13, 312. doi:10.1186/1471-2164-13-312

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61, 539–542. doi:10.1093/sysbio/sys029

Rüber, L., Kottelat, M., Tan, H., Ng, P.K., Britz, R., 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. BMC Evolutionary Biology 7, 38. doi:10.1186/1471-2148-7-38

Saitoh, K., Sado, T., Doosey, M.H., Bart, H.L., Jr, Inoue, J.G., Nishida, M., Mayden, R.L., Miya, M., 2011. Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). Zoological Journal of the Linnean Society 161, 633–662. doi:10.1111/j.1096-3642.2010.00651.x

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M., 2006. Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. Journal of Molecular Evolution 63, 826–841. doi:10.1007/s00239-005-0293-y

Sakai, H., Ito, Y., Shedko, S.V., Safronov, S.N., Frolov, S.V., Chereshnev, I.A., Jeon, S.-R., Goto, A., 2006. Phylogenetic and taxonomic relationships of northern Far Eastern phoxinin minnows, *Phoxinus* and *Rhynchocypris* (Pisces, Cyprinidae), as inferred from allozyme and mitochondrial 16S rRNA sequence analyses. Zool. Sci. 23, 323–331.

Sasaki, T., Kartavtsev, Y.P., Chiba, S.N., Uematsu, T., Sviridov, V.V., Hanzawa, N., 2007. Genetic divergence and phylogenetic independence of Far Eastern species in subfamily Leuciscinae (Pisces: Cyprinidae) inferred from mitochondrial DNA analyses. Genes Genet. Syst. 82, 329–340.

Schartl, M., Walter, R.B., Shen, Y., Garcia, T., Catchen, J., Amores, A., Braasch, I., Chalopin, D., Volff, J.-N., Lesch, K.-P., Bisazza, A., Minx, P., Hillier, L., Wilson, R.K., Fuerstenberg, S., Boore, J., Searle, S., Postlethwait, J.H., Warren, W.C., 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nature Publishing Group 45, 567–572. doi:10.1038/ng.2604

Schönhuth, S., Mayden, R.L., 2010. Phylogenetic relationships in the genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences. Molecular Phylogenetics and Evolution 55, 77–98. doi:10.1016/j.ympev.2009.10.030

Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: Species TRee Analysis Web server. Nucleic Acids Research 41, W238–W241. doi:10.1093/nar/gkt377

Slechtová, V., Bohlen, J., Perdices, A., 2008. Molecular phylogeny of the freshwater fish family Cobitidae (Cypriniformes: Teleostei): Delimitation of genera, mitochondrial introgression and evolution of sexual dimorphism. Molecular Phylogenetics and Evolution 47, 812–831. doi:10.1016/j.ympev.2007.12.018

Strange, R.M., Mayden, R.L., 2009. Phylogenetic Relationships and a Revised Taxonomy for North American Cyprinids Currently Assigned to *Phoxinus* (Actinopterygii: Cyprinidae). Copeia 2009, 494–501. doi:10.1643/CI-07-070

Swofford, D.L., 2002. Phylogenetic analysis using parsimony (PAUP), version 4.0 b10. Illinois Natural History Survey.

Tang, K.L., Agnew, M.K., Chen, W.J., Hirt, M.V., Raley, M.E., Sado, T., Schneider, L.M., Yang, L., Bart, H.L., He, S., Liu, H., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2011. Phylogeny of the gudgeons (Teleostei: Cyprinidae: Gobioninae). Molecular Phylogenetics and Evolution 61, 103–124. doi:10.1016/j.ympev.2011.05.022

Tang, K.L., Agnew, M.K., Hirt, M.V., Lumbantobing, D.N., Raley, M.E., Sado, T., TEOH, V.-H., Yang, L., Bart, H.L., Harris, P.M., He, S., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2013a. Limits and phylogenetic relationships of East Asian fishes in the subfamily Oxygastrinae (Teleostei: Cypriniformes: Cyprinidae). Zootaxa 3681, 101. doi:10.11646/zootaxa.3681.2.1

Tang, K.L., Agnew, M.K., Hirt, M.V., Sado, T., Schneider, L.M., Freyhof, J., Sulaiman, Z., Swartz, E., Vidthayanon, C., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2010. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). Molecular Phylogenetics and Evolution 57, 189–214. doi:10.1016/j.ympev.2010.05.021

Tang, K.L., Lumbantobing, D.N., Mayden, R.L., 2013b. The Phylogenetic Placement of Oxygastervan Hasselt, 1823 (Teleostei: Cypriniformes: Cyprinidae) and the Taxonomic Status of the Family-Group Name Oxygastrinae Bleeker, 1860. Copeia 2013, 13–22. doi:10.1643/CG-10-121

Tang, Q., Xiong, B., Yang, X., Liu, H., 2005. Phylogeny of the East Asian botiine loaches (Cypriniformes, Botiidae) inferred from mitochondrial cytochrome b gene sequences. Hydrobiologia 544, 249–258. doi:10.1007/s10750-005-0863-4

Tao, W., Mayden, R.L., He, S., 2013. Remarkable phylogenetic resolution of the most complex clade of Cyprinidae (Teleostei: Cypriniformes): A proof of concept of homology assessment and partitioning sequence data integrated with mixed model Bayesian analyses. Molecular Phylogenetics and Evolution 66, 603–616. doi:10.1016/j.ympev.2012.09.024

Tao, W., Zou, M., Wang, X., Gan, X., Mayden, R.L., He, S., 2010. Phylogenomic Analysis Resolves the Formerly Intractable Adaptive Diversification of the Endemic Clade of East Asian Cyprinidae (Cypriniformes). PLoS ONE 5, e13508. doi:10.1371/journal.pone.0013508.t004

Thai, B.T., Si, V.N., Phan, P.D., Austin, C.M., 2007. Phylogenetic evaluation of subfamily classification of the Cyprinidae focusing on Vietnamese species. Aquat. Living Resour. 20, 143–153. doi:10.1051/alr:2007025

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Orti, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr. doi:10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be

van der Laan, R., Eschmeyer, W.N., Fricke, R., 2014. Family-group names of Recent fishes. Zootaxa 3882, 1. doi:10.11646/zootaxa.3882.1.1

Wang, C., Wang, J., Yang, J., Lu, G., Song, X., Chen, Q., Xu, J., Yang, Q., Li, S., 2012. Complete mitogenome sequence of black carp (*Mylopharyngodon piceus*) and its use for molecular phylogeny of leuciscine fishes. Mol Biol Rep 39, 6337–6342. doi:10.1007/s11033-012-1455-9

Wang, X., Gan, X., Li, J., Mayden, R.L., He, S., 2012. Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. Sci. China Life Sci. 55, 761–773. doi:10.1007/s11427-012-4366-z

Wang, X., Li, J., He, S., 2007. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. Molecular Phylogenetics and Evolution 42, 157–170. doi:10.1016/j.ympev.2006.06.014

Wright, A.M., Hillis, D.M., 2014. Bayesian Analysis Using a Simple Likelihood Model Outperforms Parsimony for Estimation of Phylogeny from Discrete Morphological Data. PLoS ONE 9, e109210. doi:10.1371/journal.pone.0109210.s002

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Cao, D., Lu, C., Li, C., Zhou, Y., Liu, Z., Fan, Z., Shan, G., Li, X., Wu, S., Song, L., Hou, G., Jiang, Y., Jeney, Z., Yu, D., Wang, L., Shao, C., Song, L., Sun, J., Ji, P., Wang, J., Li, Q., Xu, L., Sun, F., Feng, J., Wang, C., Wang, S., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang, Z., Yu, J., Sun, X., 2014. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nature Publishing Group 46, 1212–1219. doi:10.1038/ng.3098

Yang, J., Chen, X., Bai, J., Fang, D., Qiu, Y., Jiang, W., Yuan, H., Bian, C., Lu, J., He, S., Pan, X., Zhang, Y., Wang, X., You, X., Wang, Y., Sun, Y., Mao, D., Liu, Y., Fan, G., Zhang, H., Chen, X., Zhang, X., Zheng, L., Wang, J., Le Cheng, Chen, J., Ruan, Z., Li, J., Yu, H., Peng, C., Ma, X., Xu, J., He, Y., Xu, Z., Xu, P., Wang, J., Yang, H., Wang, J., Whitten, T., Xu, X., Shi, Q., 2016. The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. BMC Biology 14, 1–13. doi:10.1186/s12915-015-0223-4

Yang, J., He, S., Freyhof, J., Witte, K., Liu, H., 2006. The phylogenetic relationships of the Gobioninae (Teleostei: Cyprinidae) inferred from mitochondrial cytochrome b gene sequences. Hydrobiologia 553, 255–266. doi:10.1007/s10750-005-1301-3

Yang, L., Arunachalam, M., Sado, T., Levin, B.A., Golubtsov, A.S., Freyhof, J., Friel, J.P., Chen, W.J., Hirt, M.V., Manickam, R., Agnew, M.K., Simons, A.M., Saitoh, K., Miya, M., Mayden, R.L., He, S., 2012. Molecular phylogeny of the cyprinid tribe Labeonini (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 65, 362–379. doi: 10.1016/j.ympev.2012.06.007

Yang, L., Sado, T., Hirt, M.V., Pasco-Viel, E., Arunachalam, M., Li, J., Wang, X., Freyhof, J., Saitoh, K., Simons, A.M., Miya, M., He, S., Mayden, R.L., 2015a. Phylogeny and polyploidy: resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 85, 97–116. doi:10.1016/j.ympev.2015.01.014

Yang, L., Wang, Y., Zhang, Z., He, S., 2015b. Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet fish, *Gymnodiptychus pachycheilus*. Genome Biol Evol 7, 251–261. doi:10.1093/gbe/evu279

Zhang, X., Yue, B., Jiang, W., Song, Z., 2008. The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. Mol Biol Rep 36, 981–991. doi:10.1007/s11033-008-9271-y

Table 1.1. Specimens used in Anchored Enrichment study. Authors of family group names are indicated with number of species in parentheses. Taxonomy assignments follow Mayden and W.J. Chen (2010), Tang et al. (2013a), Kottelot (2013), van der Laan et al. (2014) and L. Yang et al. (2015a) with one new subfamily, Eosominae. AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; IHB = Institute of Hydrobiology, Chinese Academy of Sciences; SLUM = St. Louis University Museum; CTOL = Cypriniformes Tree of Life Project; NCMNS = North Carolina Museum of Natural Sciences.

Order	Suborder	**Family (# of spp.)**	Subfamily	Tribe	Species	Specimen voucher
Outgroup						
Characiformes					*Pygocentrus nattereti*	AUFT 3043
Gymnotiformes					*Electrophorus electricus*	AUFT 3843
Siluriformes					*Callichthys callichthys*	AUFT 4774
Ingroup						
Cypriniformes						
	Cobitoidei					
		Balitoridae Swainson 1839 (~92)				
					Homaloptera ogilviei	SLUM B90.T114
		Botiidae Borg 1940 (56)				
					Sinibotia robusta	UAIC 14182.21
					Yasuhikotakia lecontei	AUFT 5061
		Catostomidae Agassiz 1850 (83)				
			Catostominae Agassiz 1850			
				Catostomini Agassiz 1850		
					Catostomus bernardini	SLUM 1558.04
					Catostomus cahita	SLUM 1562.03
					Catostomus leopoldi	SLUM MXSp09-674
					Catostomus platyrhynchus	AUFT 0183
					Catostomus plebeius	SLUM 1633.02
					Catostomus wigginsi	SLUM MXSp09-574
				Erimyzonini Hubbs 1930		
					Erimyzon oblongus	SLUM B21.T-1553
				Moxostomatini Bleeker 1863		
					Minytrema melanops	SLUM B46.T-4436
				Thoburnini Hubbs 1930		
					Thoburnia atripinnis	SLUM B21.T-1616
					Thoburnia rhothoeca	SLUM B21.T-1619
			Ictiobinae Bleeker 1863			
					Ictiobus niger	SLUM B78.10088
		Cobitidae Swainson 1838 (198)				
					Acantopsis sp.	UAIC 14310
					Cobitis biwae	CTOL 00224
					Lepidocephalichthys hasselti	CTOL 03230
					Misgurnus bipartitus	IHB 0411008
					Pangio anguillaris	SLUM B90.T139
		Gastromyzontidae Fowler 1905 (137)				
					Beaufortia kweichowensis	SLUM B90.T118
					Pseudogastromyzon myersi	UAIC 14169.22
		Gyrinocheilidae Gill 1905 (3)				
					Gyrinocheilus aymonieri	AUFT 5008
		Nemacheilidae Regan 1911 (658)				
					Lefua echigonia	SLUM B89.T006

Table 1.1 (*continued*). Specimens used in Anchored Enrichment study. Authors of family group names are indicated with number of species in parentheses. Taxonomy assignments follow Mayden and W.J. Chen (2010), Tang et al. (2013a), Kottelot (2013), van der Laan et al. (2014) and L. Yang et al. (2015a) with one new subfamily, Eosominae. AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; IHB = Institute of Hydrobiology, Chinese Academy of Sciences; SLUM = St. Louis University Museum; CTOL = Cypriniformes Tree of Life Project; NCMNS = North Carolina Museum of Natural Sciences.

Order	Suborder	**Family (# of spp.)**	Subfamily	Tribe	Species	Specimen voucher
					Nemacheilus corica	UAIC 14167.55
					Paracanthocobitis botia	CTOL 03287
					Schistura fasciolata	CTOL 00257
		Vaillantellidae Nalbant and Bănărescu 1977 (26)				
					Vaillantella maassi	CTOL 03437
	Cyprinoidei					
		Acheilognathidae Bleeker 1863 (75)				
					Acheilognathus tonkinensis	AUFT 6614
		Cyprinidae Rafinesque 1815 (1,623)				
			Acrossocheilinae L. Yang et al. 2015a			
					Acrossocheilus monticola	CTOL 00272
			Barbinae Bleeker 1859			
					Barbus barbus	UAIC 14167.25
					Capoeta aculeata	CTOL 03281
					Cyprinion semiplotum	CTOL 01499
			Cyprininae Rafinesque 1815			
					Carassioides acuminatus	SLUM B89.T037
					Cyprinus carpio	SLUM B89.T027
			Labeoninae Bleeker 1859			
					Akrokolioplax bicornis	SLUM B69.D8
					Barbichthys laevis	CTOL 02310
					Cirrhinus cirrhosus	SLUM B91.T178
					Cirrhinus microlepis	CTOL 01558
					Crossocheilus latius	CTOL 01569
					Crossocheilus reticulatus	CTOL 01561
					Garra flavatra	SLUM B91.T179
					Garra rufa	CTOL 03282
					Garra waterloti	CTOL 03174
					Gibelion catla	SLUM B90.T092
					Gymnostomus siamensis	CTOL 02856
					Labeo rohita	CTOL 01610
					Labeo senegalensis	CTOL 03175
					Labiobarbus leptochilus	CTOL 03347
					Lobocheilos melanotaenia	CTOL 01612
					Osteochilus vittatus	CTOL 01697
					Placocheilus dulongensis	SLUM B69.B6
					Schismatorhynchos nukta	CTOL 03180
			Poropuntiinae Menon 1999			
					Albulichthys albuloides	CTOL 01543
					Amblyrhynchichthys truncatus	CTOL 01545
					Barbonymus gonionotus	CTOL 01550
					Barbonymus schwanenefeldii	CTOL 01652
					Cosmochilus harmandi	CTOL 01560
					Cyclocheilichthys enolplos	CTOL 01495

Table 1.1 (*continued*). Specimens used in Anchored Enrichment study. Authors of family group names are indicated with number of species in parentheses. Taxonomy assignments follow Mayden and W.J. Chen (2010), Tang et al. (2013a), Kottelot (2013), van der Laan et al. (2014) and L. Yang et al. (2015a) with one new subfamily, Eosominae. AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; IHB = Institute of Hydrobiology, Chinese Academy of Sciences; SLUM = St. Louis University Museum; CTOL = Cypriniformes Tree of Life Project; NCMNS = North Carolina Museum of Natural Sciences.

Order	Suborder	**Family (# of spp.)**	Subfamily	Tribe	Species	Specimen voucher
					Dischorodontus ashmeadi	CTOL 03207
					Mystacoleucus obtusirostris	CTOL 01618
					Poropuntius normani	CTOL 3918
					Sawbwa resplendens	SLUM B89.T050
			Probarbinae Yang et al. 2015			
					Catlocarpio siamensis	CTOL 01557
					Probarbus jullieni	CTOL 01623
			Schizopygopsinae Mirza 1991			
					Gymnodiptychus integrigymnatus	SLUM B69.DC0376
			Schizothoracinae McClelland 1842			
					Oreinus dulongensis	SLUM B69.B4
					Percocypris tchangi	SLUM B69.DC0344
			Smiliogastrinae Bleeker 1863			
					Chagunius chagunio	SLUM B90.T093
					Dawkinsia filamentosus	CTOL 01511
					Haludaria fasciata	UAIC 14169.14
					Hampala dispar	UAIC 14167.43
					Oreichthys cosuatis	UAIC 14167.48
					Pethia nigrofasciata	CTOL 01514
					Puntius sophore	SLUM B90.T121
					Rohtee ogilbii	CTOL 00449
			Spinibarbinae L. Yang et al. 2015a			
					Spinibarbus caldwelli	CTOL 03193
			Torinae Karaman 1971			
					Labeobarbus compiniei	SLUM B90.T152
					Tor tambroides	UAIC 14182.02
		Danionidae Bleeker 1863 (~330)				
			Chedrinae Bleeker 1863			
					Chelaethips bibie	CTOL 03156
					Leptocypris niloticus	CTOL 03165
					Luciosoma setigerum	CTOL 01614
					Opsaridium ubangiense	AUFT 5799
					Opsarius koratensis	CTOL 03285
					Opsarius koratensis	AUFT 6617
					Opsarius pulchellus	SLUM B87.D5
					Opsarius tileo	AUFT 3793
					Raiamas senegalensis	AUFT 5433
					Salmostoma phulo	CTOL 00316
					Securicula gora	CTOL 03439
			Danioninae Bleeker 1863			
					Chela cachius	CTOL 00329
					Danio feegradei	CTOL 03198
					Danio margaritatus	AUFT 6618
					Danio rerio	Reference Genome

Table 1.1 (*continued*). Specimens used in Anchored Enrichment study. Authors of family group names are indicated with number of species in parentheses. Taxonomy assignments follow Mayden and W.J. Chen (2010), Tang et al. (2013a), Kottelot (2013), van der Laan et al. (2014) and L. Yang et al. (2015a) with one new subfamily, Eosominae. AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; IHB = Institute of Hydrobiology, Chinese Academy of Sciences; SLUM = St. Louis University Museum; CTOL = Cypriniformes Tree of Life Project; NCMNS = North Carolina Museum of Natural Sciences.

Order	Suborder	**Family (# of spp.)**	Subfamily	Tribe	Species	Specimen voucher
					Danio tinwini	AUFT 6619
					Danionella mirifica	CTOL 01954
					Danionella priapus	AUFT 6620
					Devario aequipinnatus	AUFT 6615
					Inlecypris auropurpurea	CTOL 01582
					Laubuka caeruleostigmata	CTOL 03205
					Laubuka laubuca	SLUM 06-060 (#081)
					Microdevario kubotai	UAIC 14166.24
					Microdevario nanus	CTOL 01616
					Microrasbora rubescens	CTOL 01583
					Neochela dadiburjori	CTOL 00330
			Esominae New Subfamily			
					Esomus danrica	AUFT 3811
			Rasborinae Günther 1868			
					Amblypharyngodon mola	SLUM B91.T198
					Horadandia atukorali	CTOL 01604
					Rasbora borapetensis	AUFT 6621
					Rasbora rubrodorsalis	UAIC 14175.07
					Trigonopoma pauciperforatum	AUFT 6622
		Gobionidae Bleecker 1863 (206)				
					Abbottina rivularis	CTOL 00259
					Coreoleuciscus splendidus	CTOL 01559
					Gnathopogon strigatus	CTOL 01759
					Gobio gobio	SLUM B12.T61
					Pseudorasbora parva	CTOL 00478
					Pungtungia herzi	CTOL 00483
					Rhinogobio typus	CTOL 00536
					Romanogobio albipinnatus	SLUM B12.T053
					Squalidus chankaensis	CTOL 01739
		Leuciscidae Bonaparte 1835 (657)				
					Acrocheilus alutaceus	AUFT 0194
					Alburnoides bipunctatus	CTOL 01752
					Alburnus alburnus	SLUM B12.T047
					Campostoma anomalum	AUFT 6108
					Chrosomus eos	AUFT 6624
					Clinostomus funduloides	AUFT 6616
					Cyprinella callistia	AUFT 6628
					Ericymba amplamala	AUFT 0033
					Erimonax monachus	NCMNS 61165
					Erimystax insignis	AUFT 6631
					Exoglossum maxillingua	AUFT 6627
					Gila nigrescens	SLUM
					Hybognathus hankinsoni	AUFT 6625
					Hybopsis amblops	AUFT 6633

Table 1.1 (*continued*). Specimens used in Anchored Enrichment study. Authors of family group names are indicated with number of species in parentheses. Taxonomy assignments follow Mayden and W.J. Chen (2010), Tang et al. (2013a), Kottelot (2013), van der Laan et al. (2014) and L. Yang et al. (2015a) with one new subfamily, Eosominae. AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; IHB = Institute of Hydrobiology, Chinese Academy of Sciences; SLUM = St. Louis University Museum; CTOL = Cypriniformes Tree of Life Project; NCMNS = North Carolina Museum of Natural Sciences.

Order	Suborder	**Family (# of spp.)**	Subfamily	Tribe	Species	Specimen voucher
					Leuciscus leuciscus	SLUM B12.T33
					Luxilus chrysocephalus	AUFT 5982
					Lythrurus bellus	AUFT 0593
					Macrhybopsis storeriana	AUFT 0007
					Nocomis biguttatus	AUFT 6626
					Notemigonus crysoleucas	AUFT 6632
					Notropis longirostris	AUFT 0048
					Opsopoeodus emiliae	SLUM B43.T4247
					Oreoleuciscus humilis	CTOL 00446
					Phenacobius catostomus	AUFT 6629
					Phoxinus oxycephalus jouyi	CTOL 00469
					Phoxinus phoxinus	SLUM B91.T187
					Pimephales vigilax	AUFT 6630
					Ptychocheilus oregonensis	AUFT 0202
					Rhinichthys atratulus	SLUM B58.T6246
					Richardsonius balteatus	AUFT 0166
					Rutilus rutilus	SLUM B12.T041
					Semotilus atromaculatus	AUFT 5949
					Squalius lepidus	CTOL 03284
		Paedocyprididae Mayden and W.J. Chen 2010 (3)				
					Paedocypris cf. *progenetica*	AUFT 6623
		Sundadanionidae Mayden and W.J. Chen 2010 (8)				
					Sundadanio axelrodi "red"	CTOL 01723
		Tanichthyidae Mayden and Chen 2009 (3)				
					Tanichthys micagemmae	SLUM B91.T205
		Xenocyprididae Günther 1868 (159)				
					Aphyocypris normalis	CTOL 01619
					Chanodichthys erythropterus	SLUM 06-093
					Ctenopharyngodon idella	CTOL 00337
					Elopichthys bambusa	CTOL 03186
					Hemigrammocypris neglectus	CTOL 03199
					Hypophthalmichthys molitrix	CTOL 03276
					Macrochirichthys macrochirus	CTOL 01615
					Metzia lineata	SLUM B89.T58
					Nipponocypris sieboldii	CTOL 00604
					Nipponocypris temmincki	CTOL 00605
					Opsariichthys bidens	CTOL 00448
					Parabramis pekinensis	CTOL 00459
					Parachela siamensis	CTOL 03246
					Paralaubuca sp.	SLUM B87.TA5
					Squaliobarbus curriculus	CTOL 00569
					Yaoshanicus arcus	CTOL 01747
					Zacco platypus	CTOL 00602

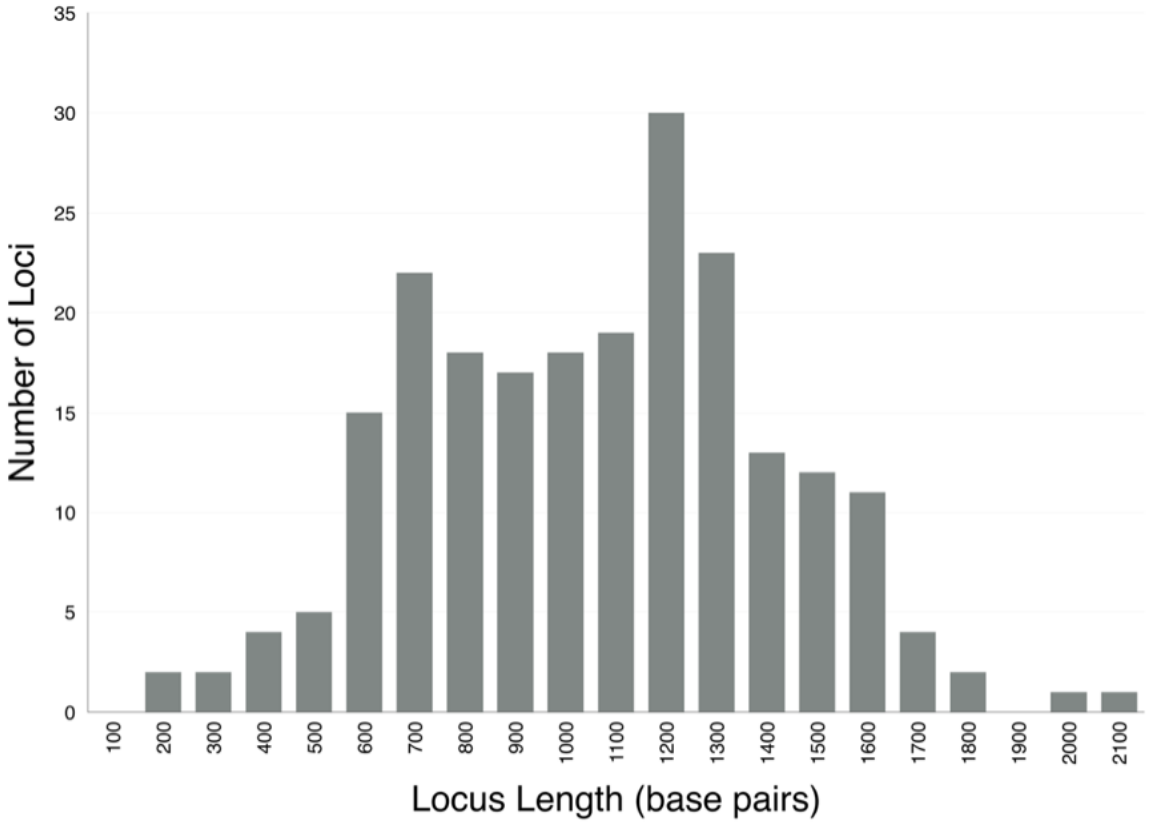

Figure 1.1. Histogram showing lengths of loci in base pairs.

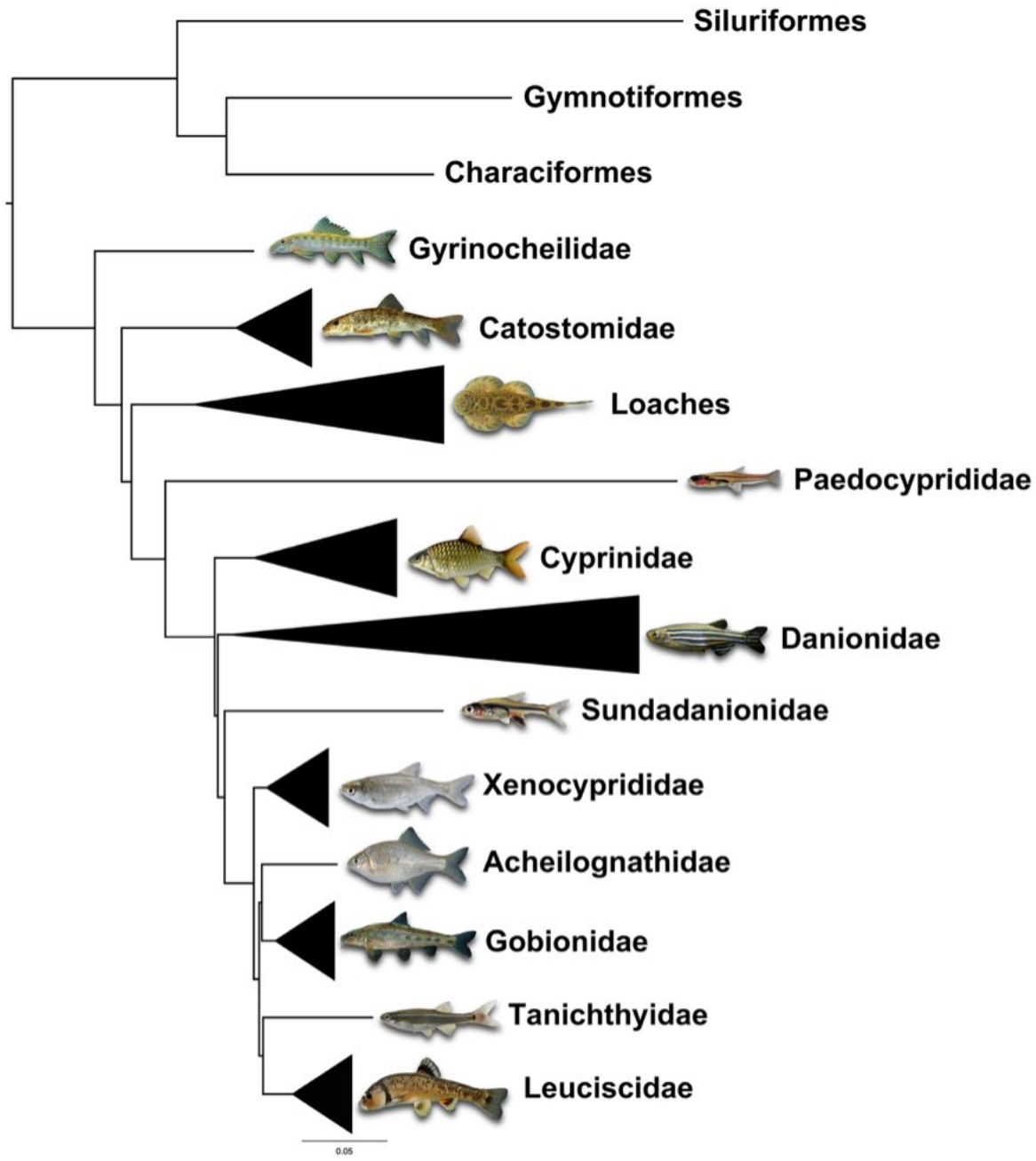

Figure 1.2. Maximum likelihood tree based on concatenation and collapsed into major clades. All nodes shown are 100% bootstrap supported unless otherwise indicated. Scale bar represents the number of nucleotide substitutions per site.

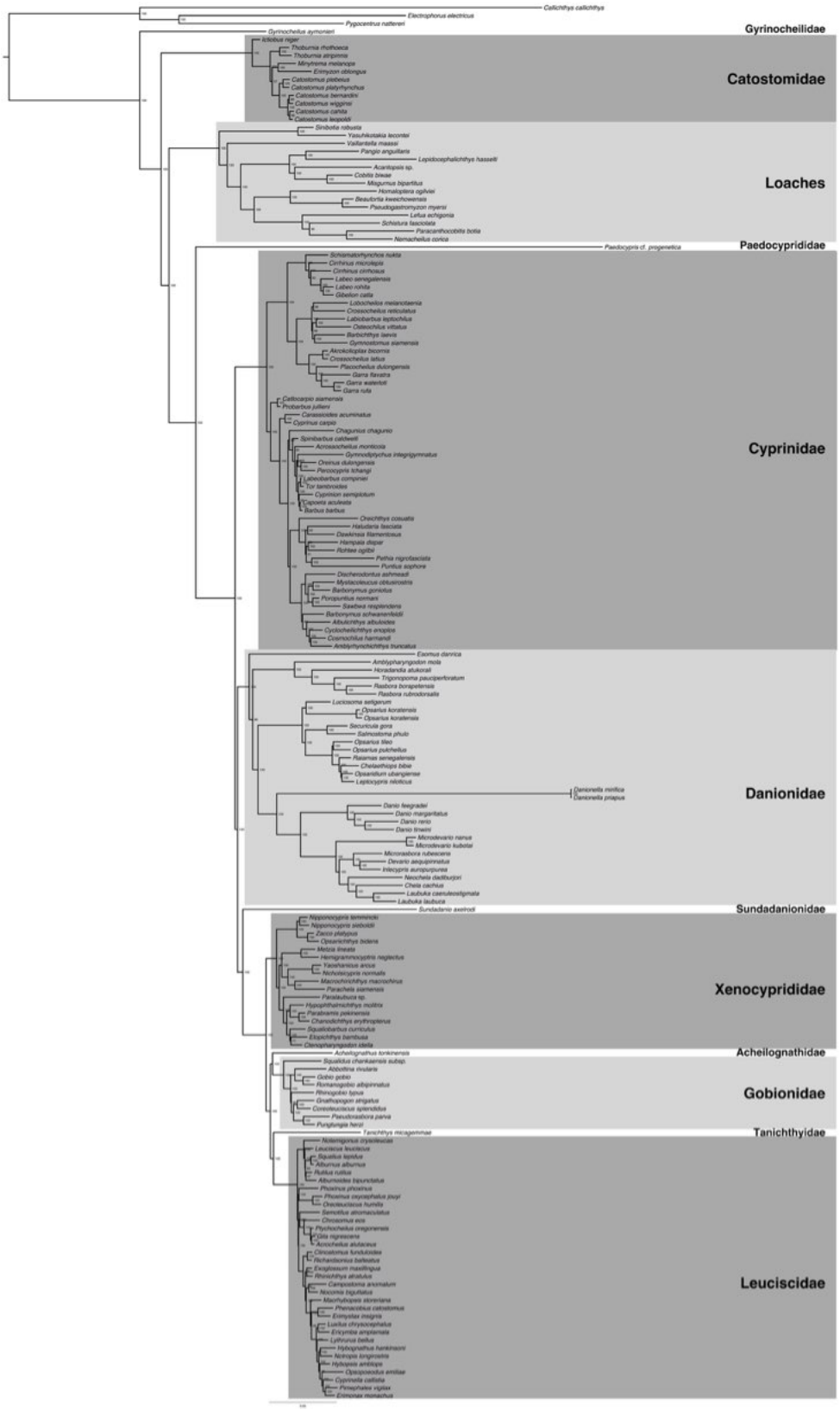

Figure 1.3. Maximum likelihood tree for concatenated dataset of 172 ingroup and three outgroup taxa, fully expanded.

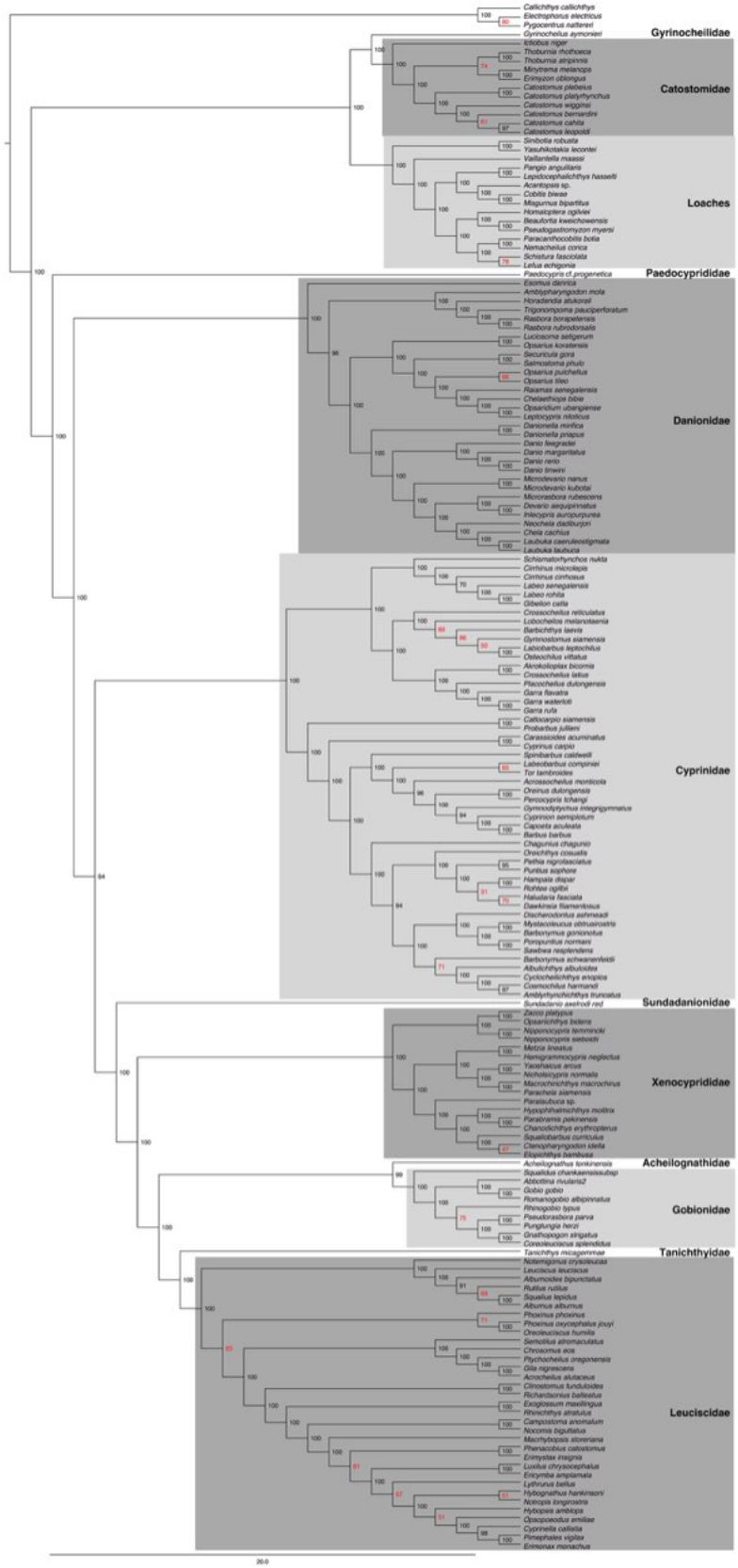

Figure 1.4. Species tree for all taxa, fully expanded, using STAR.

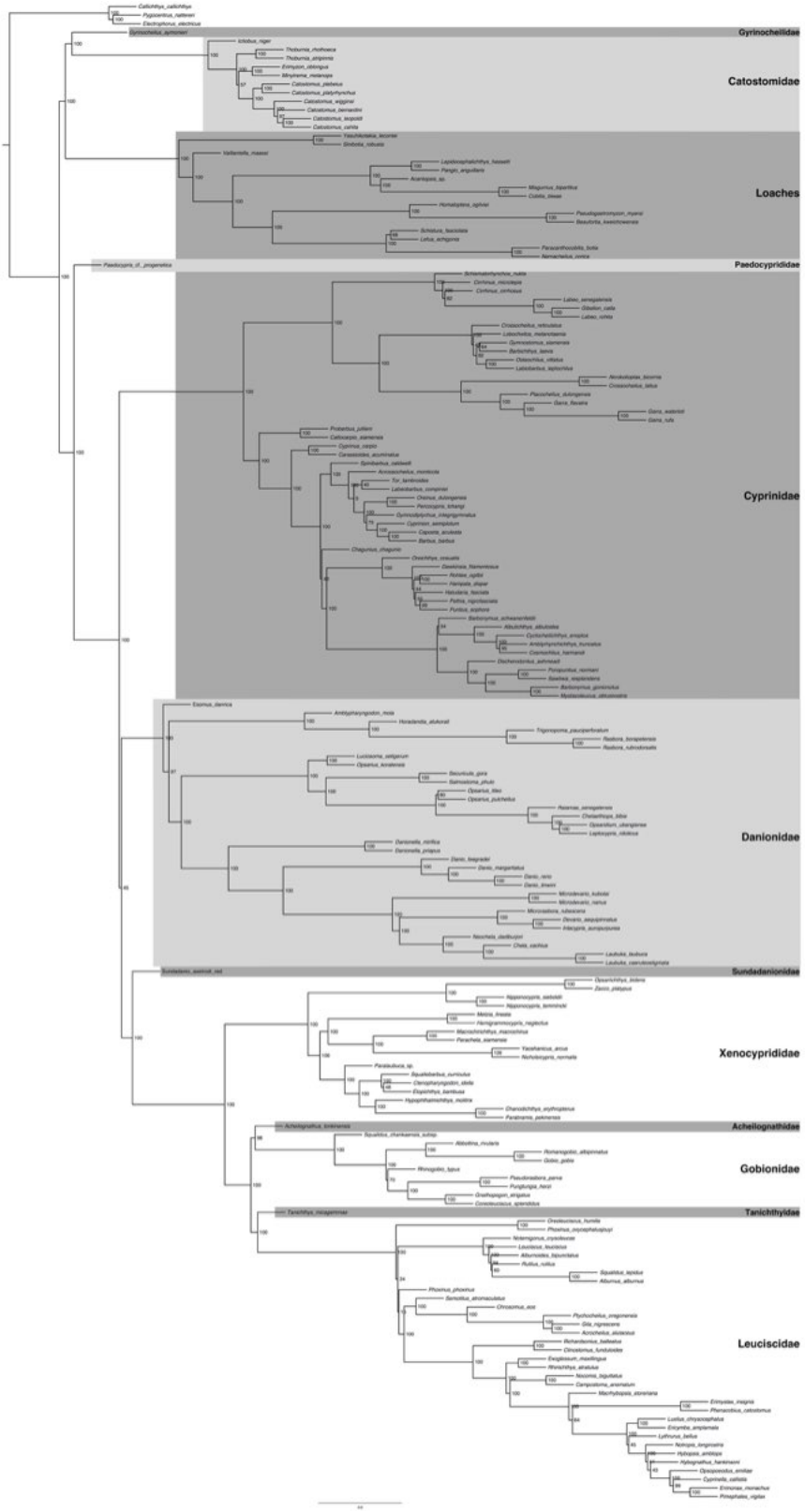

Figure 1.5. Species tree for all taxa, using ASTRAL. Internal branch lengths are in coalescent units and branches that lead to tips are not calculated by ASTRAL but instead arbitrarily displayed.

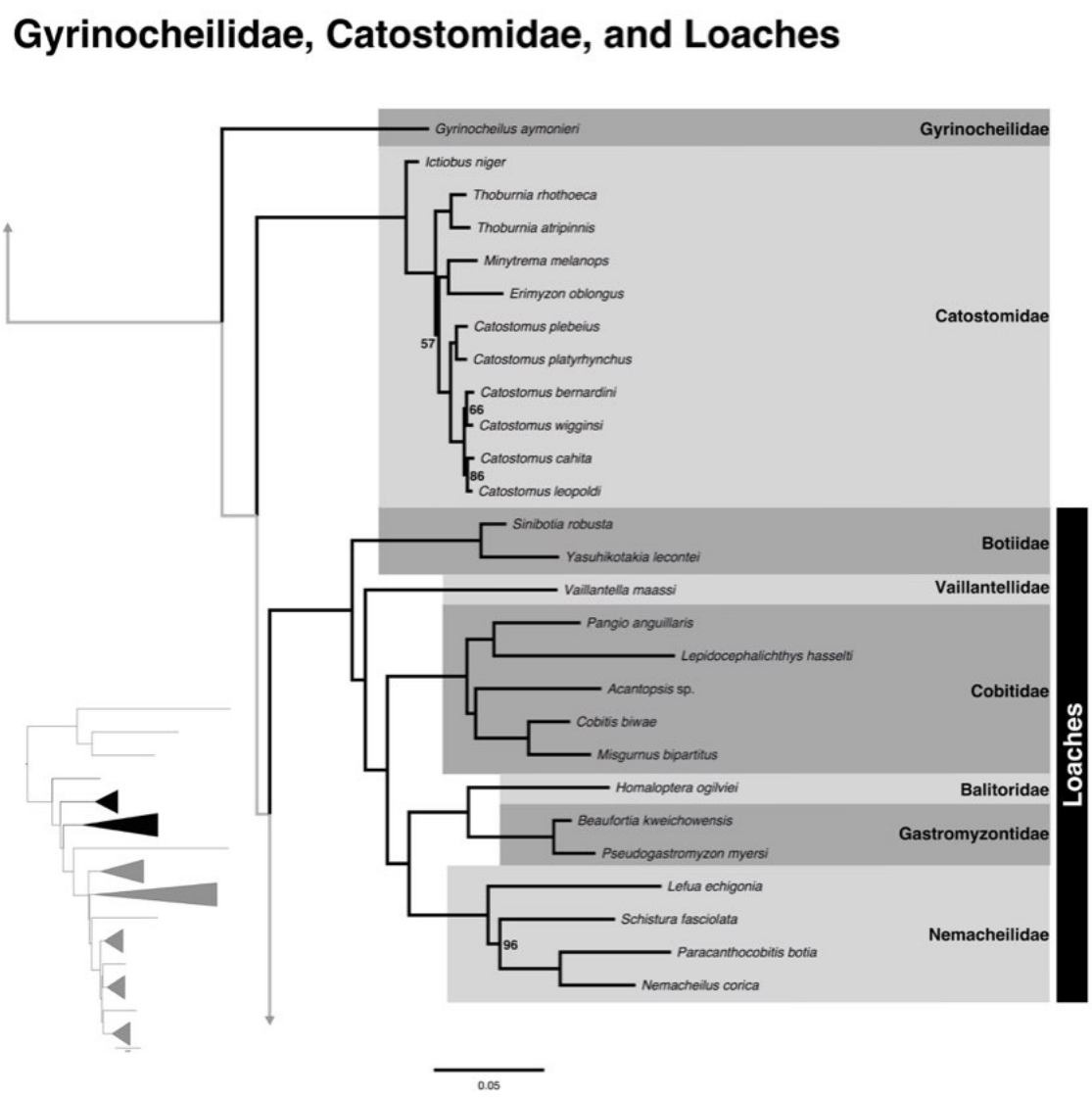

Figure 1.6. Expansion of Cobitoidei families from Figure 1.2 (inset). All nodes are 100% bootstrap supported unless otherwise indicated.

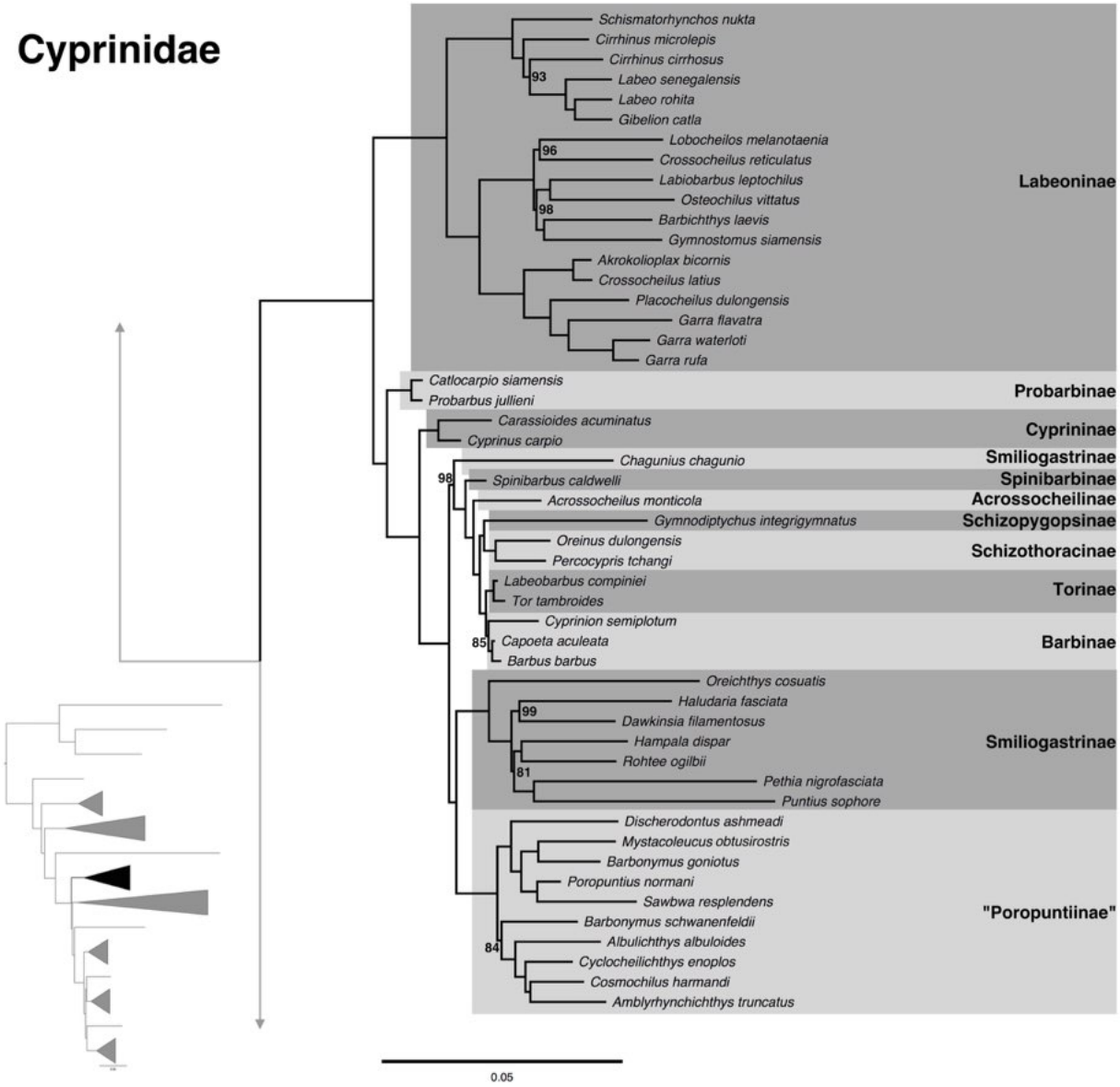

Figure 1.7. Expansion of Cyprinidae from Figure 1.2 (inset). All nodes are 100% bootstrap supported unless otherwise indicated.

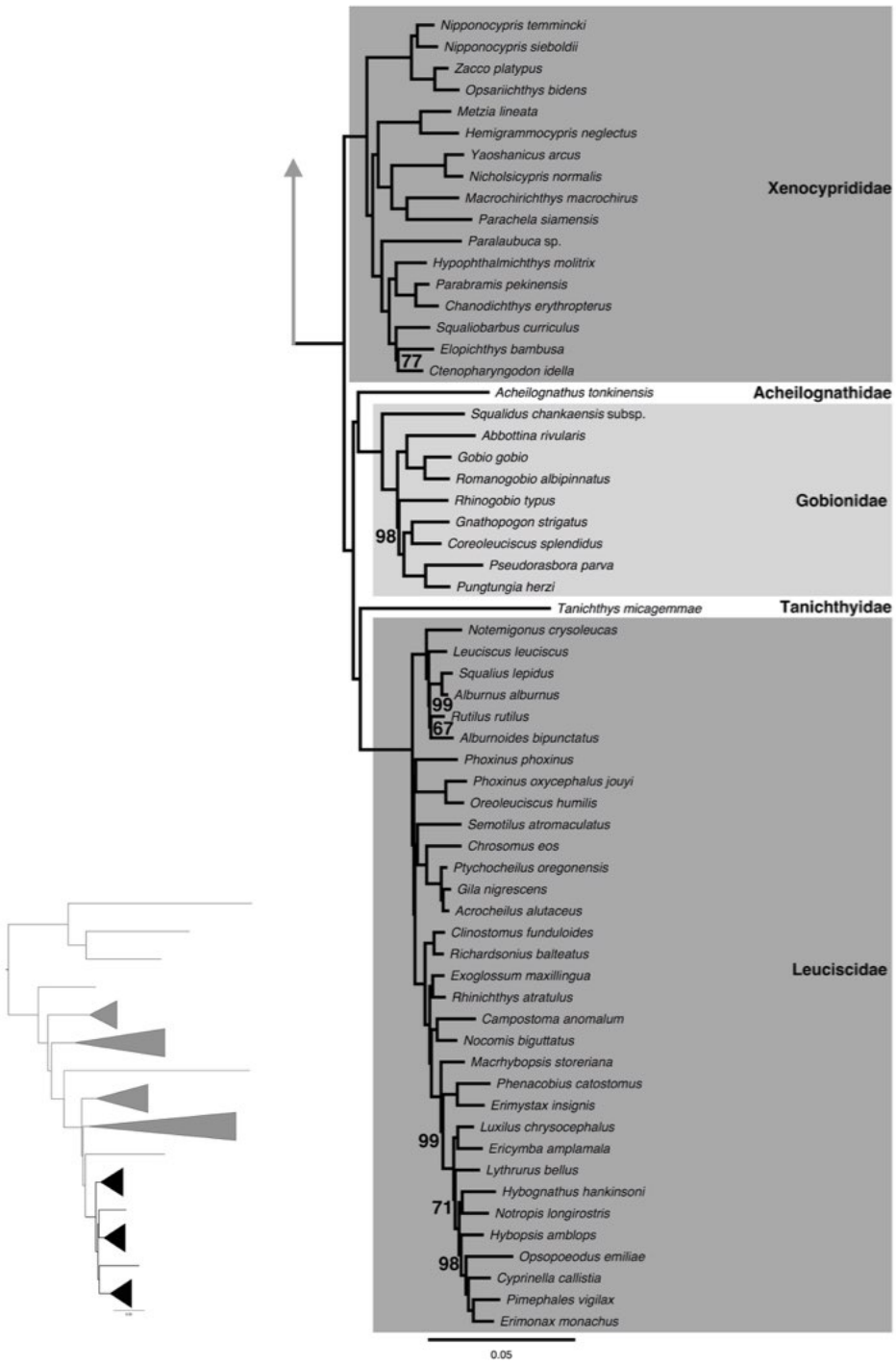

Figure 1.8. Expansion of Xenocyprididae, Acheilognathidae, Gobionidae, Tanichthyidae, and Leuciscidae from Figure 1.2 (inset). All nodes are 100% bootstrap supported unless otherwise indicated.

CHAPTER TWO

AN EVALUATION OF RELATIONSHIPS WITHIN THE FAMILY LEUCISCIDAE (CYPRINIFORMES: CYPRINOIDEI) WITH INSIGHT INTO BIOGEOGRAPHICAL PATTERNS

INTRODUCTION

The North American flora and fauna was enriched by the movement of organisms from Europe and Asia through various connections that existed between the continents. Three different connections have existed between the continents, and these vary in timing, extent, and latitude. Asia and North America have been variously connected to one another through the Bering land bridge (or Beringia) during many periods since the Cretaceous. Beringia connected Siberia and Alaska, and has gone through successive periods of exposure since the Cretaceous. Europe and North America have been connected by the De Geer and Thulean Bridges (Brikiatis, 2014). The De Geer Bridge was a northerly route from Scandinavia, through the Barents Sea, and across northern Greenland. The De Geer route is thought to have been exposed from the late Cretaceous to the early Paleocene [~71 to 63 million years ago (mya)] (McKenna, 1983; Tiffney, 1985; Brikiatis, 2014). The Thulean Route is a connection across England and Ireland to southern Greenland and northern Canada. The Thulean Route was likely exposed for brief periods in the late Paleocene (~57 mya) and early Eocene (~56 mya) (McKenna, 1983; Brikiatis, 2014), and there is paleontological and geological evidence that land connections between Scotland, the Faroe Islands, Iceland, Greenland, and North America may have been intermittently exposed even as late as 20-6 mya (Denk et al., 2011). Dated phylogenies can help to determine which routes were taken by organisms colonizing North American by correlating

molecular clock estimates with estimates of the exposure of the various land bridges. Given the more common exposure of Beringia, molecular dates that do not include exposure of the European routes preclude movement of organisms from Europe into North America. We examine the effect of these land bridges on the formation of the North American fish fauna by providing a dated analysis of minnows based on the phylogeny of Stout et al. (2016; Chapter 1), and also examine other aspects of their biogeography.

With ~310 species, the Leuciscidae accounts for a large portion of the freshwater fish diversity in North America. A further ~340 species occur in Europe and Asia, and the species are commonly referred to as chubs, shiners, and minnows. While the family as a whole has been supported as monophyletic across many studies (Briolay et al., 1998; Cunha et al., 2002; Liu and Chen, 2003; Saitoh et al., 2006; Thai et al., 2007; Wang et al., 2007; Chen et al., 2008; Fang et al., 2009; Gaubert et al., 2009; Mayden et al., 2009; Mayden and Chen, 2010; Tang et al., 2010; Wang et al., 2012), relationships among the genera and clades of leuciscids have had differing results. Certain clades have been recovered in multiple studies, including the (1) Eurasian phoxinines, (2) leuciscines, (3) open posterior myodome (OPM), (4) creek chub - plagopterin (CC-P), and (5) western North America (WNA) clades. The leuciscines are found primarily in Europe with a distribution that extends into western Asia. Interestingly, within the leuciscines, there is one species, *Notemigonus crysoleucas*, found in North America. Eurasian phoxinines are found in Asia and extend into eastern Europe. The three remaining clades of phoxinines (OPM, CC-P, and WNA) are all found exclusively in North America (Figures 2.1−2.6). Relationships among these clades have varied in different analyses, and many hypotheses have been proposed based on both molecular and morphological data [Cavender and Coburn, 1992; Cunha et al.,

2002; Saitoh et al., 2006; Sakai et al., 2006; Rüber et al., 2007; Sasaki et al., 2007; Wang et al., 2007; Zhang et al., 2008; Strange and Mayden, 2009; Bufalino and Mayden, 2010a; 2010b; 2010c; Saitoh et al., 2011; Imoto et al., 2013; see Imoto et al. (2013) for summary]. Of the most recent analyses, Bufalino and Mayden (2010b) used two mitochondrial (12S and 16S) and two nuclear (RAG1 and S7) loci while Imoto et al. (2013) used entire mitogenomes to infer relationships among these clades. Both resulted in conflicting topologies with low to moderate support for nodes among the major clades. The mitogenomic analysis (Imoto et al., 2013) inferred biogeographical patterns based on these weakly supported relationships and failed to include the CC-P clade in their analysis. Having a clear, well-supported hypothesis for these relationships is vital to inferring the biogeographical history of the group and for explaining their current distributions.

Cavender (1991) reviewed the fossil record for Cyprinoidei (previously Cyprinidae) and reported that the oldest North American leuciscid fossil was 31 million years old. Based on the fossil evidence, he hypothesized, along with several zoogeographers, an Asian origin for North American leuciscids with subsequent movement of Leuciscidae ancestors into the more northern latitudes of Europe and northern Asia and eastward across Beringia through one of its exposures in the mid-Oligocene.

Imoto et al. (2013), using their mitogenomic assessment of the relationships within the Leuciscidae, inferred a completely different biogeographic explanation that directly contradicted Cavender's (1991) hypothesis. Imoto et al. (2013) used ancestral state reconstruction of range for their taxa, but exclude the outgroup as is considered standard practice in biogeographical analyses. This is generally recommended because phylogenetic distance between the ingroup and

outgroup may be large or unknown, and because outgroup taxa may be widely distributed. Imoto et al. (2013) proposed a European origin for Leuciscidae (~71 mya), followed by movement into North America (~68 mya) and finally westward movement across Beringia into Asia (~62 mya) (Figure 2.7).

The ability to infer biogeographical patterns requires a strongly supported evolutionary hypothesis of relationships among the taxa of interest. With the advent of phylogenomics, relationships among problematic taxa are beginning to find resolution. Included among these taxa are the relationships among members of Cypriniformes (Stout et al., 2016; Chapter 1). The objective of this study is to use the phylogenomic assessment of Cypriniformes presented by Stout et al. (2016) with a focus on the relationships recovered among members of Leuciscidae to estimate divergence times and reevaluate the biogeography of this family in order to make comparisons with the results presented by Imoto et al. (2013). Given that the analysis of Stout et al. (2016) shows strong support for the outgroup taxa to the Leuciscidae and that all outgroup members are found in one geographic region, we further test the effect of including and excluding the outgroup in ancestral state reconstructions of range and find dramatic differences in interpretation.

METHODS

Divergence Time Estimation

The phylogeny used for this study was published by Stout et al. (2016) and is based on 219 concatenated single-copy nuclear loci with an average length of 1011 bp for a total of 315,288 bp. A subset of this dataset was used to focus on the Leuciscidae, with inclusion of taxa

from Acheilognathidae, Gobionidae, and Tanichthyidae in order to include more fossil calibration points. This resulted in 33 ingroup and 11 outgroup taxa. BEAST v1.8 (Drummond and Rambaut, 2007) was used for divergence time estimation with an alignment that was partitioned by the 219 loci. The tree topology was constrained to reflect the relationships for Leuciscidae previously established by the full Cypriniformes phylogeny (Stout et al., 2016). Four chains with a length of 60 million generations were run and sampled every 6000 generations. Fossil calibration included the oldest North American fossil dated at 31.1 mya (Cavender, 1991), a *Gnathopogon* (Gobionidae) fossil from the Miocene (23–5.3 mya) (Jiajian, 1990), the oldest Gobionidae fossil from 33.9 mya (Jiajian, 1990), and the oldest known cyprinoid fossil, *Parabarbus* (Cyprinidae) from the Eocene (Sytchevskaya, 1986) as a maximum calibration point for priors (55.8–33.9 mya).

Ancestral State Reconstructions

Reconstruction of ancestral geographic distributions was carried out on the BEAST maximum clade credibility (MCC) tree using both a maximum likelihood (DEC; dispersal-extinction cladogenesis) and Bayesian (BBM; bayesian binary MCMC) approach in RASP (Yu et al., 2015). For both analyses, the maximum number of ancestral areas was set at two and the regions were defined as 1) Asia, 2) Europe, and 3) North America. BBM analysis was carried out using 50,000 generations with 10 chains sampled every 100 generations (discarding 100 trees as burn-in) under the Jukes-Cantor fixed-state frequencies model with among-site variation set to equal. Traditionally, exclusion of outgroup taxa has been recommended for historical biogeographic analysis for two reasons; outgroup taxa may represent widely distributed species,

and phylogenetic distance from the ingroup may either be unknown or large. For this dataset, all outgroup taxa can be coded under the Asia region, and phylogenetic distance is known based on the overall topology recovered by Stout et al. (2016) that sampled across the entire order. Because of these factors, analyses were also conducted using the same parameters above, but with the inclusion of the outgroup.

RESULTS

Divergence Time Estimation

Our analysis results in an estimated age of 41 mya [47.1−36.1 mya 95% highest posterior density (HPD) of divergence time estimates] for the Leuciscidae (Figure 2.8). Divergences of the major subclades all occur within a very short time span ranging from within zero to 17 million years. The split between leuciscines and all other members of Leuciscidae occurred approximately 38 mya (43.8−34.0 HPD). The divergence between Eurasian phoxinines and the remaining North American taxa occurred just half a million years later (37.5 mya; 43.2−33.6 HPD), and between the OPM clade and CC-P + WNA clade half a million years after that (37 mya; 42.8−33.0 HPD). The most recent divergence between subclades (CC-P and WNA clades) is estimated to have occurred approximately 34 mya (41.2−26.3 HPD). The estimate for all of these diversification events (~45−25 mya) corresponds primarily to the late Eocene, with ranges extending through the Oligocene.

Ingroup-only ancestral state reconstructions

The historical biogeographical distributions inferred from both the maximum likelihood

(DEC) and bayesian (BBM) analyses on only the ingroup (Figure 2.9) reflect similar patterns with a few differences at key nodes. In the DEC analysis, nodes A (origin of Leuciscidae) and B (most recent common ancestor of all North American taxa except *Notemigonus crysoleucas*) are recovered as 100% probability for an Asian/North American distribution. For node C (most recent common ancestor of leuciscines), DEC recovers 100% probability of a European/North American distribution. BBM analysis recovers node A as 86.18% probability of a North American, 6.27% probability of a European, and 1.93% probability of an Asian distribution. The distribution probabilities for node B under BBM are 67.70% North America, 24.13% Asia, and 7.27% Asia/North America. BBM Node C distribution probabilities are: 64.81% North America, 26.64% Europe, and 7.79% Europe/North America.

Ancestral state reconstructions with outgroup inclusion

Unsurprisingly, inclusion of the outgroup in both DEC and BBM analyses recover very different distribution probabilities for key nodes A, B, and C compared to ingroup-only analyses (Figure 2.10). For the origin of Leuciscidae, both recover a partial Asian distribution probability (DEC: 100% Asia/Europe; BBM: 71.89% Asia, 15.06% North America, 11.02% Asia/North America). Both also recover an Asian/North American distribution for the most recent common ancestor to North American taxa (except *Notemigonus crysoleucas*; node B). The results of DEC and BBM differ at node C, however, with DEC recovering 100% European distribution probability and BBM recovering 56.50% North America, 19.68% Europe, 12.79% Asia, and 5.88% Europe/Asia distribution probabilities.

DISCUSSION

Incongruence of ancestral state reconstruction analyses

For analyses excluding the outgroup, DEC infers an ancestral population for Leuciscidae in Asia and North America, while BBM strongly suggests North America (Figure 2.9, node A), which is similar to that found by Imoto et al. (2013) using similar methodology. A North American origin hypothesis is contradictory to studies that have a broader scope across Cyprinoidei (Cavender, 1991, Saitoh et al., 2011) using both molecular and fossil data that hypothesize an Asian or Eurasian origin for Leuciscidae. Various morphological and molecular studies across Cypriniformes have confirmed a sister relationship between Leuciscidae and Asian taxa, with the only variable being which family (Gobionidae, Acheilognathidae, or Tanichthyidae) is sister (for example Chen et al., 1984; Cavender and Coburn, 1992; Wang et al., 2007; Chen and Mayden, 2009; Tang et al., 2010; Chen et al., 2013; Dahanukar et al., 2013; Tang et al., 2013; Tao et al., 2013). We include representatives for all of these families in our outgroup, and the relationships are well supported based on the overall Cypriniformes topology (Stout et al., 2016). This leads us to question the generalized practice of excluding outgroup taxa, particularly in our analyses where phylogenetic distance from the outgroup is well established and all outgroup taxa are found in one region instead of widely dispersed. Interpretation of ancestral biogeographical ranges seems highly dependent on the scope of the question; for example, if our original question had not focused on Leuciscidae alone but on the entire clade leading to families Acheilognathidae, Gobionidae, Tanichthyidae, and Leuciscidae (as our results in Figure 2.10), we would recover quite different probabilities and biogeographical patterns for Leuciscidae than the analyses using only Leuciscidae as the ingroup. For these reasons, we feel

justified in focusing further discussion on analyses that include the outgroup (Figure 2.10), but recognize the need for further research, perhaps with simulation studies, regarding the influence of outgroup exclusion on ancestral state reconstruction analyses when phylogenetic distance is known and outgroup taxa are not widely dispersed.

Congruencies between the outgroup-included analyses include support for an Asian origin for Leuciscidae and an Asia/North America distribution for ancestors of North American clades. They differ at node C (ancestors of the leuciscine clade), with DEC reporting a European ancestral distribution with the ancestor of *Notemigonus crysoleucas* dispersing to North America, and BBM results suggesting a primarily North American distribution for ancestors, followed by colonization and subsequent diversification in Europe. The latter is less likely, however, unless it is assumed that the ancestors of *Notemigonus* failed to diversify at a time when other North American and European clades were diversifying, or if there were diversification events, that all other species went extinct except for the ancestor to *Notemigonus*. Further, *Notemigonus* is generally found nested within the European taxa of Leuciscinae (for example Sasaki et al., 2007; Perea et al., 2010; Dahanukar et al., 2013; Tang KL et al., 2013), but taxa that would demonstrate that were not included in our analyses. The dramatic effect of the lack of broad sampling in the leuciscinae on ancestral state reconstruction of ranges can be strongly seen in the BBM analysis without the outgroup, which suggests a high probability for a North American origin for the Leuciscidae. Thus, ancestral state reconstructions of range are also sensitive to taxon sampling, and more complete taxon sampling would serve to strengthen arguments as to the origins of the clades of the Leuciscidae.

Asian origin for Leuciscidae followed by expansion into Europe and northern Asia

The relationships recovered by Stout et al. (2016), in conjunction with divergence times and ancestral distributions established by this study, suggest a biogeographical and temporal pattern more congruent with that of Cavender (1991) than that of Imoto et al. (2013; Figure 2.6). This includes an Asian origin for Leuciscidae, followed by expansions into western Europe and northern Asia and finally from northern Asia across Beringia into North America. Our age estimates are much younger than those proposed by Imoto et al. (2013; 71 mya) and span the Eocene and Oligocene. Geologic and climatic characteristics of these epochs are very plausible in helping to explain how members of Leuciscidae may have come to have their current distributions.

During the Eocene (55−34 mya), higher latitudes experienced warmer climates that could have allowed species to migrate further into Europe and northern Asia. Laurasia began to break up, although there remained a land connection between Europe, Greenland, and North America until approximately 50 mya (with support for intermittent connections from 20−6 mya; see below). In Europe, the Eocene/Oligocene boundary is marked by the Grande Coupure extinction event where much of the European fauna were replaced by Asian species, and this could have further facilitated leuciscine movement into Europe.

Movement across Beringia and diversification in North America

The warmer climates at higher latitudes that allowed migration into northern Asia also allowed movement across Beringia into North America, and faunal exchanges are known to have occurred during the Eocene (Tiffney, 1985; Sanmartín et al., 2001). Transition to cooler climates

at the end of the Eocene may have halted this migration. Proposed dates in the late Eocene correlate with the end of the thermal optimum of the Eocene and the growing glacial cycle that began at this time. Growing mountains in the western part of the continent during the Oligocene may have played a role in the subsequent diversifications of the North American clades (WNA, OPM, and CC-P). During the late Oligocene, volcanic activity and tectonic movement resulting in rifts along far western North America produced fragmentation and constant habitat shifts, perhaps leading to extinctions that could explain the relative paucity of Leuciscidae species in this area today compared to what is found east of the Mississippi (Willis, 1909).

A second migration to North America

The placement of the North American species, *Notemigonus crysoleucas*, in our phylogeny as more closely related to European leuciscids than to other North American taxa in not unusual. This species is repeatedly recovered in the European clade across many different studies (Cavender and Coburn, 1992; Cunha et al., 2002; Saitoh et al., 2006; Sakai et al., 2006; Rüber et al., 2007; Sasaki et al., 2007; Wang et al., 2007; Zhang et al., 2008; Strange and Mayden, 2009; Bufalino and Mayden, 2010a; 2010b; 2010c; Saitoh et al., 2011; Imoto et al., 2013). The strong consensus of its phylogenetic position and diversification date led Böhme (2000) to infer that a transatlantic route must have existed during the early Miocene, and there are some geological and paleontological studies that support the intermittent existence of a Scotland-Faroe-Iceland-Greenland-North America land bridge anywhere from 20−6 mya (see Denk et al., 2011 for review). This aligns closely with our estimate that *Notemigonus* split from its European sisters ~23.6 mya (15−35 HPD) with a range that overlaps with the estimate given

by Perea et al. (2010; 29.07 mya), as opposed to the older date given by Imoto et al (2013; 37.1 mya). The intermittent nature of each component of this land bridge could explain why only one leuciscid species was successful at using this route for colonization of North America from Europe. The timing of this diversification also eliminates the possibility that this species could have used the Thulean route (~56 mya) or the Van Geer route (~62 mya) for movement into North America. However, Beringia experienced several exposures as well. Few leuciscines are in eastern Asia suggesting that the movement of the ancestor of *Notemigonus* was from Europe.

CONCLUSIONS

Our findings support the biogeographical hypothesis first proposed by Cavender (1991) that included Asian origin for Leuciscidae, expansion into Europe and northern Asia, and finally movement across Beringia into North America, with divergences of major clades occurring in quick succession of approximately half a million years starting in the late Eocene. The ancestor to *Notemigonus* came later, most likely through a European route. The distribution of these major clades across the northern hemisphere has sparked much interest in elucidating the biogeographical history of the family, but hypotheses were either hampered or obscured by the difficulties associated with producing a robust phylogenetic framework for a rapidly diversifying group. This study illustrates the necessity for such a robust framework and the importance of phylogenomic data for clades traditionally difficult to resolve, such as those within Cypriniformes.

References

Böhme, M., 2000. Die Cypriniden (Teleostei: Cypriniformes) des oberoligozänen Maares von Enspel nebst Bemerkungen zur Phylogenie und Biogeographie der Phoxininae. Paläontologische Zeitschrift 74, 99–112. doi:10.1007/BF02987956

Brikiatis, L., 2014. The De Geer, Thulean and Beringia routes: key concepts for understanding early Cenozoic biogeography. J. Biogeogr. 41, 1036–1054. doi:10.1111/jbi.12310

Briolay, J., Galtier, N., Brito, R.M., Bouvet, Y., 1998. Molecular Phylogeny of Cyprinidae Inferred from cytochrome b DNA Sequences. Molecular Phylogenetics and Evolution 9, 100–108.

Bufalino, A.P., Mayden, R.L., 2010a. Molecular phylogenetics of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) based on RAG1 and S7 nuclear DNA sequence data. Molecular Phylogenetics and Evolution 55, 274–283. doi:10.1016/j.ympev.2009.12.017

Bufalino, A.P., Mayden, R.L., 2010b. Phylogenetic evaluation of North American Leuciscidae (Actinopterygii: Cypriniformes: Cyprinoidea) as inferred from analyses of mitochondrial and nuclear DNA sequences. Systematics and Biodiversity 8, 493–505. doi:10.1080/14772000.2010.516029

Bufalino, A.P., Mayden, R.L., 2010c. Phylogenetic relationships of North American phoxinins (Actinopterygii: Cypriniformes: Leuciscidae) as inferred from S7 nuclear DNA sequences. Molecular Phylogenetics and Evolution 55, 143–152. doi:10.1016/j.ympev.2009.11.001

Cavender, T.M., 1991. The fossil record of the Cyprinidae, in: Winfield, I., Nelson, J.S. (Eds.), Cyprinid Fishes: Systematics, Biology, and Exploitation. Chapman & Hall, pp. 1–12.

Cavender, T.M., Coburn, M.M., 1992. Systematics, Historical Ecology, and North American Freshwater Fishes. Ecology.

Chen, W.J., Mayden, R.L., 2009. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: Further evidence from six nuclear genes. Molecular Phylogenetics and Evolution 52, 544–549. doi:10.1016/j.ympev.2009.01.006

Chen, W.J., Lavoué, S., Mayden, R.L., 2013. Evolutionary origin and early biogeography of otophysan fishes (Ostariophysi: Teleostei). Evolution 67, 2218–2239. doi:10.1111/evo.12104

Chen, W.J., Miya, M., Saitoh, K., Mayden, R.L., 2008. Phylogenetic utility of two existing and four novel nuclear gene loci in reconstructing Tree of Life of ray-finned fishes: The order

Cypriniformes (Ostariophysi) as a case study. Gene 423, 125–134. doi:10.1016/j.gene.2008.07.016

Chen, X.L., Yue, P.Q., Lin, R.D., 1984. Major groups within the family Cyprinidae and their phylogenetic relationships. Acta Zootaxon. Sinica 9, 424-440.

Cunha, C., Mesquita, N., Dowling, T.E., Gilles, A., Coelho, M.M., 2002. Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences 61, 1–16. doi:10.1006/jfbi.2002.2105

Dahanukar, N., Philip, S., Krishnakumar, K., Ali, A., Raghavan, R., 2013. The phylogenetic position of Lepidopygopsis typus (Teleostei: Cyprinidae), a monotypic freshwater fish endemic to the Western Ghats of India. Zootaxa 3700, 113. doi:10.11646/zootaxa.3700.1.4

Denk, T., Grímsson, F., Zetter, R., Símonarson, L.A., 2011. The Biogeographic History of Iceland – The North Atlantic Land Bridge Revisited, in: Late Cainozoic Floras of Iceland, Topics in Geobiology. Springer Netherlands, Dordrecht, pp. 647–668. doi:10.1007/978-94-007-0372-8_12

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7, 214. doi:10.1186/1471-2148-7-214

Fang, F., Norén, M., Liao, T.Y., Källersjö, M., Kullander, S.O., 2009. Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). Zoologica Scripta 38, 237–256. doi:10.1111/j.1463-6409.2008.00373.x

Gaubert, P., Denys, G., Oberdorff, T., 2009. Genus-level supertree of Cyprinidae (Actinopterygii: Cypriniformes), partitioned qualitative clade support and test of macro-evolutionary scenarios. Biological Reviews 84, 653–689. doi:10.1111/j.1469-185X.2009.00091.x

Imoto, J.M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y.P., Miya, M., Nishida, M., Hanzawa, N., 2013. Phylogeny and biogeography of highly diverged freshwater fish species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. Gene 514, 112–124. doi:10.1016/j.gene.2012.10.019

Jiajian, Z., 1990. The Cyprinidae fossils from the middle Miocene of Shanwang Basin. Certebrata Palasiatica. 2:002

Liu, H., Chen, Y., 2003. Phylogeny of the East Asian cyprinids inferred from sequences of the mitochondrial DNA control region. Can. J. Zool. 81, 1938–1946. doi:10.1139/z03-169

Mayden, R.L., Chen, W.J., 2010. The world“s smallest vertebrate species of the genus *Paedocypris*: a new family of freshwater fishes and the sister group to the world”s most

diverse clade of freshwater fishes (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 57, 152–175. doi:10.1016/j.ympev.2010.04.008

Mayden, R.L., Chen, W.J., Bart, H.L., Doosey, M.H., Simons, A.M., Tang, K.L., Wood, R.M., Agnew, M.K., Yang, L., Hirt, M.V., Clements, M.D., Saitoh, K., Sado, T., Miya, M., Nishida, M., 2009. Reconstructing the phylogenetic relationships of the earth's most diverse clade of freshwater fishes—order Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and the mitochondrial genome. Molecular Phylogenetics and Evolution 51, 500–514. doi:10.1016/j.ympev.2008.12.015

McKenna, M.C., 1983. Cenozoic paleogeography of North Atlantic land bridges. Structure and Development of the Greenland-Scotland Ridge 351–399. doi: 10.1007/978-1-4613-3485-9_19

Perea, S., Böhme, M., Zupančič, P., Freyhof, J., Šanda, R., Özuluğ, M., Abdoli, A., Doadrio, I., 2010. Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. BMC Evol Biol 10, 265. doi:10.1186/1471-2148-10-265

Rüber, L., Kottelat, M., Tan, H., Ng, P.K., Britz, R., 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. BMC Evolutionary Biology 7, 38. doi:10.1186/1471-2148-7-38

Saitoh, K., Sado, T., Doosey, M.H., Bart, H.L., Jr, Inoue, J.G., Nishida, M., Mayden, R.L., Miya, M., 2011. Evidence from mitochondrial genomics supports the lower Mesozoic of South Asia as the time and place of basal divergence of cypriniform fishes (Actinopterygii: Ostariophysi). Zoological Journal of the Linnean Society 161, 633–662. doi:10.1111/j. 1096-3642.2010.00651.x

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M., 2006. Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. Journal of Molecular Evolution 63, 826–841. doi:10.1007/s00239-005-0293-y

Sakai, H., Ito, Y., Shedko, S.V., Safronov, S.N., Frolov, S.V., Chereshnev, I.A., Jeon, S.-R., Goto, A., 2006. Phylogenetic and taxonomic relationships of northern Far Eastern phoxinin minnows, *Phoxinus* and *Rhynchocypris* (Pisces, Cyprinidae), as inferred from allozyme and mitochondrial 16S rRNA sequence analyses. Zool. Sci. 23, 323–331.

Sasaki, T., Kartavtsev, Y.P., Chiba, S.N., Uematsu, T., Sviridov, V.V., Hanzawa, N., 2007. Genetic divergence and phylogenetic independence of Far Eastern species in subfamily Leuciscinae (Pisces: Cyprinidae) inferred from mitochondrial DNA analyses. Genes Genet. Syst. 82, 329–340.

Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biological Journal of the Linnean Society 73, 345–390.

Stout, C.C., Tan, M., Lemmon, A.R., Lemmon, E.M., armbruster, J.W., 2016. Resolving Cypriniformes relationships using an anchored enrichment approach. BMC Evolutionary Biology 1–13. doi:10.1186/s12862-016-0819-5

Strange, R.M., Mayden, R.L., 2009. Phylogenetic Relationships and a Revised Taxonomy for North American Cyprinids Currently Assigned to *Phoxinus* (Actinopterygii: Cyprinidae). Copeia 2009, 494–501. doi:10.1643/CI-07-070

Sytchevskaya, E.K., 1986. Paleogene freshwater fish fauna of the USSR and Mongolia. Proceedings of the Joint Soviet-Mongolian Paleontological Expedition.

Tang, K.L., Agnew, M.K., Hirt, M.V., Lumbantobing, D.N., Sado, T., Raley, M.E., Teoh, V.-H., Bart, H.L., Yang, L., Harris, P.M., He, S., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2013. Limits and phylogenetic relationships of East Asian fishes in the subfamily Oxygastrinae (Teleostei: Cypriniformes: Cyprinidae). Zootaxa 3681, 101–135. doi:10.11646/zootaxa.3681.2.1

Tang, K.L., Agnew, M.K., Hirt, M.V., Sado, T., Schneider, L.M., Freyhof, J., Sulaiman, Z., Swartz, E., Vidthayanon, C., Miya, M., Saitoh, K., Simons, A.M., Wood, R.M., Mayden, R.L., 2010. Systematics of the subfamily Danioninae (Teleostei: Cypriniformes: Cyprinidae). Molecular Phylogenetics and Evolution 57, 189–214. doi:10.1016/j.ympev.2010.05.021

Tao, W., Mayden, R.L., He, S., 2013. Remarkable phylogenetic resolution of the most complex clade of Cyprinidae (Teleostei: Cypriniformes): a proof of concept of homology assessment and partitioning sequence data integrated with mixed model Bayesian analyses. Molecular Phylogenetics and Evolution 66, 603–616. doi:10.1016/j.ympev.2012.09.024

Thai, B.T., Si, V.N., Phan, P.D., Austin, C.M., 2007. Phylogenetic evaluation of subfamily classification of the Cyprinidae focusing on Vietnamese species. Aquat. Living Resour. 20, 143–153. doi:10.1051/alr:2007025

Tiffney B.H. 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. Journal of the Arnold Arboretum.

Willis, B., 1909. Paleogeographic Maps of North America. The Journal of Geology. doi: 10.1086/621609

Wang, X., Gan, X., Li, J., Mayden, R.L., He, S., 2012. Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. Sci. China Life Sci. 55, 761–773. doi:10.1007/s11427-012-4366-z

Wang, X., Li, J., He, S., 2007. Molecular evidence for the monophyly of East Asian groups of Cyprinidae (Teleostei: Cypriniformes) derived from the nuclear recombination activating gene 2 sequences. Molecular Phylogenetics and Evolution 42, 157–170. doi:10.1016/j.ympev.2006.06.014

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Molecular Phylogenetics and Evolution 87, 46–49. doi: 10.1016/j.ympev.2015.03.008

Zhang, X., Yue, B., Jiang, W., Song, Z., 2008. The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. Mol Biol Rep 36, 981–991. doi:10.1007/s11033-008-9271-y

Figure 2.1. Approximate distribution of phoxinine species within Leuciscidae.

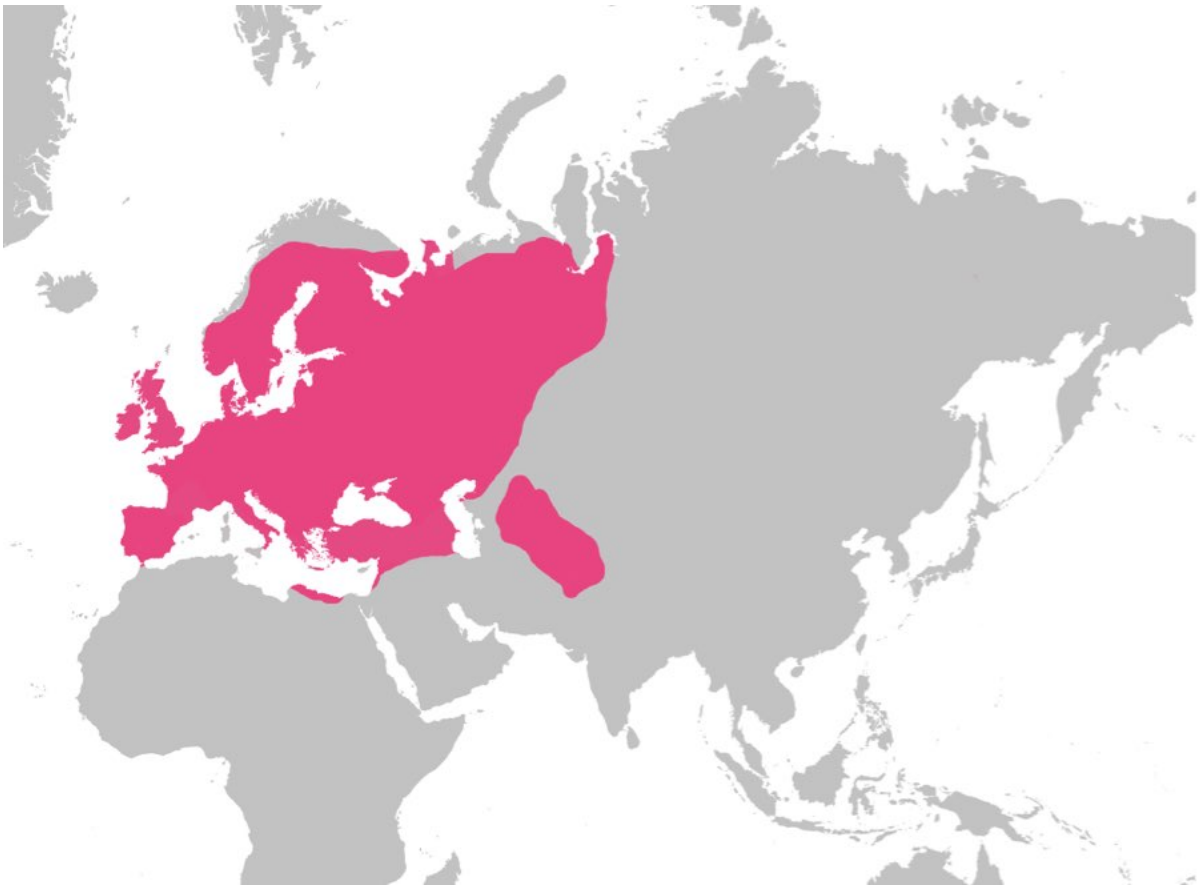

Figure 2.2. Current distribution of leuciscine species within Leuciscidae, with the exception of *Notemigonus crysoleucas* (see Figure 2.3).

Figure 2.3. Current distribution of *Notemigonus crysoleucas* (inset), the only leuciscine species found in North America. Native range is east of the Mississippi RIver, but the species has been introduced into other regions.

Figure 2.4. Current distribution of WNA species within Leuciscidae.

Figure 2.5. Current distribution of CC-P species within Leuciscidae.

Figure 2.6. Current distribution of OPM species within Leuciscidae.

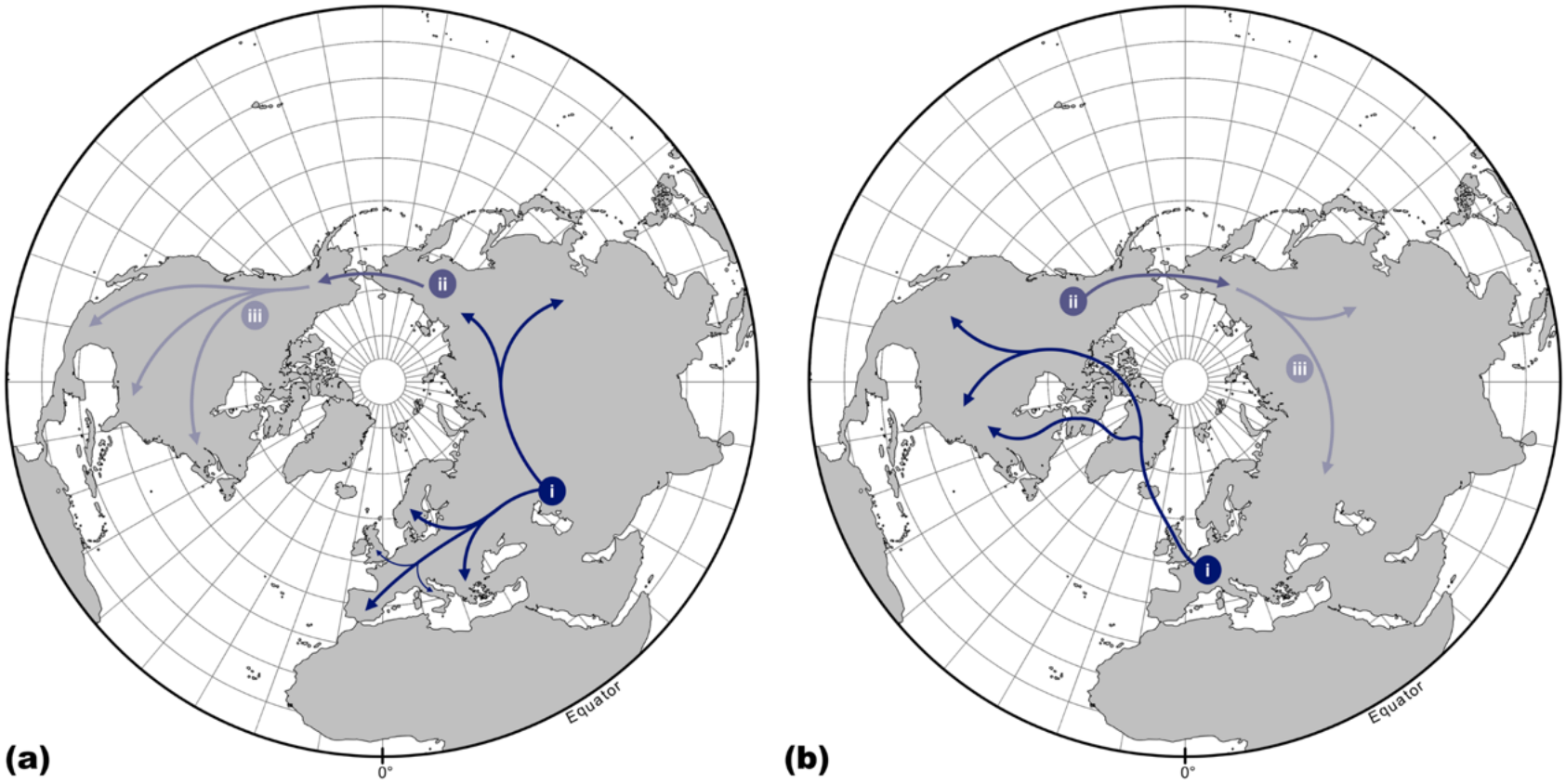

Figure 2.7. Biogeographical hypotheses according to (a) Cavender (1991) and (b) Imoto et al. (2013). Map projection is Albers equal area with the North Pole in the center, and earlier organismal movements are indicated with darker arrows. Cavender [1991;(a)] proposed a Eurasian origin, followed by expansions into Europe and Asia (i), eastward movement across Beringia (ii) and diversification in North America (iii). Imoto et al. [2013; (b)] proposed a European origin, followed by westward movement into North America (i), then across Beringia (ii), and finally into Asia (iii).

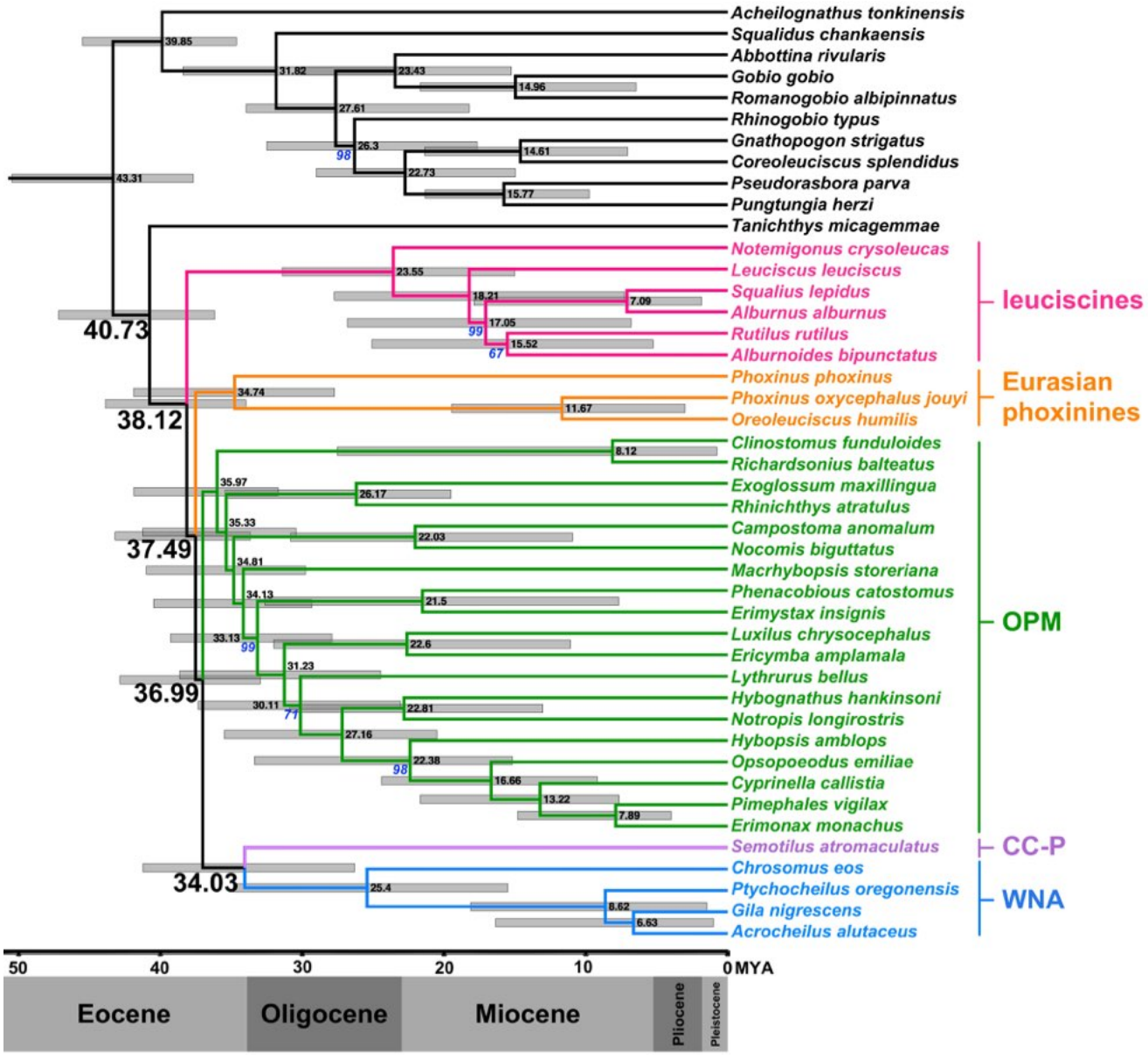

Figure 2.8. Time-calibrated phylogeny based on the topology recovered by Stout et al. (2016). All nodes are 100% bootstrap supported unless otherwise indicated by white numbers. Estimated divergence times in mya are represented by black numbers at nodes. Larger black numbers represent divergence times in mya for major clades of Leuciscidae.

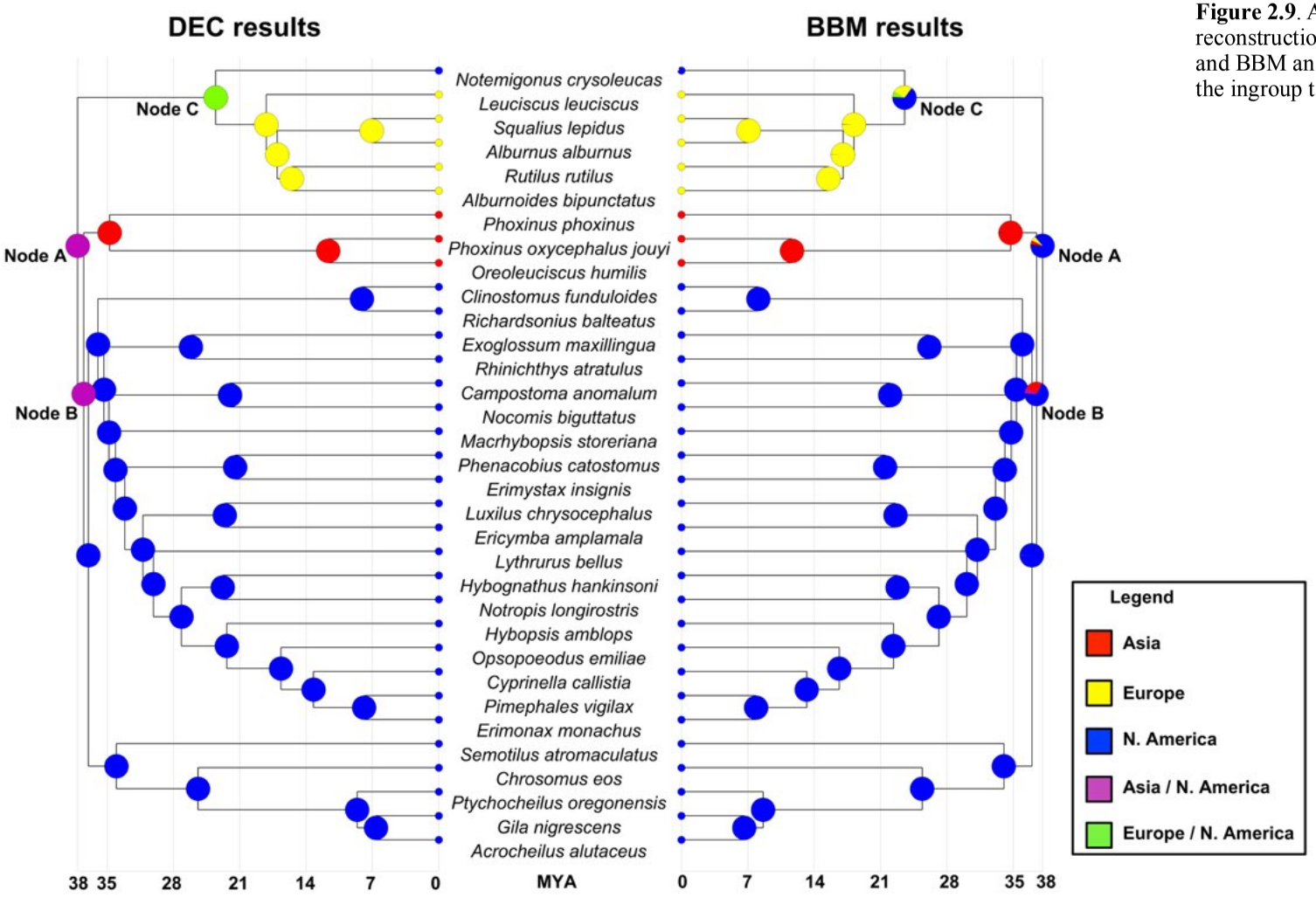

Figure 2.9. Ancestral state reconstruction using DEC and BBM analyses on just the ingroup taxa.

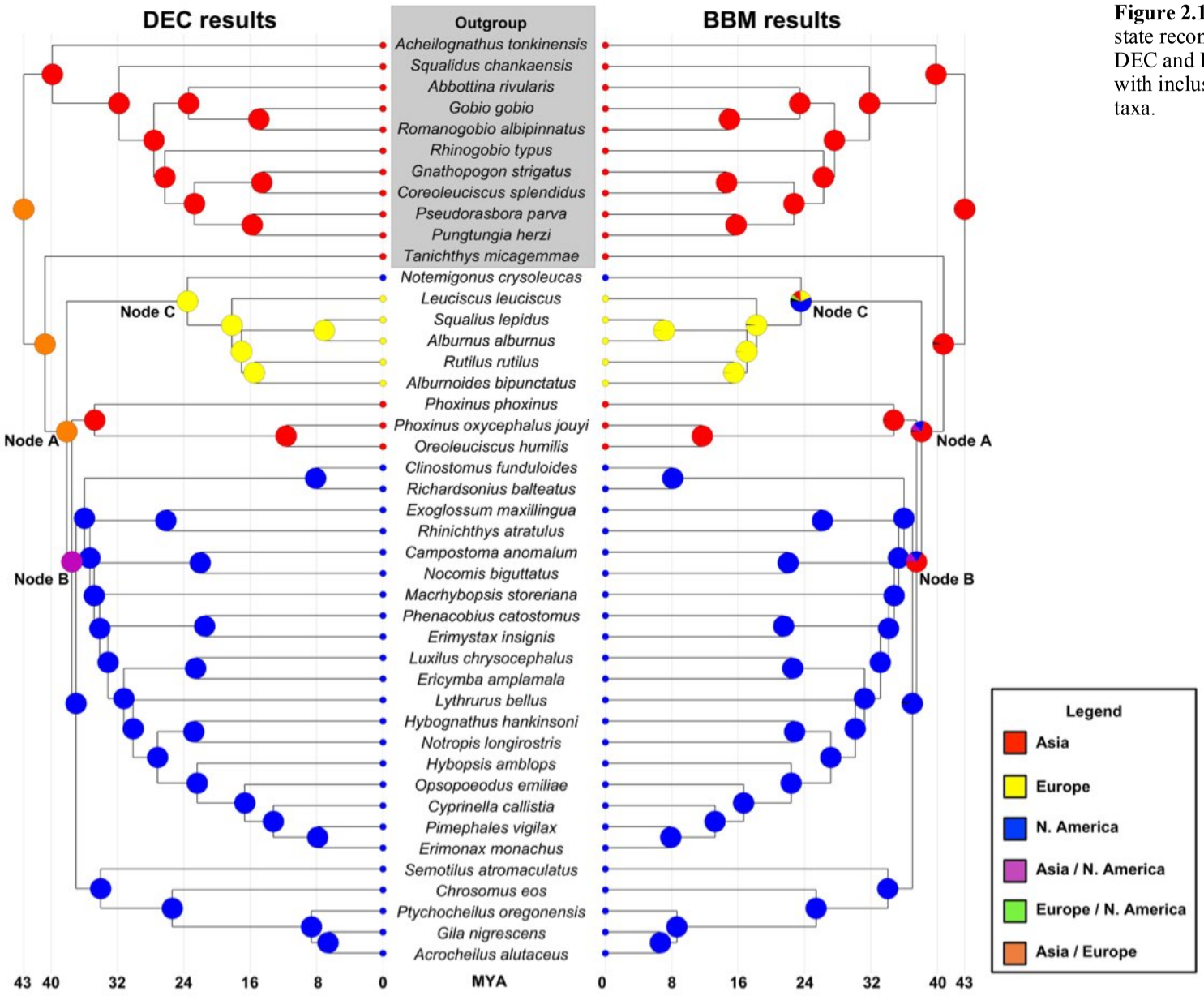

Figure 2.10. Ancestral state reconstruction using DEC and BBM analyses with inclusion of outgroup taxa.

CHAPTER THREE

MOLECULAR SYSTEMATICS OF THE SHINER CLADE (CYPRINIFORMES: LEUCISCIDAE)

INTRODUCTION

Among North American fishes, the shiners and related minnows have been among the most difficult for students of ichthyology to learn. Ichthyologists tasked with assembling the species into meaningful genera initially described a dizzying array of genera and subgenera. Species were moved between these various categories seemingly at random until the community decided to lump all of the taxa into one genus, *Notropis*, which at one time held at least 213 described species. Starting with Mayden (1989), *Notropis* began to be separated into other genera, such as *Cyprinella, Luxilus, Lythrurus,* and *Pimephales.* Still, *Notropis* remained as a "taxonomic repository for small, silvery fishes of unknown relationship" (Gidmark and Simons, 2014:379) of approximately 91 species loosely organized into subgenera (Jordan 1885). Primarily because of the large number of taxa, coupled with conserved morphologies, few have attempted to tackle the remaining species allocated to the genus or other orphaned taxa of unknown taxonomic placement. Even when taxonomic decisions are made (like in Mayden et al., 2006 and Gidmark and Simons, 2014), they are largely ignored (Eschmeyer et al., 2017), leaving a taxonomy that is illogical and in disarray.

Phylogenetically, most previous studies have focused on resolving relationships within subgenera (for example Snelson, 1972; Buth, 1979; Raley et al., 2001; Cashner et al., 2011) with varied results, and without investigation into relationships among the subgenera or to genera that have been removed from *Notropis*. Mayden et al. (2006) attempted one of the most

comprehensive studies to try to resolve these relationships using cytb (mitochondrial marker) and made several conclusions, the most notable being nonmonophyly of *Notropis*, in addition to the recognition of the following genera: *Agosia*, *Alburnops*, *Aztecula*, *Cyprinella*, *Ericymba*, *Graodus*, *Hudsonius*, *Hybognathus*, *Hybopsis*, *Lythrurus*, *Miniellus*, *Pimephales*, and *Yuriria*. Still, many species were relegated to the status of '*Notropis'* because of their uncertain placement due to weak support, and relationships among the genera listed above remained unclear. Despite the recognition of additional genera from within the nonmonophyletic *Notropis*, most subsequent studies reverted back to a larger encompassing *Notropis*, with perhaps recognition of some of these genera as subgenera (for example Bird and Hernandez, 2007; Ruber et al., 2007; Zhang et al., 2008; Chen and Mayden, 2009; Fang et al., 2009; Gaubert et al., 2009; Scott et al., 2009; Bufalino and Mayden, 2010; Houston et al., 2010; Cashner et al., 2011; Wang et al., 2012; Hollingsworth et al., 2013; Imoto et al., 2013; Eschmeyer et al., 2017).

Hollingsworth et al. (2013) expanded upon the cytb study by adding the RAG1 (nuclear) molecular marker to test for a correlation between a shift from benthic to pelagic lifestyles with increased diversification rates based on a recovered phylogenetic reconstruction. Unsurprisingly, this analysis also resulted in a relatively poorly resolved overall phylogeny with moderate support for non-monophyly of *Notropis*, and illustrates the importance of understanding these relationships to better inform our understanding of ecological and evolutionary processes. Morphological analyses have also been attempted, with the most influential being Mayden (1989), who recognized the monophyletic OPM (Open Posterior Myodome) clade, which includes *Notropis* and related genera.

To promote further study into this group, Gidmark and Simons (2014) amassed much of

the knowledge reported for the shiners (distributions, histories, ecologies, etc.) and proposed using the designations made by Mayden et al. (2006) with the understanding that the relationships among them still remain unclear, despite support for the shiner clade as a whole (Simons et al. 2003; Mayden et al. 2006; Schönhuth et al. 2008).

Recent advances in sequencing technologies have provided the opportunity to re-examine the shiner clade using phylogenomic markers. Most phylogenomic-scale studies thus far have focused on higher taxonomic levels (Lemmon et al. 2012; Bond et al. 2014; Eytan et al. 2015; Prum et al. 2015; Hamilton et al. 2016), but decreases in costs and the establishment of universal loci specifically for fishes (Betancur-R et al. 2013; Arcila et al. 2017) have helped overcome the hurdles associated with applying a phylogenomic approach to the shiner clade. In this study, we employ the probes developed by Arcila et al. (2017) in an attempt to tackle the systematic problem of *Notropis* and related genera, and follow the taxonomy discussed in (Gidmark and Simons 2014). Although the exon-capture method of Arcila et al. (2017) showed excellent utility at higher taxonomic scales, this is the first test of the markers at lower taxonomic scales and in a group with what appears to be very rapid divergence.

METHODS

Taxon selection, tissue preparation, and sequencing

Every effort was made to acquire broad representation across shiner genera. Table 3.1 shows genera with number of recognized species, type species, and species sampled. Genera not available included *Tampichthys* (6 species)*, Yuriria* (1 species), *Algansea* (7 species), *Aztecula* (2 species), *Agosia* (2 species), *Dionda* (6 species), and *Erimonax* (1 species). Except for *Agosia*,

Dionda, and *Erimonax* (currently listed as threatened), all other unsampled genera are found exclusively in Mexico, although the Mexican genera *Graodus* and *Codoma* are represented in this study. To test the utility of the markers at a smaller taxonomic scale, we include two specimens of *Notropis atherinoides*, *Ericymba amplamala* and *Pimephales notatus*.

DNA was extracted from 93 ethanol-preserved muscle or fin clips representing 89 ingroup and four outgroup taxa using the Omegabiotek E.Z.N.A. animal tissue extraction kit (product #D3396-02) following manufacturer protocols. Extracted DNA was checked for quality using electrophoresis and quantity using nanodrop. After ensuring high molecular weight and a minimum of 2 µg total DNA, samples were sent for library preparation and Illumina sequencing to MYcroarray (mycroarray.com). Probes developed by (Arcila et al., 2017) were used to target 1060 loci.

Bioinformatics and tree reconstruction

FASTQ files were uploaded to the Alabama Supercomputer Center (ASC) for preliminary quality control processing. Trimmomatic (Bolger et al., 2014) was used to remove adapters and remove leading and trailing low quality bases in the paired end reads, as well as to remove reads with a length less than 36 base pairs. Resulting reads were then imported into Geneious v 6.1.8 (www.geneious.com), set as paired reads, and assembled using the zebrafish (*Danio rerio*) reference for the concatenated loci using five iterations and trimmed to each reference locus. The loci for each species were then concatenated and all concatenations were aligned in Geneious v 6.1.8 (www.geneious.com). Tree reconstruction was performed on the Center for for Advanced Science Innovation and Commerce (CASIC) computer cluster at Auburn University, Auburn, AL,

USA. RAxML was implemented using GTR + G model of evolution on the partitioned loci and the resulting tree then subjected to 500 bootstrap replicates. Species tree reconstruction was conducted using ASTRAL-II (Mirarab and Warnow, 2015) on individual RAxML gene trees that were subjected to 100 bootstrap replicates. Approximately-unbiased (AU) tests were conducted using CONSEL v.0.20 (Shimodaira 2002) to specifically test the unconstrained maximum likelihood best tree topology against trees that were constrained to force monophyly for three genera: *Cyprinella*, *Hudsonius*, and *Luxilus*.

RESULTS

The final alignment yielded 1004 loci, 286,445 base pairs, and only 0.42% missing data. Of those sites, 32,466 (11.33%) were phylogenetically informative. The range for locus size was 196−1748 bp, with an average bp length of 285 (Figure 3.1). In the resulting ML tree, 78% of nodes are 100% bootstrap supported with only six nodes collapsing below the 70% bootstrap threshold. (Figure 3.2). Species tree analysis produced highly congruent results, particularly at the genus level. At deeper nodes there is less support for the placement of a few clades (i.e. *Hudsonius hudsonius* + *H. altipinnis*; '*Notropis*' *atrocaudalis* + '*N.*' *bifrenatus* + '*N.*' *heterolepis*), resulting in remaining uncertainty as to the relationships among the genera. Nevertheless, with our focus on resolving within-genera relationships, both concatenation and species tree approaches resolve the same patterns with strong support.

Unsurprisingly, *Notropis* is found as nonmonophyletic, with *N. jemezanus*, *N. amabilis*, *N. micropteryx*, *N. rubellus*, and *N. amoenus* forming a monophyletic clade with the type species, *N. atherinoides*, while *N. buchanani*, *N. wickliffi*, *N. volucellus*, and *N. spectrunculus* form

another clade. Specimens designated as '*Notropis*' by Mayden et al., (2006) are found throughout the tree. Other nonmonophyletic genera include *Hudsonius*, *Pteronotropis*, *Luxilus*, and *Alburnops*. The majority of *Cyprinella* forms a supported monophyletic clade, with the exception of *C. callistia* forming a polytomy with *Opsopoeodus* + *Pimephales* and the clade containing the remaining members of *Cyprinella* + *Codoma*. The results of the AU test were all significant (*Cyprinella* constrained, p=3e-06; *Hudsonius* constrained, p=6e-08; *Luxilus* constrained, p=2e-18), indicating that all of the constrained topologies can be rejected as alternative tree hypotheses. A list of all species, genera, and proposed taxonomic changes discussed below are given in Table 3.2.

DISCUSSION

Pteronotropis and *Hudsonius*

All three species of *Hudsonius* were included in the analysis but were not recovered as monophyletic, with *H. cummingsae* grouping with *Pteronotropis* and rendering *Pteronotropis* paraphyletic. The range for all three species of *Hudsonius* overlaps with that of *Pteronotropis* across the southeastern states of North and South Carolina, Georgia, and Florida, but only *H. hudsonius* extends northward up through the Great Lakes and across much of Canada. Mayden et al. (2006) found support for a monophyletic *Hudsonius*, but individuals of *H. altipinnis* were not monophyletic, suggesting cryptic speciation. In our analysis, *Hudsonius cummingsae* instead forms a monophyletic clade with members of *Pteronotropis*, while *H. altipinnis* (collected in South Carolina) and *H. hudsonius* (collected in Wisconsin) were found as sister to each other. Because *H. hudsonius* is the type species, we propose moving *Hudsonius cummingsae* to

Pteronotropis to maintain monophyly of *Pteronotropis*.

Luxilus

Our analysis includes seven of the nine recognized species of *Luxilus* and recovers two distinct clades. *Luxilus chrysocephalus* (type species) forms a distinct clade with *L. zonatus*, *L. pilsbryi*, *L. albeolus*, and *L. cornutus* that is sister to *Ericymba* + '*Notropis*' *dorsalis*. *Luxilus coccogenis* and *L. zonistius*, however, are found as a clade distant to other members of *Luxilus* and instead sister to *Hybopsis*. Mayden (1989) removed *Luxilus* from *Notropis*, considering it sister to *Cyprinella* and monophyletic based on three morphological characters, while Coburn and Cavender (1992) considered *Luxilus* to be sister to a clade comprised of *Lythrurus*, *Cyprinella*, *Pimephales*, and *Opsopoeodus*. Molecular studies have primarily focused on members within *Luxilus*, assuming monophyly of the genus instead of including other shiner genera, and have consistently found a sister relationship for *L. coccogenis* + *L. zonistius*, which is supported by our findings (Gilbert 1964; Buth, 1979; Dowling and Naylor, 1997; Mayden et al., 2006), or that *Luxilus* is not monophyletic (Schönhuth and Mayden 2010). Because we include a variety of other shiner taxa, we find that these two species should no longer be considered as part of *Luxilus*, and the genus *Coccotis* Jordan 1882 is resurrected to reflect their distinct placement in our phylogeny, with *Coccotis coccogenis* Jordan 1882 as the type species.

Lythrurus

Lythrurus has long been considered to be monophyletic (Snelson 1972; Schmidt et al., 1998; Mayden et al., 2006; Pramuk et al., 2007), and our findings support monophyly. What has

been more problematic, however, is determining the clade's relationship to other genera. It was considered sister to a *Luxilus* + *Cyprinella* clade by Mayden (1989), but later poorly resolved by Mayden et al., (2006) in a clade with various '*Notropis*' species. Coburn and Cavender (1992) determined *Lythrurus* was sister to a clade comprised of *Cyprinella*, *Pimephales*, and *Opsopoeodus*. We find strong support for *Lythrurus* as sister to true *Notropis* (the clade containing *Notropis atherinoides*, the type species of *Notropis*; more discussion on *Notropis* below).

Cyprinella

One of the most extensive and recent molecular studies concerning *Cyprinella* (Schönhuth and Mayden, 2010) found that the genus was not monophyletic. Most of the species comprised a monophyletic clade, but a sister relationship between these species and *Codoma* + *Tampichthys* placed *Cyprinella callistia* outside of *Cyprinella*, although its exact placement was not fully resolved. Our analysis shows the same pattern. While we do not include *Tampichthys*, we also show that *Codoma* is more closely related to all other representatives of *Cyprinella* than *Cyprinella callistia* is. We could not resolve the node leading to *Cyprinella callistia*, *Opsopoeodus* + *Pimephales*, and *Codoma* + *Cyprinella*, but we clearly show *Cyprinella callistia* should not be included in *Cyprinella*. *Cyprinella callistia* was originally described as *Photogenis callistius* (Jordan 1877), but we are hesitant to resurrect this genus to apply to *C. callistia.* We did not include the type of the genus, *Photogenis photogenis* (*Notropis photogenis*), in our analysis, and *Photogenis* was recognized with a mix of species that are currently in *Notropis* and *Cyprinella.* With no name available for the species, we refer to it as '*Cyprinella*' *callistia* until

such time that a broader analysis can be completed. This name will reflect that this species is clearly divergent from other *Cyprinella*, both morphologically (Mayden 1989) and genetically (Schonhuth and Mayden, 2010; this study).

Alburnops

Gidmark and Simons (2014) resurrected *Alburnops* based on the monophyly recovered by Mayden et al. (2006). We do not recover monophyly of the species placed in *Alburnops*, however, and instead find primarily two non-sister clades. The type species, *Alburnops blennius*, is recovered in a clade with *A. chalybaeus*, *A. petersoni*, *A. baileyi*, *A. xaenocephalus*, and *A. texanus*, and thus these should retain the genus name. The other clade is comprised of *A. chrosomus*, *A. rubricroceus*, *A. chiliticus*, *A. chlorocephalus*, and *A. lutipinnis*, and is more closely related to species currently recognized under *Notropis*, '*Notropis*', and *Miniellus*. Cashner et al. (2011) recognized these five species as the only members of the subgenus *Hydrophlox* and our results confirm this finding. We resurrect *Hydrophlox* as a genus to represent this clade.

Miniellus

Miniellus is currently recognized as containing four species: *Miniellus procne* (type species), *M. heterodon*, *M. stramineus*, and *M. topeka*. We did not include *M. procne* in our analysis, but did include the similar *M. stramineus*, which was not sister to the other species of *Miniellus* we included, *M. heterodon*. Several '*Notropis*' species were found to be more closely related to *Miniellus* species than they are to each other. Many of these '*Notropis*' were considered

by Mayden et al. (2006) as belonging to a '*Notropis*' *longirostris* clade. Given strong support for the monophyly of *Miniellus,* the '*N.*' *longirostris* clade, and these other species of '*Notropis*', we extend the genus *Miniellus* to include '*Notropis*' *greenei*, '*N.*' *scabriceps*, '*N.*' *sabinae*, '*N.*' *longirostris*, '*N.*' *ammophilus*, '*N.*' *chihuahua*, '*N.*' *melanostomus*, and '*N.*' *nubilus*, and, although not included in our analysis, additionally '*N.*' *rafinesquei*, based on its original description (Suttkus 1991) and its strongly supported position as part of the '*Notropis*' *longirostris* clade (Suttkus and Boschung, 1990). These species are all ventrally flattened, benthic fishes that prefer sand substrates.

'Notropis'

Besides the species listed above that we now consider under *Miniellus*, several other '*Notropis*' are found throughout our phylogeny. '*Notropis*' *scepticus* is found sister to *Hudsonius* in the concatenated analysis but its placement remains unresolved in the species tree. The position of '*N.*' *scepticus* varies in different studies, and likely the best solution would be to describe a separate genus for the species. We retain it under '*Notropis*' for the time being. We recover another '*Notropis*' clade sister to (*Ericymba* + '*Notropis*' *dorsalis*) + *Luxilus* composed of '*Notropis*' *heterolepis*, '*N.*' *bifrenatus*, and '*N.*' *atrocaudalis*. Jordan (1878) described *Chriope* for '*N.*' *bifrenatus.* Because *Chriope* is feminine, '*N.*' *bifrenatus* would be recognized as *C. bifrenata* while the other two species are nouns in apposition, and would not be changed.

Interestingly, we do not recover the '*Notropis*' *dorsalis* group (Mayden, 1989; Raley et al., 2001) as monophyletic. This group was composed of '*Notropis*' *dorsalis*, '*N.*' *ammophilus*, '*N.*' *longirostris*, '*N.*' *rafinesquei*, and '*N.*' *sabinae.* Instead we find '*N.*' *ammophilus*, '*N.*'

longirostris, and '*N.*' *sabinae* to group with *Miniellus* (see above) and '*N.*' *dorsalis* as sister to *Ericymba*. Currently, *Ericymba* is diagnosed by the presence of enlarged infraorbital canal scales (Pera and Armbruster, 2001), which are not found in '*N.*' *dorsalis;* however, '*N.*' *dorsalis* is otherwise very similar in morphology to the species of *Ericymba,* having a large mouth and ventrally flattened body. We propose to include '*N.*' *dorsalis* in *Ericymba*.

Notropis

We find two distinct and distant clades that include species regarded as true *Notropis* (Mayden et al., 2006; Gidmark and Simons, 2014). The type species, *Notropis atherinoides,* is found in a clade that is sister to *Lythrurus* and contains *N. jemezanus*, *N. amabilis*, *N. micropteryx*, *N. rubellus*, and *N. amoenus*, and this clade should retain the genus name *Notropis.* The other clade includes *N. buchanani*, *N. wickliffi*, *N. volucellus*, and *N. spectrunculus*, and this clade forms a polytomy with the *Hydrophlox* clade and the *Miniellus* clade. These species were either originally described as *Notropis*, or have been moved to *Notropis* from *Alburnops*, *Hybognathus*, or *Hybopsis*. *Notropis leucidous* is a very similar species considered to be closely related to this clade (Simons et al., 2003), and it is the type species of *Paranotropis* Fowler 1904, and we refer these species to *Paranotropis.*

Intraspecies utility of FishLife markers

This study included two specimens of three species: *Notropis atherinoides, Ericymba amplamala,* and *Pimephales vigilax*. *Notropis atherinoides* specimens, one from Wisconsin and the other from Arkansas, exhibited 99.6% sequence similarity with a pairwise distance of 0.003

and a total of 1,086 nucleotide differences across the entire 286,455 bp alignment. The specimens of *E. amplamala* were from Alabama and Mississippi, populations that were not found to be morphologically distinguishable in a detailed analysis (Pera and Armbruster, 2001), and had 99.5% sequence similarity, a pairwise distance of 0.005, and 1,424 differences. Our samples of *Pimephales vigilax* were collected from Paint Rock River in Tennessee and the Uphapee River in Alabama and had 99.7% sequence similarity, a pairwise distance of 0.002, and 845 nucleotide differences. These results suggest two things: there may be cryptic diversity within shiner clade species, and the FishLife Markers are likely of utility at the the population level, despite their initial development for use across a very broad taxonomic scale (Arcila et al., 2017).

One of the targeted sequences was COI, a popular mitochondrial marker that is often used to delineate fish species. We find a wide range of infraspecific differences in the 703 bp of the partial COI sequences examined. *Notropis atherinoides*, despite disparate collection sites, has only a 2 bp difference (0.4% divergence). *Pimephales vigilax* from the neighboring Tennessee and Mobile River systems had a 16 bp difference (2.3% divergence). *Ericymba amplamala,* however, had a 54 bp difference (7.7% divergence), a degree of difference often associated with species-level differentiation, and there needs to be further investigation into the genetic structure of the species. COI alone may be suitable for identification of cryptic diversity for shiners, but the full phylogenomic dataset adds a considerable number of characters for elucidating population structure.

CONCLUSIONS

This study provides an important first step in using phylogenomics to resolve the problematic shiner clade. By employing a publicly available probe set (Arcila et al., 2017), future research can include more specimens that were not sampled in this study and easily be combined with our dataset. Our phylogenies help in understanding why this group has been difficult to resolve and requires a phylogenomic approach. Not only has the group been described as "morphologically conserved" (Gidmark and Simons, 2014), thus hampering morphological interpretations of relationships, but we would argue that the same is true genetically. Single or sub-ten locus phylogenies would most likely never be able to provide robust resolution when we find over 88% similarity (or uninformativeness) in a dataset comprised of over 288,000 base pairs. Problems with elucidating shiner relationships have been exacerbated by studies focusing only on subsets of the shiner clade due to sampling or cost restrictions. We demonstrate the utility of the exon capture method of Arcila et al., (2017) to elucidate relationships of rapidly evolving clades, and demonstrate that the markers may be of use at the population level as well. With the continuing decrease in cost of phylogenomic methods, the demonstrable utility of the FishLife markers at many phylogenetic levels, and the soon to be large number of fish taxa sampled using the FishLife markers, we would encourage researchers to add to this dataset. Numerous issues remain in the taxonomy and systematics of North American leuciscids, and we will continue to add species to the analysis. This study continues the trend at subtending the shiner clade into genera, but several important clades still need to be resolved and described.

References

Arcila D., Orti G., Vari R., armbruster J.W., Stiassny M.L.J., Ko K.D., Sabaj M.H., Lundberg J., Revell L.J., Betancur-R R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. :1–10.

Betancur-R R., Broughton R.E., Wiley E.O., Carpenter K., López J.A., Li C., Holcroft N.I., Arcila D., Sanciangco M., Cureton J.C. II, Zhang F., Buser T., Campbell M.A., Ballesteros J.A., Roa-Varon A., Willis S., Borden W.C., Rowley T., Reneau P.C., Hough D.J., Lu G., Grande T., Arratia G., Orti G. 2013. The tree of life and a new classification of bony fishes. PLoS Curr. 5. Edition 1. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Bird, N.C., Hernandez, L.P., 2007. Morphological variation in the Weberian apparatus of Cypriniformes. J. Morphol. 268, 739–757. doi:10.1002/jmor.10550

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Bond J.E., Garrison N.L., Hamilton C.A., Godwin R.L., Hedin M., Agnarsson I. 2014. Phylogenomics Resolves a Spider Backbone Phylogeny and Rejects a Prevailing Paradigm for Orb Web Evolution. Current Biology. 24:1765–1771.

Bufalino, A.P., Mayden, R.L., 2010. Phylogenetic evaluation of North American Leuciscidae (Actinopterygii: Cypriniformes: Cyprinoidea) as inferred from analyses of mitochondrial and nuclear DNA sequences. Systematics and Biodiversity 8, 493–505. doi: 10.1080/14772000.2010.516029

Buth D.G. 1979. Biochemical systematics of the cyprinid genus *Notropis*—I. The subgenus *Luxilus*. Biochemical Systematics and Ecology. 7:69–79.

Cashner, M.F., Piller, K.R., Bart, H.L., 2011. Phylogenetic relationships of the North American cyprinid subgenus *Hydrophlox*. Molecular Phylogenetics and Evolution 59, 725–735. doi: 10.1016/j.ympev.2011.03.019

Chen, W.J., Mayden, R.L., 2009. Molecular systematics of the Cyprinoidea (Teleostei: Cypriniformes), the world's largest clade of freshwater fishes: Further evidence from six nuclear genes. Molecular Phylogenetics and Evolution 52, 544–549. doi:10.1016/j.ympev.2009.01.006

Coburn M.M., Cavender T.M. 1992. Interrelationships of North American Cyprinid Fishes. In: Mayden R.L., editor. Systematics, Historical Ecology, and North American Freshwater Fishes. Systematics.

Dowling T.E., Naylor G.J.P. 1997. Evolutionary Relationships of Minnows in the Genus *Luxilus* (Teleostei: Cyprinidae) as Determined from Cytochrome b Sequences. Copeia. 1997:758–765.

Eschmeyer, W.N., Fong, J.D., n.d. Catalog of Fishes [WWW Document]. researcharchive.calacademy.org. URL http://researcharchive.calacademy.org/research/ichthyology/catalog/speciesbyfamily.asp (accessed 2.17.17).

Eytan R.I., Evans B.R., Dornburg A., Lemmon A.R., Lemmon E.M., Wainwright P.C., Near T.J. 2015. Are 100 enough? Inferring acanthomorphteleost phylogeny using Anchored HybridEnrichment. BMC Evolutionary Biology. 15:113.

Fang, F., Norén, M., Liao, T.Y., Källersjö, M., Kullander, S.O., 2009. Molecular phylogenetic interrelationships of the south Asian cyprinid genera *Danio*, *Devario* and *Microrasbora* (Teleostei, Cyprinidae, Danioninae). Zoologica Scripta 38, 237–256. doi:10.1111/j.1463-6409.2008.00373.x

Geneious (v.6) created by Biomatters. Available from http://www.geneious.com.

Gaubert, P., Denys, G., Oberdorff, T., 2009. Genus-level supertree of Cyprinidae (Actinopterygii: Cypriniformes), partitioned qualitative clade support and test of macro-evolutionary scenarios. Biological Reviews 84, 653–689. doi:10.1111/j.1469-185X.2009.00091.x

Gidmark N.J., Simons A.M. 2014. Cyprinidae: Carps and Minnows. In: Warren M.L. Jr, M B.B., editors. Freshwater Fishes of North America. Baltimore: Johns Hopkins University Press. p. 334–450.

Gilbert C.R. 1964. The American Cyprinid Fishes of the Sub-genus *Luxilus* (genus *Notropis*). Bulletin of the Florida State Museum.

Hamilton C.A., Lemmon A.R., Lemmon E.M., Bond J.E. 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evolutionary Biology.:1–20.

Hollingsworth P.R., Simons A.M., Fordyce J.A., Hulsey C.D. 2013. Explosive diversification following a benthic to pelagic shift in freshwater fishes. BMC Evolutionary Biology. 13:272.

Houston, D.D., Shiozawa, D.K., Riddle, B.R., 2010. Phylogenetic relationships of the western North American cyprinid genus *Richardsonius*, with an overview of phylogeographic structure. Molecular Phylogenetics and Evolution 55, 259–273. doi:10.1016/j.ympev.2009.10.017

Imoto, J.M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y.P., Miya, M., Nishida, M., Hanzawa, N., 2013. Phylogeny and biogeography of highly diverged freshwater fish

species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. Gene 514, 112–124. doi:10.1016/j.gene.2012.10.019

Jordan D.S. 1885. A Catalogue of the Fishes Known to Inhabit the Waters of North America, North of the Tropic of Cancer, With Notes on the Species Discovered in 1883 and 1884. Washington: US Government Printing Office.

Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. Systematic Biology. 61:727–744.

Mayden R.L. 1989. Phylogenetic studies of North American minnows, with emphasis on the genus *Cyprinella* (Teleostei, Cypriniformes).

Mayden R.L., Simons A.M., Wood R.M., Harris P.M., Kuhajda B.R. 2006. Molecular systematics and classification of north american notropin shiners and minnows. p. 1–30.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–52. doi:10.1093/bioinformatics/btv234

Pera, T.P., Armbruster, J.W., 2006. A new species of *Notropis* (Cypriniformes: Cyprinidae) from the southeastern United States. Copeia 2006, 423–430.

Pramuk J.B., Grose M.J., Clarke A.L., Greenbaum E., Bonaccorso E., Guayasamin J.M., Smith-Pardo A.H., Benz B.W., Harris B.R., Siegfreid E., Reid Y.R., Holcroft-Benson N., Wiley E.O. 2007. Phylogeny of finescale shiners of the genus *Lythrurus* (Cypriniformes: Cyprinidae) inferred from four mitochondrial genes. Molecular Phylogenetics and Evolution. 42:287–297.

Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 526:569–573.

Raley M.E., Wood R.M., McEachran J.D. 2001. Molecular systematics of members of the *Notropis dorsalis* species group (Actinopterygii: Cyprinidae). Copeia. 2001:638–645.

Rüber, L., Kottelat, M., Tan, H., Ng, P.K., Britz, R., 2007. Evolution of miniaturization and the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate. BMC Evolutionary Biology 7, 38. doi:10.1186/1471-2148-7-38

Schmidt T.R., Bielawski J.P., Gold J.R. 1998. Molecular Phylogenetics and Evolution of the Cytochrome b Gene in the Cyprinid Genus *Lythrurus* (Actinopterygii: Cypriniformes). Copeia. 1998:14–22.

Schönhuth S., Doadrio I., Domínguez-Domínguez O., Hillis D.M., Mayden R.L. 2008. Molecular evolution of southern North American Cyprinidae (Actinopterygii), with the description of the new genus *Tampichthys* from central Mexico. Molecular Phylogenetics and Evolution. 47:729–756.

Schönhuth S., Mayden R.L. 2010. Phylogenetic relationships in the genus *Cyprinella* (Actinopterygii: Cyprinidae) based on mitochondrial and nuclear gene sequences. Molecular Phylogenetics and Evolution. 55:77–98.

Scott, C.H., Cashner, M., Grossman, G.D., Wares, J.P., 2009. An awkward introduction: phylogeography of *Notropis lutipinnis* in its "native" range and the Little Tennessee River. Ecology of Freshwater Fish 18, 538–549. doi:10.1111/j.1600-0633.2009.00368.x

Shimodaira, H., 2002. An Approximately Unbiased Test of Phylogenetic Tree Selection. Systematic Biology 51, 492–508. doi:10.1080/10635150290069913

Simons, A.M., Berendzen, P.B., Mayden, R.L., 2003. Molecular systematics of North American phoxinin genera (Actinopterygii: Cyprinidae) inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Zoological Journal of the Linnean Society 139, 63–80.

Snelson F.F. 1972. Systematics of the subgenus *Lythrurus*, genus Notropis (Pisces: Cyprinidae).

Suttkus, R.D., 1991. *Notropis rafinesquei*, a new cyprinid fish from the Yazoo River system in Mississippi. Bulletin of the Alabama Museum of Natural History. 10:1-9

Suttkus, R.D., Boschung, H.T., 1990. *Notropis ammophilus*, a new cyprinid fish from Southeastern United States. Tulane Stud. Zool 27, 49–63.

Wang, C., Wang, J., Yang, J., Lu, G., Song, X., Chen, Q., Xu, J., Yang, Q., Li, S., 2012. Complete mitogenome sequence of black carp (*Mylopharyngodon piceus*) and its use for molecular phylogeny of leuciscine fishes. Mol Biol Rep 39, 6337–6342. doi:10.1007/s11033-012-1455-9

Zhang, X., Yue, B., Jiang, W., Song, Z., 2008. The complete mitochondrial genome of rock carp *Procypris rabaudi* (Cypriniformes: Cyprinidae) and phylogenetic implications. Mol Biol Rep 36, 981–991. doi:10.1007/s11033-008-9271-y

Table 3.1. Tissues used in this study. Taxonomy follows Simons and Gidmark (2014). AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; SLUM = St. Louis University Museum; SELU = Southeastern Louisiana University. Type species for each genus are indicated with an asterisk.

Species	Tissue Voucher
Outgroup	
Notemigonus crysoleucas	AUFT 6632
Chrosomus eos	AUFT 6624
Phoxinus phoxinus	SLUM B.91.T187
Semotilus atromaculatus	AUFT 5949
Ingroup	
***Alburnops* Girard 1856** (20 spp.)	
Alburnops baileyi	SELU 87
*Alburnops blennius**	SELU 451
Alburnops chalybaeus	SLUM 2046
Alburnops chiliticus	AUFT 1726
Alburnops chlorocephalus	SLUM 2054
Alburnops chrosomus	SLUM 2056
Alburnops lutipinnis	SELU 492
Alburnops petersoni	SELU 945
Alburnops potteri	SLUM 2203
Alburnops rubricroceus	SLUM 2243
Alburnops texanus	AUFT 0579
Alburnops xaenocephalus	SLUM 2310
***Codoma* Girard 1856** (1 sp.)	
*Codoma ornata**	CBD09-26.03
***Cyprinella* Girard 1856** (30 spp.)	
Cyprinella analostana	SLUM 1771
Cyprinella caerulea	SLUM 13330.01
Cyprinella callisema	SLUM 1797
Cyprinella callistia	SLUM RLM5462
Cyprinella camura	SLUM 333.01
Cyprinella chloristia	SLUM 1802
Cyprinella galactura	SLUM RLM5936
Cyprinella gibbsi	AUFT 5938
*Cyprinella lutrensis**	SLUM RLM3451
Cyprinella nivea	SLUM 1821
Cyprinella pyrrhomelas	SLUM 1828
Cyprinella trichroistia	AUFT 5933
Cyprinella venusta	AUFT 5922
Cyprinella whipplei	SLUM 211.02
Cyprinella xaenura	SLUM 1850

Table 3.1 (*continued*). Tissues used in this study. Taxonomy follows Simons and Gidmark (2014). AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; SLUM = St. Louis University Museum; SELU = Southeastern Louisiana University. Type species for each genus are indicated with an asterisk.

Species	Tissue Voucher
***Ericymba* Cope 1865** (2 spp., type = *E. buccata*)	
Ericymba amplamala	AUFT 1085
Ericymba amplamala	SLUM 675.04
***Erimystax* Jordan 1882** (5 spp., type = *E. dissimilis*)	
Erimystax insignis	AUFT 6631
***Graodus* Günther 1868** (3 spp., type = *G. boucardi*)	
Graodus moralesi	SELU 4688
***Hudsonius* Girard 1856** (3 spp.)	
Hudsonius altipinnis	SLUM 1968
Hudsonius cummingsae	SLUM 2060
*Hudsonius hudsonius**	SELU 917
***Hybognathus* Agassiz 1855** (7 spp., type = *H. nuchalis*)	
Hybognathus hankinsoni	AUFT 6625
Hybopsis hypsinotus	AUFT 1786
Hybopsis lineapunctata	AUFT 5999
Hybopsis winchelli	AUFT 1077
***Luxilus* Rafinesque 1820** (9 spp.)	
Luxilus albeolus	SLUM 1901
*Luxilus chrysocephalus**	AUFT 5962
Luxilus coccogenis	AUFT 1691
Luxilus cornutus	AUFT 6122
Luxilus pilsbryi	SLUM RLM10208
Luxilus zonatus	SLUM RLM4106
Luxilus zonistius	AUFT 1140
***Lythrurus* Jordan 1876** (11 spp., type = *L. umbratilis*)	
Lythrurus ardens	SLUM 1915
Lythrurus atrapiculus	AUFT 1079
Lythrurus bellus	AUFT 0647
Lythrurus fasciolaris	SLUM 1922
Lythrurus fumeus	SLUM 1928
Lythrurus lirus	SLUM 1930
Lythrurus roseipinnis	AUFT 0568
***Miniellus* Jordan 1888** (4 spp., type = *M. procne*)	
Miniellus heterodon	SELU 991
Miniellus stramineus	AUFT 0061

Table 3.1 (*continued*). Tissues used in this study. Taxonomy follows Simons and Gidmark (2014). AUFT = Auburn University Fish Tissue Collection; UAIC =University of Alabama Ichthyological Collection; SLUM = St. Louis University Museum; SELU = Southeastern Louisiana University. Type species for each genus are indicated with an asterisk.

Species	Tissue Voucher
***Notropis* Rafinesque 1818** (21 spp.)	
Notropis amabilis	SLUM NAFF400
Notropis amoenus	SLUM 2003
*Notropis atherinoides**	SELU 1298
*Notropis atherinoides**	SLUM 2018
Notropis buchanani	SLUM RLM3440
Notropis jemezanus	UAIC 13508.03
Notropis micropteryx	SLUM 617.03
Notropis rubellus	SELU 1034
Notropis spectrunculus	AUFT 1693
Notropis volucellus	SELU 218
Notropis wickliffi	SLUM 2308
Incertae sedis	
'Notropis' ammophilus	AUFT 0014
'Notropis' atrocaudalis	SLUM 2023
'Notropis' bifrenatus	SLUM 2027
'Notropis' chihuahua	SLUM 5085
'Notropis' dorsalis	AUFT 0125
'Notropis' greenei	SLUM RLM4330
'Notropis' heterolepis	SELU 1000
'Notropis' longirostris	AUFT 0048
'Notropis' melanostomus	UAIC 12075.01
'Notropis' nubilus	SLUM RLM10074
'Notropis' sabinae	SLUM RLM6455
'Notropis' scabriceps	AUFT 1660
'Notropis' scepticus	SELU 153
***Opsopoeodus* Hay 1881** (1 sp.)	
*Opsopoeodus emiliae**	AUFT 0638
***Pimephales* Rafinesque 1820** (4 spp., type = *P. promelas*)	
Pimephales vigilax	AUFT 6630
Pimephales vigilax	AUFT 5925
***Pteronotropis* Fowler 1935** (10 spp.)	
Pteronotropis euryzonas	AUFT 1136
Pteronotropis grandipinnis	AUFT 1123
Pteronotropis harperi	AUFT 0585
Pteronotropis hubbsi	SLUM 4379
*Pteronotropis hypselopterus**	SLUM RLM5732
Pteronotropis merlini	AUFT 1855
Pteronotropis signipinnis	AUFT 1861

Table 3.2. Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Agosia chrysogaster		Girard, 1856	
Algansea aphanea		Barbour & Miller, 1978	
Algansea avia		Barbour & Miller, 1978	
Algansea barbata		Álvarez & Cortés, 1964	
Algansea lacustris		Steindachner, 1895	
Algansea monticola		Barbour & Contreras-Balderas, 1968	
Algansea popoche		(Jordan & Snyder, 1899)	
Algansea tincella		(Valenciennes, 1844)	
Aztecula sallaei		(Günther, 1869	
Codoma ornata		Girard, 1856	✔
Cyprinella alvarezdelvillari		Contreras-Balderas & Lozano-Vilano, 1994	
Cyprinella analostana		Girard, 1859	✔
Cyprinella bocagrande		(Chernoff & Miller, 1982)	
Cyprinella caerulea		(Jordan, 1877)	✔
Cyprinella callisema		(Jordan, 1877)	✔
Cyprinella callistia	*'Cyprinella' callistia*	(Jordan, 1877)	✔
Cyprinella callitaenia		(Bailey & Gibbs, 1956)	
Cyprinella camura		(Jordan & Meek, 1884)	✔
Cyprinella chloristia		(Jordan & Brayton, 1878)	✔
Cyprinella eurystoma		(Jordan, 1877)	
Cyprinella formosa		(Girard, 1856)	
Cyprinella galactura		(Cope, 1868)	✔
Cyprinella garmani		(Jordan, 1885)	
Cyprinella gibbsi		(Howell & Williams, 1971)	✔
Cyprinella labrosa		(Cope, 1870)	
Cyprinella leedsi		(Fowler, 1942)	
Cyprinella lepida		Girard, 1856	
Cyprinella lutrensis		(Baird & Girard, 1853)	✔
Cyprinella monacha		(Cope, 1868)	

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Cyprinella monacha		(Cope, 1868)	
Cyprinella nivea		(Cope, 1870)	✔
Cyprinella panarcys		(Hubbs & Miller, 1978)	
Cyprinella proserpina		(Girard, 1856)	
Cyprinella pyrrhomelas		(Cope, 1870)	✔
Cyprinella rutila		(Girard, 1856)	
Cyprinella spiloptera		(Cope, 1867)	
Cyprinella stigmatura		(Jordan, 1877)	
Cyprinella trichroistia		(Jordan & Gilbert, 1878)	✔
Cyprinella venusta		Girard, 1856	✔
Cyprinella whipplei		Girard, 1856	✔
Cyprinella xaenura		(Jordan, 1877)	✔
Cyprinella xanthicara		(Minckley & Lytle, 1969)	
Cyprinella zanema		(Jordan & Brayton, 1878)	
Dionda argentosa		Girard, 1856	
Dionda diaboli		Hubbs & Brown, 1957	
Dionda episcopa		Girard, 1856	
Dionda melanops		Girard, 1856	
Dionda nigrotaeniata		(Cope, 1880)	
Dionda serena		Girard, 1856	
Ericymba amplamala		(Pera & Armbruster, 2006)	✔
Ericymba buccata		Cope, 1865	
Erimonax monachus		(Cope, 1868)	
Erimystax cahni		Hubbs & Crowe, 1956	
Erimystax dissimilis		(Kirtland, 1840)	
Erimystax harryi		(Hubbs & Crowe, 1956)	
Erimystax insignis		(Hubbs & Crowe, 1956)	✔
Erimystax x-punctatus		(Hubbs & Crowe, 1956)	
Hybognathus amarus		(Girard, 1856)	
Hybognathus argyritis		Girard, 1856	

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Hybognathus hankinsoni		Hubbs, 1929	✔
Hybognathus hayi		Jordan, 1885	
Hybognathus nuchalis		Agassiz, 1855	
Hybognathus placitus		Girard, 1856	
Hybognathus regius		Girard, 1856	
Hybopsis amblops		(Rafinesque, 1820)	
Hybopsis amnis		(Hubbs & Greene, 1951)	
Hybopsis hypsinotus		(Cope, 1870)	✔
Hybopsis lineapunctata		Clemmer & Suttkus, 1971	✔
Hybopsis rubrifrons		(Jordan, 1877)	
Hybopsis winchelli		Girard, 1856	✔
Luxilus albeolus		(Jordan, 1889)	✔
Luxilus cardinalis		(Mayden, 1988)	
Luxilus cerasinus		(Cope, 1868)	
Luxilus chrysocephalus		Rafinesque, 1820	✔
Luxilus cornutus		(Mitchill, 1817)	✔
Luxilus pilsbryi		(Fowler, 1904)	✔
Luxilus zonatus		(Putnam, 1863)	✔
Lythrurus alegnotus		(Snelson, 1972)	
Lythrurus ardens		(Cope, 1868)	✔
Lythrurus atrapiculus		(Snelson, 1972)	✔
Lythrurus bellus		(Hay, 1881)	✔
Lythrurus fasciolaris		(Gilbert, 1891)	✔
Lythrurus fumeus		(Evermann, 1892)	✔
Lythrurus lirus		(Jordan, 1877)	✔
Lythrurus matutinus		(Cope, 1870)	
Lythrurus roseipinnis		(Hay, 1885)	✔
Lythrurus snelsoni		(Robison, 1985)	
Lythrurus umbratilis		(Girard, 1856)	
Luxilus coccogenis	*Coccotis coccogenis* (type sp.)	(Cope, 1868)	✔

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Luxilus zonistius	*Coccotis zonistius*	Jordan, 1880	✔
Notropis asperifrons		Suttkus & Raney, 1955	
Notropis baileyi		Suttkus & Raney, 1955	✔
Notropis bairdi		Hubbs & Ortenburger, 1929	
Notropis blennius		(Girard, 1856)	✔
Notropis buccula		Cross, 1953	
Notropis candidus		Suttkus, 1980	
Notropis chalybaeus		(Cope, 1867)	✔
Notropis edwardraneyi		Suttkus & Clemmer, 1968	
Notropis hypsilepis		Suttkus & Raney, 1955	
Notropis petersoni		Fowler, 1942	✔
Notropis potteri		Hubbs & Bonham, 1951	✔
Notropis shumardi		Girard, 1856	
Notropis texanus		(Girard, 1856)	✔
Notropis xaenocephalus		(Jordan, 1877)	✔
Notropis calientis		Jordan & Snyder, 1899	
Notropis boucardi		(Günther, 1868)	
Notropis cumingii		(Günther, 1868)	
Notropis moralesi		de Buen, 1955	✔
Notropis altipinnis		(Cope, 1870)	✔
Notropis hudsonius		(Clinton, 1824)	✔
Notropis heterodon		(Cope, 1865)	✔
Notropis procne		(Cope, 1865)	
Notropis stramineus		(Cope, 1865)	✔
Notropis topeka		(Gilbert, 1884)	
Notropis amabilis		(Girard, 1856)	✔
Notropis amoenus		(Abbott, 1874)	✔
Notropis ariommus		(Cope, 1867)	
Notropis atherinoides		Rafinesque, 1818	✔
Notropis cahabae		Mayden & Kuhajda, 1989	

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Notropis girardi		Hubbs & Ortenburger, 1929	
Notropis jemezanus		(Cope, 1875)	✔
Notropis micropteryx		(Cope, 1868)	✔
Notropis ozarcanus		Meek, 1891	
Notropis percobromus		(Cope, 1871)	
Notropis perpallidus		Hubbs & Black, 1940	
Notropis rubellus		(Agassiz, 1850)	✔
Notropis stilbius		Jordan, 1877	
Notropis suttkusi		Humphries & Cashner, 1994	
Notropis oxyrhynchus		Hubbs & Bonham, 1951	
Notropis chiliticus	*Hydrophlox chiliticus* (type sp.)	(Cope, 1870)	✔
Notropis chlorocephalus	*Hydrophlox chlorocephalus*	(Cope, 1870)	✔
Notropis chrosomus	*Hydrophlox chrosomus*	(Jordan, 1877)	✔
Notropis lutipinnis	*Hydrophlox lutipinnis*	(Jordan & Brayton, 1878)	✔
Notropis rubricroceus	*Hydrophlox rubricroceus*	(Cope, 1868)	✔
Notropis leuciodus	*Paranotropis leuciodus* (type sp.)	(Cope, 1868)	
Notropis buchanani	*Paranotropis buchanani*	Meek, 1896	✔
Notropis spectrunculus	*Paranotropis spectrunculus*	(Cope, 1868)	✔
Notropis volucellus	*Paranotropis volucellus*	(Cope, 1865)	✔
Notropis wickliffi	*Paranotropis wickliffi*	Trautman, 1931	✔
Notropis cummingsae	*Pteronotropis cummingsae*	Myers, 1925	✔
Notropis aguirrepequenoi		Contreras-Balderas & Rivera-Teillery, 1973	
Notropis albizonatus		Warren & Burr, 1994	
Notropis alborus		Hubbs & Raney, 1947	
Notropis amecae		Pérez-Rodríguez, Pérez-Ponce de León, Domínguez-Domínguez & Doadrio, 2009	
Notropis anogenus		Forbes, 1885	
Notropis aulidion		Chernoff & Miller, 1986	
Notropis boops		Gilbert, 1884	
Notropis braytoni		Jordan & Evermann, 1896	

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Notropis dorsalis	*Ericymba dorsalis*	(Agassiz, 1854)	✔
Notropis maculatus		(Hay, 1881)	
Notropis mekistocholas		Snelson, 1971	
Notropis nazas		Meek, 1904	
Notropis orca		Woolman, 1894	
Notropis ortenburgeri		Hubbs, 1927	
Notropis photogenis		(Cope, 1865)	
Notropis rafinesquei		Suttkus, 1991	
Notropis rupestris		Page, 1987	
Notropis saladonis		Hubbs & Hubbs, 1958	
Notropis scepticus	New genus description req'd	(Jordan & Gilbert, 1883)	✔
Notropis semperasper		Gilbert, 1961	
Notropis simus		(Cope, 1875)	
Notropis telescopus		(Cope, 1868)	
Notropis tropicus		Hubbs & Miller, 1975	
Notropis uranoscopus		Suttkus, 1959	
Notropis heterolepis	*Chriope heterolepis*	Eigenmann & Eigenmann, 1893	✔
Notropis atrocaudalis	*Chriope atrocaudalis*	Evermann, 1892	✔
Notropis bifrenatus	*Chriope bifrenata*	(Cope, 1867)	✔
Notropis chihuahua	*Miniellus chihuahua*	Woolman, 1892	✔
Notropis greenei	*Miniellus greenei*	Hubbs & Ortenburger, 1929	✔
Notropis longirostris	*Miniellus longirostris*	(Hay, 1881)	✔
Notropis melanostomus	*Miniellus melanostomus*	Bortone, 1989	✔
Notropis nubilus	*Miniellus nubilus*	(Forbes, 1878)	✔
Notropis sabinae	*Miniellus sabinae*	Jordan & Gilbert, 1886	✔
Notropis scabriceps	*Miniellus scabriceps*	(Cope, 1868)	✔
Notropis ammophilus	*Miniellus ammophilus*	Suttkus & Boschung, 1990	✔
Notropis calabazas		Lyons & Mercado-Silva, 2004	
Notropis grandis		Domínguez-Domínguez, Pérez-Rodríguez, Escalera-Vázquez & Doadrio, 2009	

Table 3.2 (*continued*). Current taxonomy of species in the shiner clade and proposed revisions.

Current Taxonomy	Proposed taxonomic changes (this study)	Author	Included in this study
Notropis imeldae		Cortés, 1968	
Notropis marhabatiensis		Domínguez-Domínguez, Pérez-Rodríguez, Escalera-Vázquez & Doadrio, 2009	
Opsopoeodus emiliae		Hay, 1881	✔
Pimephales notatus		(Rafinesque, 1820)	
Pimephales promelas		Rafinesque, 1820	
Pimephales tenellus		(Girard, 1856)	
Pimephales vigilax		(Baird & Girard, 1853)	✔
Pteronotropis euryzonus		(Suttkus, 1955)	✔
Pteronotropis grandipinnis		(Jordan, 1877)	✔
Pteronotropis harperi		(Fowler, 1941)	✔
Pteronotropis hubbsi		(Bailey & Robison, 1978)	✔
Pteronotropis hypselopterus		(Günther, 1868)	✔
Pteronotropis merlini		(Suttkus & Mettee, 2001)	✔
Pteronotropis metallicus		(Jordan & Meek, 1884)	
Pteronotropis signipinnis		(Bailey & Suttkus, 1952)	✔
Pteronotropis stonei		(Fowler, 1921)	
Pteronotropis welaka		(Evermann & Kendall, 1898)	
Tampichthys catostomops		(Hubbs & Miller, 1977)	
Tampichthys dichromus		(Hubbs & Miller, 1977)	
Tampichthys erimyzonops		(Hubbs & Miller, 1974)	
Tampichthys ipni		(Álvarez & Navarro, 1953)	
Tampichthys mandibularis		(Contreras-Balderas & Verduzco-Martínez, 1977)	
Tampichthys rasconis		(Jordan & Snyder, 1899)	
Yuriria alta		(Jordan, 1880)	
Yuriria amatlana		Domínguez-Domínguez, Pompa-Domínguez & Doadrio, 2007	
Yuriria chapalae		(Jordan & Snyder, 1899)	

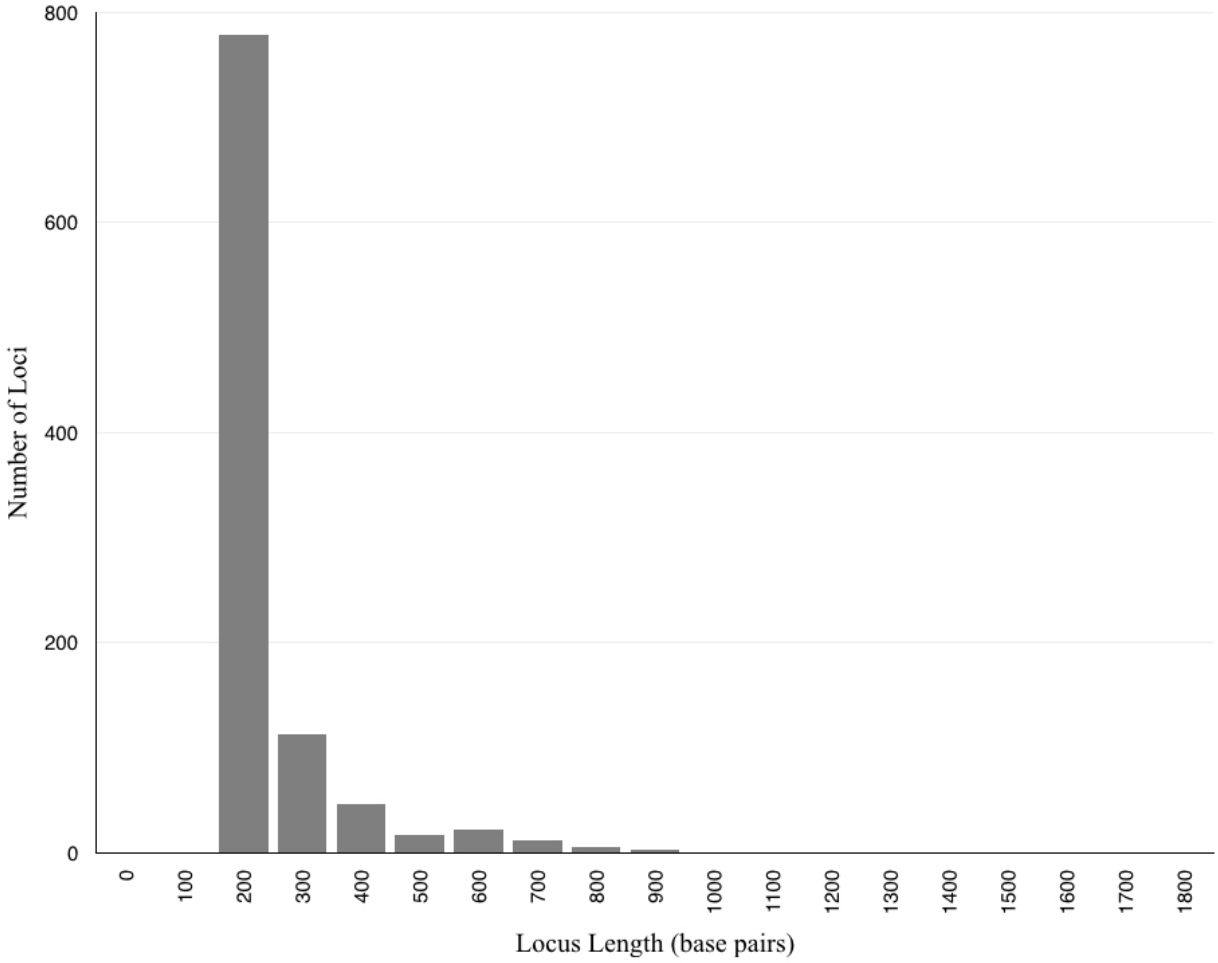

Figure 3.1. Histogram showing lengths of loci in base pairs.

Figure 3.2. ML tree based on concatenated alignment. Numbers at nodes represent bootstrap support, with nodes less than 70% supported collapsed. Scale bar represents number of substitutions per site.

Figure 3.3. Species tree using ASTRAL-II. Internal branch lengths are in coalescent units and branches that lead to tips are not calculated by ASTRAL II but instead arbitrarily displayed. Branch support values indicate the support for a quadipartition (instead of bipartitions).

CHAPTER FOUR

GENETIC DIFFERENTIATION WITHOUT MORPHOLOGICAL DIFFERENTIATION IN THE LONGNOSE SHINER (*MINIELLUS LONGIROSTRIS*)

INTRODUCTION

Miniellus longirostris, the longnose shiner, is currently recognized as a single species with a southeastern US range extending from the Mississippi River drainage in the west to the Apalachicola River drainage in the east with records of collections extending to isolated pockets of the upper Etowah and Flint rivers in Georgia (Bart et al., 1995; Suttkus and Boschung, 1990; Boschung and Mayden, 2004). The species has had quite a convoluted taxonomic history, originally being described in the genus *Alburnops* (Hay, 1881), then moved to *Notropis* (Gilbert, 1978), then recognized as *Hybopsis* (Boschung, 1992), until consensus settled on back on *Notropis* due to the use of *Notropis* as a repository for many shiner species of uncertain placement. Previous to the work completed in Chapter three, relationships among the shiners had been difficult to resolve and thus the taxonomy remained uninformed. Gidmark and Simons (2014) evaluated the history of the genus and followed relationships established by Mayden et al. (2006) to break up the group. Many species were still left with an uncertain placement and designated as '*Notropis'*, including *Miniellus longirostris*. Chapter three has revealed that this species belongs to a clade with members assigned by Gidmark and Simons (2014) to *Miniellus*, and the conclusions of Chapter three to expand the name to this entire clade are followed for this chapter.

The *Miniellus longirostris* species group consists of four species: *M. ammophilu*s, *M. longirostris*, *M. sabinae*, and *M. rafinesquei* (Suttkus and Boschung, 1990; Suttkus, 1991). The

species are all small and tan with their dorsal scales faintly outlined in black and they generally have yellow to orange fins. They are ventrally flattened and are found over sand or fine gravel close to the substrate in small- to medium-sized streams.

The recognition of *Miniellus ammophilus* (Suttkus and Boschung, 1990) essentially bisected the range of *M. longirostris* into areas east and west of the Mobile River basin; however, some localities for *M. longirostris* are known from the lower Mobile River. This distribution is unusual, and suggests potential bifurcation of the range of *M. longirostris* by the Mississippi embayment to the Gulf of Mexico. This makes *M. longirostris* a suitable species to examine for potential speciation across the gulf coast. Many studies have shown that despite morphological similarities, significant genetic divergence is present in a variety of southeastern freshwater fishes (April et al., 2011; Berendzen et al., 2008; Butler and Mayden, 2003; Schneider et al., 2012). We hypothesized that the wide distribution of *M. longirostris* makes this species a prime candidate to investigate the possibility for the presence of diversity. In this study we employ genetic techniques using the vertebrate bar-coding marker, partial cytochrome oxidase subunit I (COI) mitochondrial gene, to identify potential distinct lineages, as well as geometric morphometrics to help elucidate possible previously unrecognized shape distinctions across the drainages. While COI has been used successfully across a wide variety of taxa for evaluating cryptic diversity (for example King et al., 2008; Ståhls and Savolainen, 2008; Witt et al., 2006), the addition of geometric morphometric data for comparison can allow us to make inferences about the similarities or differences we may see in the genetic versus morphometric results.

METHODS

Molecular Analyses

A total of 192 samples of *Miniellus longirostris* were preserved in ethanol from 36 locations representing four geographical groupings, supplemented by 11 sequences acquired from GenBank (Table 4.1). These groupings are based on patterns reported for other southeastern taxa (see Swift et al., 1985; Wiley and Mayden, 1985; Bermingham and Avise, 1986; Soltis et al., 2006) and consist of the Mississippi drainage, Western drainage (between Mississippi and Alabama rivers), Eastern drainages (from the Alabama River through Choctawhatchee River), and the Apalachicola drainage (Figure 4.1A). Outgroup taxa for phyloegenetic tree reconstruction included four specimens of *Miniellus ammophilus* and two specimens of *M. rafinesquei*.

DNA was extracted using the Omegabiotek E.Z.N.A. animal tissue extraction kit (product #D3396-02) following manufacturer protocols. PCR primers and conditions follow Ivanova et al. (2007) to amplify a 648 bp region of the protein-coding mtDNA COI gene. Sequences were blasted, aligned, and checked for an open reading frame in Geneious v. 6.1.8 (http:// www.geneious.com). Haplotype networks were constructed using TCS v.1.2.1 (Clement et al., 2000). One reticulation in the network was broken according to rules established by (Crandall et al., 1994). Phylogenetic reconstruction was conducted using RAxML using the GTR+ Γ model and subjected to 1000 bootstrap replicates.

Geometric Morphometrics

A total of 171 formalin-preserved specimens representing various *Miniellus longirostris* collections throughout the recognized range were laterally photographed (Table 4.2). Eighteen

homologous landmarks were digitally placed on each photo using the software package tpsDIG2 (http://life.bio.sunysb.edu/morph/) according to the methods developed by Armbruster (2012; http://www.auburn.edu/~armbrjw/gmguide/Geometric_Morphometrics_Guide/Introduction.html). MorphoJ (http://www.flywings.org.uk/MorphoJ_page.htm) was used for a general procrustes analysis (GPA) that aligns, resizes, and removes slight curvature in the specimens. It also generates a consensus with a spread of points and performs principal components analysis (PCA) for examination of shape-space groupings (to compare with geographical distributions and genetic results) and canonical variates analysis (CVA) for visualization of morphological features that can distinguish *a priori* groupings (in this case drainage groupings).

RESULTS

Molecular Results

The haplotype analysis shows four distinct, unconnected haplotype networks that mirror almost exactly our geographical drainage groupings (Figure 4.1B). The exceptions include a particular population from the Mississippi drainage (indicated with a triangle) that has haplotypes that fall in both the Mississippi network and the Western network, and the individuals from the Etowah (indicated with a star) that, although geographically part of the Eastern drainage system, groups genetically with the Apalachicola. These patterns are also present in the phylogenetic analysis (Figure 4.1C), with bootstrap support for all of the clades that represent the haplotype networks with the exception of only moderate support for the Western clade, but with support for a Mississippi/Western clade. Monophyly of *Miniellus longirostris* is also supported.

Genetic distances (Table 4.3) were calculated based on the haplotype network structuring (Mississippi/Western calculated both together and separate; Etowah population grouped with Apalachicola) and show relatively low within group mean distances (0.55%-1.68% compared to between group mean distances (range of 3.24% between Mississippi and Western to 7.34% between the Mississippi and Apalachicola). Distances varied from the outgroup taxa by as little as 7.19% up to 11.05%.

Geometric Morphometrics Results

The spread of all points on the consensus wireframe (Figure 4.2) shows low variation across all 171 individuals. The PCA (Figure 4.3) shows that PC1, which explains 34.44% of the variation, has so little useful variation that the small degree of warping seen in a few individuals is affecting the PCA substantially; thus, artifacts of preservations explain more of the shape difference than anything biological. The spread of individuals from all the drainage groupings across this axis shows that all groupings were subject to this artifact. PC2 (17.22% of the variation) shows a shortening of the caudal peduncle and elongation of the anterior region for individuals that are high along this axis. PC 3 (8.18% of the variation) primarily describes the body depth, with individuals high on this axis exhibiting a slightly deeper body. Removing PC 1 from interpretation and focusing on PC 2 versus PC 3 reveals that these morphological features do not provide any separation among the drainage groupings and that the variation in shape is present among all the groupings. While a CVA can provide insight into which shape changes can separate *a priori* groupings, it can be seen in this analysis (Figure 4.4A) that although certain shape changes can pull apart the drainage groupings to a certain degree, there is still a lot of

overlap in shape space. The shifts in shape are almost indistinguishable when laid over the consensus wireframe (Figure 4.4B-C), making them virtually useless for separating populations.

DISCUSSION

Morphological stasis despite genetic divergence

Molecular studies using mitochondrial markers have been useful in delimiting species and evolutionary significant units. One of the primary beliefs in evolutionary biology is that geographically separated populations will gain physical differences from one another due to selection to local conditions or due to random factors that cause the populations to differentiate However, a growing body of literature indicates that speciation may not include morphological distinction because of strong selective pressure maintaining a common form (Avise et al., 1994; Peterson et al., 1999; Kuraku and Kuratani, 2006; Lavoué et al., 2011). These previous studies, however, have focused on divergences at much deeper time scales than those examined here. Factors that may play a role in morphological conservatism across disjunct populations include stabilizing selection, ecological niche conservatism, and genetic and developmental constraints (Erwin, 2007). Any one of these factors could explain the patterns among populations of *Miniellus longirostris* revealed in this study.

The paradox of morphological stasis has been controversial and difficult to adequately explain (Eldridge and Gould, 1972; Futuyma, 2010; Gould and Eldredge, 1977; Wake et al., 1983), but is likely due to strong stabilizing selection (Haller and Hendry, 2014). In the case of *Miniellus longirostris*, the populations cannot be distinguished from one another, even using CVA which is a test designed to separate *a priori* established groups. The lack of distinction

across analyses demonstrates that form has not changed among the populations of *M. longirostris* despite having levels of genetic difference commonly seen between freshwater fish species (Hubert et al., 2008). Without any differences in flow regime, there is no selective pressure present to drive populations to different shapes, and there may be selective pressure to maintain shapes within the habitat. Indeed, *Ericymba amplamala*, a species found sympatrically and in the same habitats with *M. longirostris*, has a very similar body shape, and the results in Chapter 3 suggest that there is likely significant genetic differences between populations of *E. amplamala* while a traditional morphometric study (Pera and Armbruster, 2001) did not find significant differences.

Species of the *Miniellus longirostris* group are very similar to one another, varying mainly in color of the fins and some minor mensural differences (Suttkus and Boschung, 1990; Suttkus, 1991). All of the species are found on a rather homogenous substrate of sand or sand with fine gravel. We believe that this habitat provides little variation and, thus, there is a strong selective force keeping populations of *M. longirostris* from differentiating morphologically from one another.

The fin coloration differences present with the *Miniellus longirostris* group are only based on the relative orange or yellow of the fins and the extent of the colored areas of the fins. Sand-dwelling fishes have to maintain crypsis against their background by matching it in color and being at least somewhat translucent. On a sandy background, there is strong selective pressure maintaining cryptic coloration and against showy colors, thus eliminating the major avenues of morphological variation seen between closely related species of North American fishes.

Alternatively, ecological niche conservatism, the concept that more closely related taxa will occupy similar niches, predicts that speciation is driven by geography, followed by ecological differences (Holt and Gaines, 1992; Peterson et al., 1999). While the similar habitats where the *Miniellus longirostris* group are found could be driving stabilizing selection, another explanation is that their shared ancestral ecological niche could have been conserved without enough time having passed for the accrual of ecological differences that would eventually manifest in morphological distinctions. Studies have shown that this concept can be a good predictor for ecological niches in sister taxa (Peterson et al., 1999; Cooper et al., 2011) and we find the same pattern for taxa closely related to *M. longirostris*.

Genetic or developmental constraints could restrict deviation from the shared morphology across the populations. Without further genetic and experimental data, this hypothesis remains untested, although it could help explain the morphological conservatism seen across the entire clade of shiners that has hampered taxonomic and phylogenetic clarity (see Chapter three).

Deviations from our a priori groupings

Of particular interest in our findings is the genetic placement of specimens from the upper Etowah River in Georgia. While the Etowah River currently flows west to eventually join the Alabama River drainage, specimens were found to be more closely related to individuals from the Apalachicola River drainage. The close proximity of the Etowah River to tributaries of the Apalachicola River provide two possible explanations. Although bait-bucket transfer is a possibility, the Etowah River is a well-known area of river capture with studies dating back to

Campbell (1896). Several species indicative of the Chattahoochee River are found in the Etowah, including *Miniellus lutipinnis, M. xaenocephalus, Ameiurus brunneus, Hypentelium etowanum,* and *Fundulus stellifer* (Ramsey, 1965; Bryant et al., 1979). Bryant et al. (1979) further speculated that the presence of *Ericymba amplamala* in the Etowah was a result of transfer from the Chattahoochee rather than from populations in the Mobile River Drainage. *Ericymba amplamala* is found in coastal plain streams, and the Etowah population is disjunct from those lower in the Mobile River drainage. *Ericymba amplamala* and *M. longitorstris* are sympatric across much of their ranges, and occur in the same habitats. Recent studies have also shown genetic connectivity between the Etowah River and either the Chattahoochee River or Atlantic drainages (Kozak et al., 2005; Scott et al., 2009).

The only other deviation from our *a priori* groupings involve specimens collected from the Big Black River in Mississippi (see triangle in Figure 1.A). This locality is represented in this study by three GenBank sequences and four new sequences (Table 1). While all the GenBank sequences grouped with the rest of the Mississippi network, the new sequences fell in both the Mississippi (one sequence) and the Western network (three sequences), despite all four of them belonging to the same collection. They shared the same haplotype as some representatives from the Pearl River. While the Big Black River collection site and the two Pearl River collections sites are over 157 km apart, the shortest distance between these two rivers is approximately 25 km. This suggests either continued, albeit somewhat restricted, gene flow between these two rivers, or a more recent connection between them, or perhaps bait bucket transfer.

Distinct species?

Whether the various populations of *Miniellus longirostris* defined by our haplotype networks should be recognized as separate species is a matter of debate. Some populations fulfill the Biological Species Concept (Mayr, 2000) by not experiencing interbreeding and the Evolutionary Species Concept (Wiley and Mayden, 2000) by having their own distinct evolutionary fates, but they are not morphologically diagnosable entities per the Phylogenetic Species Concept (Wheeler and Platnick, 2000). None of the populations are of special concern as the species exists in high numbers in appropriate habitats, so there is no conservation reason to recognize the populations as separate. We continue to recognize all of the populations as a single species and encourage the examination of life colors to determine if any of the populations deserve separate species status.

Despite our hesitation to recognize any of our groupings as distinct species, it is important to note that the genetic patterns recovered by this study are not surprising when compared to other studies that examine cryptic diversity across the southeastern U.S. (April et al., 2011; Butler and Mayden, 2003; Schneider et al., 2012; Wooten et al., 1988). For example, the Apalachicola River has been well documented as a genetic break for species such as *Amia calva*, *Micropterus salmoides,* and various *Lepomis* species (Bermingham and Avise, 1986). Swift et al. (1985) examined ranges of more than 230 recognized southeast species in an effort to understand the zoogeography of freshwater fishes and found that lowland vicariant patterns existed from the Pontchartrain to the Choctawhatchee drainages. We find the same general pattern for the Western grouping, although we have some evidence for gene flow with at least one tributary of the Mississippi River. With the recognition of *Miniellus ammophilus* (Suttkus and Boschung, 1990), the distribution of *Miniellus longirostris* was essentially bifurcated near

Mobile Bay. It is unsurprising then, to find a genetic break at this location, and indeed the Alabama/Mobile River has been recognized as a boundary for many species of fish in the southeast (Wiley and Mayden, 1985).

Although no population of *Miniellus longirostris* is imperiled, it is important to note that morphological differentiation may not be correlated with speciation. To test whether the mitochondrial lineages are deserving of specific status will require analyses of nuclear genes, however, and we do not describe separate species for each of the lineages at this time. This study illustrates the importance of identification of putative species, or populations in the process of speciation, using mtDNA sequences to identify genetic structure based on geographic locations in the absence of morphological differentiation. We suspect that genetic divergence has accrued across the populations that have limited possibility for gene flow, but that morphology, at least in shape space, remains similar due to stabilizing selection, ecological niche conservatism, and/or genetic and developmental constraints.

REFERENCES

April, J., Mayden, R.L., Hanner, R.H., Bernatchez, L., 2011. Genetic calibration of species diversity among North America's freshwater fishes. Proc. Natl. Acad. Sci. U.S.A. 108, 10602–10607. doi:10.1073/pnas.1016437108

Armbruster, J.W., 2012. Standardized measurements, landmarks, and meristic counts for cypriniform fishes. Zootaxa, 3586: 8–16

Avise, J.C., Nelson, W.S., Sugita, H., 1994. A Speciational History of "Living Fossils": Molecular Evolutionary Patterns in Horseshoe Crabs. Evolution 48, 1986–2001. doi: 10.2307/2410522?ref=search-gateway:a2b889687e82f5d6a91af04f08a8d49f

Bart, H.L., Jr, Taylor, M.S., Harbaugh, J.T., Evans, J.W., Schleiger, S.L., Clark, W., 1995. New distribution records of Gulf Slope drainage fishes in the Ocmulgee River system, Georgia. Southeastern Fishes Council Proceedings 30, 4–10.

Berendzen, P.B., Simons, A.M., Wood, R.M., Dowling, T.E., Secor, C.L., 2008. Recovering cryptic diversity and ancient drainage patterns in eastern North America: Historical biogeography of the *Notropis rubellus* species group (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 46, 721–737. doi:10.1016/j.ympev.2007.07.008

Bermingham, E., Avise, J.C., 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113, 939–965.

Boschung, H.T., 1992. Catalogue of freshwater and marine fishes of Alabama. Alabama Museum of Natural History Bulletin.

Boschung, H.T., Mayden, R.L., 2004. Fishes of Alabama. Smithsonian Books Washington, DC.

Butler, R.S., Mayden, R.L., 2003. Cryptic biodiversity. Endangered Species 28, 25.

Bryant, R.T., Bauer, B.H., Ryon, M.G., 1979. Distributional notes on fishes from Northern Georgia with comments on the status of rare species. Southeastern Fishes Proceedings 1, 8.

Campbell, M.R., 1896. Drainage Modifications and Their Interpretation. The Journal of Geology 4, 567–581. doi:10.2307/30054974?ref=search-gateway: 1ec1391d4c923c87ccc2a26903e145f0

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9, 1657–1659.

Cooper, N., Freckleton, R.P., Jetz, W., 2011. Phylogenetic conservatism of environmental niches in mammals. Proceedings of the Royal Society B: Biological Sciences 278, 2384–2391. doi:10.1073/pnas.0704469104

Crandall, K.A., Templeton, A., Sing, C., 1994. Intraspecific phylogenetics: problems and solutions. Systematics Association Special Volume.

Eldridge, N., Gould, S.J., 1972. Punctuated equilibria: an alternative to phyletic gradualism. Models in paleobiology.

Erwin, D.H., 2007. Disparity: morphological pattern and developmental context. Palaeontology 50, 57–73.

Futuyma, D.J., 2010. Evolutionary constraint and ecological consequences. Evolution 64, 1865–1884. doi:10.2307/40793081?ref=search-gateway:58d7440b4f9f583d3266dae2d0de8b43

Geneious (v.6) created by Biomatters. Available from http://www.geneious.com, 2013.

Gidmark, N.J., Simons, A.M., 2014. Cyprinidae: Carps and Minnows, in: Warren, M.L., Jr, M, B.B. (Eds.), Freshwater Fishes of North America. Johns Hopkins University Press, Baltimore, pp. 334–450.

Gilbert, C.R., 1978. Type catalogue of the North American cyprinid fish genus Notropis. Bulletin of the Florida State Museum, Biological Sciences 23, 1–104.

Gould, S.J., Eldredge, N., 1977. Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered. Paleobiology 3, 115–151. doi:10.2307/2400177?ref=search-gateway:884a4bfc5146864bcd389f712ad327e2

Haller, B.C., Hendry, A.P., 2014. Solving the paradox of stasis: squashed stabilizing selection and the limits of detection. Evolution 68, 483–500. doi:10.5061/dryad.0jj03

Hay, O.P., 1881. On a collection of fishes from eastern Mississippi. Proceedings of the United States National Museum 3, 488–515.

Holt, R.D., Gaines, M.S., 1992. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evol Ecol 6, 433–447.

Hubert, N., Hanner, R., Holm, E., Mandrak, N.E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J., Bernatchez, L., 2008. Identifying Canadian Freshwater Fishes through DNA Barcodes. PLoS ONE 3, e2490. doi:10.1371/journal.pone.0002490.s002

Ivanova, N.V., Zemlak, T.S., Hanner, R.H., HEBERT, P.D.N., 2007. Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes 7, 544–548. doi:10.1111/j.1471-8286.2007.01748.x

King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Molecular Ecology 17, 4684–4698. doi:10.1111/j.1365-294X.2008.03931.x

Kozak, K.H., Blaine, R.A., Larson, A., 2005. Gene lineages and eastern North American palaeodrainage basins: phylogeography and speciation in salamanders of the Eurycea bislineata species complex. Molecular Ecology 15, 191–207. doi:10.1111/j.1365-294X.2005.02757.x

Kuraku, S., Kuratani, S., 2006. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. Zool. Sci. 23, 1053–1064. doi:10.2108/zsj.23.1053

Lavoué, S., Miya, M., Arnegard, M.E., McIntyre, P.B., Mamonekene, V., Nishida, M., 2011. Remarkable morphological stasis in an extant vertebrate despite tens of millions of years of divergence. Proceedings of the Royal Society B: Biological Sciences 278, 1003–1008. doi:10.1146/annurev.ecolsys.31.1.533

Mayden, R.L., Simons, A.M., Wood, R.M., Harris, P.M., Kuhajda, B.R., 2006. Molecular systematics and classification of north american notropin shiners and minnows, in:. pp. 1–30.

Mayr, E., 2000. The Biological Species Concept, in: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory: a Debate. Columbia University Press, pp. 17–29.

Pera, T.P., Armbruster, J.W., 2006. A new species of Notropis (Cypriniformes: Cyprinidae) from the southeastern United States. Copeia 2006, 423–430.

Peterson, A.T., 1999. Conservatism of Ecological Niches in Evolutionary Time. Science 285, 1265–1267. doi:10.1126/science.285.5431.1265

Ramsey, J.S., 1965. Zoogeographic studies on the freshwater fish fauna of rivers draining the southern Appalachian region. PhD dissertation: Tulane University, New Orleans, LA, USA

Schneider, C.H., Gross, M.C., Terencio, M.L., Porto, J.I.R., 2012. Cryptic diversity in the mtDNA of the ornamental fish Carnegiella strigata. Journal of Fish Biology 81, 1210–1224. doi:10.1111/j.1095-8649.2012.03403.x

Scott, C.H., Cashner, M., Grossman, G.D., Wares, J.P., 2009. An awkward introduction: phylogeography of *Notropis lutipinnis* in its "native" range and the Little Tennessee River. Ecology of Freshwater Fish 18, 538–549. doi:10.1111/j.1600-0633.2009.00368.x

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., Soltis, P.S., 2006. Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15, 4261–4293. doi:10.1111/j.1365-294X.2006.03061.x

Ståhls, G., Savolainen, E., 2008. MtDNA COI barcodes reveal cryptic diversity in the Baetis vernus group (Ephemeroptera, Baetidae). Molecular Phylogenetics and Evolution 46, 82–87. doi:10.1016/j.ympev.2007.09.009

Suttkus, R.D., 1991. *Notropis rafinesquei*, a new cyprinid fish from the Yazoo River system in Mississippi. Bulletin of the Alabama Museum of Natural Museum of Natural History 10, 1-9

Suttkus, R.D., Boschung, H.T., 1990. *Notropis ammophilus*, a new cyprinid fish from Southeastern United States. Tulane Stud. Zool 27, 49–63.

Swift, C.C., Gilbert, C.R., Bortone, S.A., Burgess, G.H., Yerger, R.W., 1985. Zoogeography of the freshwater fishes of the southeastern United States: Savannah River to Lake Pontchartrain, in: Hocutt, C.H., Wiley, E.O. (Eds.), Zoogeography of North American Freshwater Fishes. John Wiley and Sons, New York, pp. 213–265.

Wake, D.B., Roth, G., Wake, M.H., 1983. On the problem of stasis in organismal evolution. Journal of Theoretical Biology.

Wheeler, Q.D., Platnick, N.I., 2000. The Phylogenetic Species Concept (sensu Wheeler and Platnick), in: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory: a Debate. Columbia University Press, pp. 55–69.

Wiley, E.O., Mayden, R.L., 2000. The Evolutionary Species Concept, in: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory: a Debate. Columbia University Press, pp. 70–89.

Wiley, E.O., Mayden, R.L., 1985. Species and speciation in phylogenetic systematics, with examples from the North American fish fauna. Annals of the Missouri Botanical Garden 72, 596–635.

Witt, J., Threloff, D.L., Hebert, P., 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Molecular Ecology 15, 3073–3082.

Wooten, M.C., Scribner, K.T., Smith, M.H., 1988. Genetic variability and systematics of Gambusia in the southeastern United States. Copeia 283–289.

Table 4.1. Specimens used for genetic analyses. MMNS = Mississippi Museum of Natural Sciences, SLUM = St. Louis University Museum, NCSM = North Carolina Museum of Natural Sciences, FLMNH = Florida Museum of Natural History, KU = University of Kansas Biodiversity Institute, AUMNH = Auburn University Museum of Natural History.

Drainage Grouping	Museum Voucher	Locality	GPS Coordinates		Number of Individuals	Genbank Accession Numbers
			Lat (N)	Long (W)		
Mississippi	MMNS 48320	Ouachita	31.81	-91.84	10	
	SLUM 1155 (JC)	Mississippi/Buffalo	31.1	-91.18	10	
	NCSM 44531	Mississippi/Homochitto	31.4	-91.12	3	
	FLMNH 172861	Mississippi/Homochitto	31.5	-90.78	2	
	NCSM HLBSC	Mississippi/Big Black	32.12	-90.77	4	
	GenBank	Mississippi	31.15	-91.54	2	JN027569-70
	GenBank	Mississippi/Big Black	32.12	-90.77	3	JN027576-8
Western	FLMNH 172695	Pontchartrain	30.5	-90.55	3	
	AUMNHl12.05	Escatawpa/Pascagoula	30.86	-88.42	2	
	KU 25792	Pontchartrain	30.46	-90.01	5	
	KU 26850	Pearl	30.77	-89.96	5	
	NCSM 31382	Pascagoula/Leaf	31.47	-89.52	6	
	NCSM HLOC	Pascagoula	31.44	-89.41	2	
	SLUM 1154	Pascagoula/Leaf	31.44	-89.3	10	
	KU 29846	Pearl	32.74	-89.26	5	
	SLUM 1140	Pascagoula	31.05	-89.18	25	
	NCSM HLBC	Pascagoula	31.05	-89.18	2	
	SLUM 1141 (BR)	Biloxi	30.49	-89.04	42	
	NCSM HLPR	Pearl	31.24	-89.85	2	
	GenBank	Pascagoula	30.77	-89.08	2	JN027571-2
Eastern	AUMNH 26764	Conecuh/Escambia	31.13	-87.09	2	
	FLMNH 172747	Alabama/Coosa/Etowah	34.29	-84.27	3	
	KU 29833	Pensacola/Blackwater	30.63	-87.04	3	
	KU 29850	Mobile/Alabama	31.3	-87.71	4	
	NCSM 31461	Mobile/Alabama	31.3	-87.71	2	
	SLUM 1002 (BEC)	Escambia	31.01	-87.26	6	
	SLUM 1003 (YR)	Yellow	31.1	-86.44	10	
	SLUM 1004 (PR)	Choctawhatchee	31.07	-86.17	6	
	SLUM 1005 (FC)	Perdido/Styx	30.7	-87.66	5	
	GenBank	Escambia	31.04	-87.22	1	JN027568
	GenBank	Escambia	30.92	-87.31	1	JN027573
	GenBank	Choctawhatchee	31.6	-85.85	2	JN027574-5

Table 4.1 (*continued*). Specimens used for genetic analyses. MMNS = Mississippi Museum of Natural Sciences, SLUM = St. Louis University Museum, NCSM = North Carolina Museum of Natural Sciences, FLMNH = Florida Museum of Natural History, KU = University of Kansas Biodiversity Institute, AUMNH = Auburn University Museum of Natural History.

Drainage Grouping	Museum Voucher	Locality	GPS Coordinates		Number of Individuals	Genbank Accession Numbers
			Lat (N)	Long (W)		
Apalachicola	FLMNH 173274	Flint	32.8	-84.5	5	
	NCSM 46033	Flint	32.01	-84.23	5	
	AUMNH 63408	Chattahoochee	31.53	-85.21	2	
	AUMNH 61275	Chattahoochee	32.27	-85.21	1	
Outgroup						
	SLUM 1144	*Miniellus rafinesquei*			2	
	NCSM 47445	*Miniellus ammophilus*			1	
	Genbank	*Miniellus ammophilus*			3	HQ 579093, JN027381, JN027390

Table 4.2. Specimens used for geometric morphometric analyses. AUMNH = Auburn University Museum of Natural History.

Drainage Grouping	Museum Voucher	Locality	Number of Individuals
Mississippi	AUMNH 26313	Homochitto	21
Western	AUMNH 26998	Pascagoula/Black	2
	AUMNH 26971	Pontchartrain	29
Eastern	AUMNH 05852	Choctawhatchee/Pea	3
	AUMNH 24166	Choctawhatchee/Pea	1
	AUMNH 31528	Choctawhatchee	12
	AUMNH 36403	Escambia/Conecuh	17
	AUMNH 41848	Perdido	30
	AUMNH 30431	Yellow	16
	AUMNH 31424	Yellow	6
Apalachicola	AUMNH 10589	Chattahoochee	8
	AUMNH 16487	Chattahoochee	4
	AUMNH 30301	Chattahoochee	15
	AUMNH 41715	Chattahoochee	3
	AUMNH 24645	Flint	2
	AUMNH 28399	Flint	2

Table 4.3. Estimates of genetic distance. Values on lower left of matrix represent between group mean distances. Values on upper right of matrix represent the net between group mean distances. Colors respond to geographic groupings indicated on Figure 4.1.

Drainage Group	Within Drainage Group Mean Distances	1	2	3	4	5	6	7
Mississippi and Western	1.55%	—	—	—	4.67%	5.10%	6.78%	8.47%
Mississippi	0.74%	—	—	2.58%	5.69%	6.57%	7.98%	9.16%
Western (other than MS)	0.55%	—	3.24%	—	4.98%	5.26%	7.02%	8.88%
Eastern	1.68%	6.29%	6.91%	6.09%	—	4.61%	7.67%	10.21%
Apalachicola+	0.78%	6.26%	7.34%	5.93%	5.84%	—	6.71%	9.51%
M. ammophilus (outgroup)	0.19%	7.65%	8.46%	7.40%	8.61%	7.19%	—	8.46%
M. rafinesquei (outgroup)	0.00%	9.25%	9.54%	9.16%	11.05%	9.90%	8.56%	—

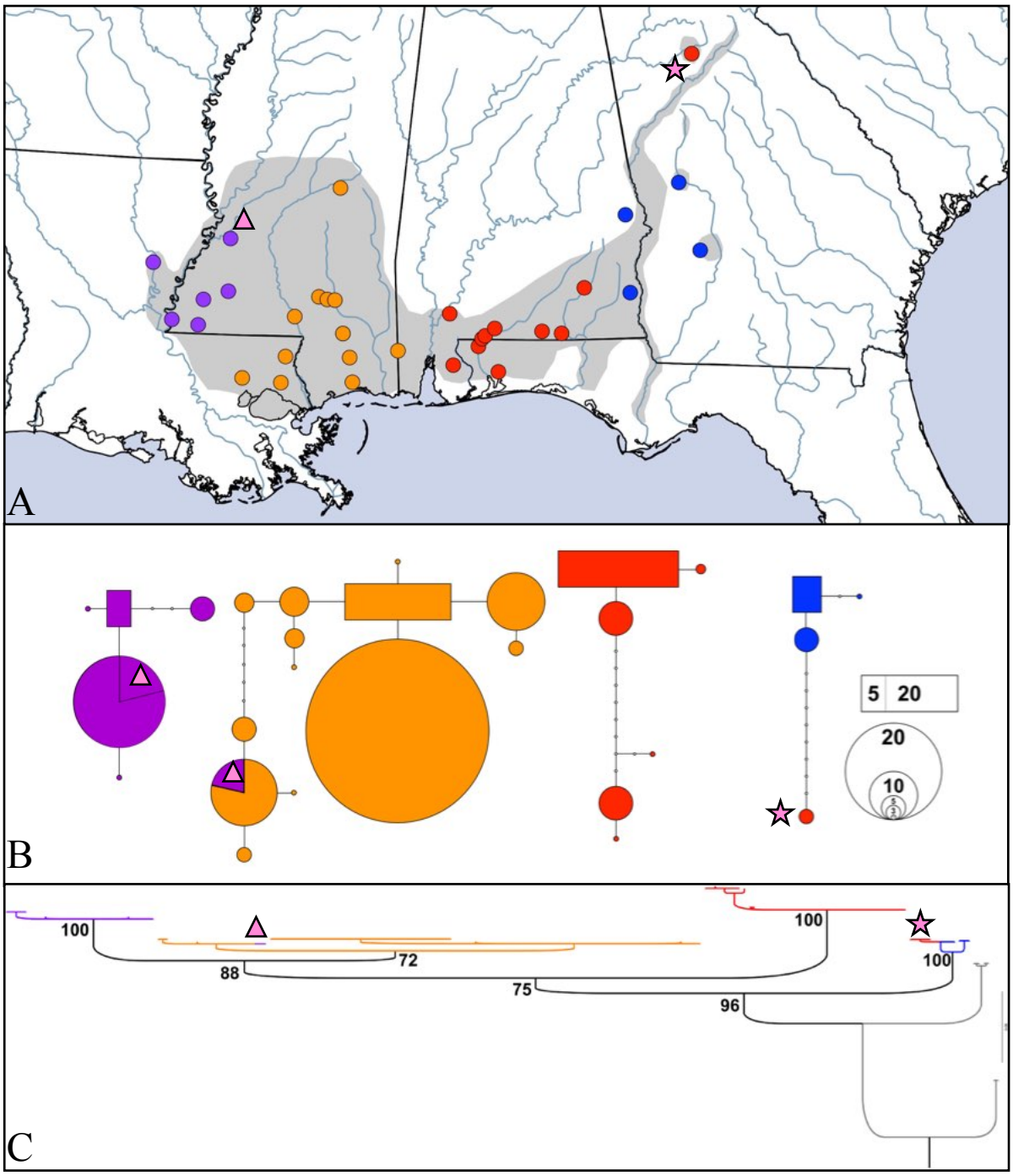

Figure 4.1. (A) Distribution of *Miniellus longirostris* shaded in gray, with sampled localities colored by drainage. Localities near triangle or star represent populations that did not group with the rest of their drainage. (B) Haplotype network showing four distinct networks with very little color overlap. Exceptions include Etowah (star) grouping with Apalachicola populations, and some individuals from the Mississippi drainage grouping with the Western drainage. (C) shows the phylogenetic relationships based on a ML analysis subjected to 1000 bs replicates.

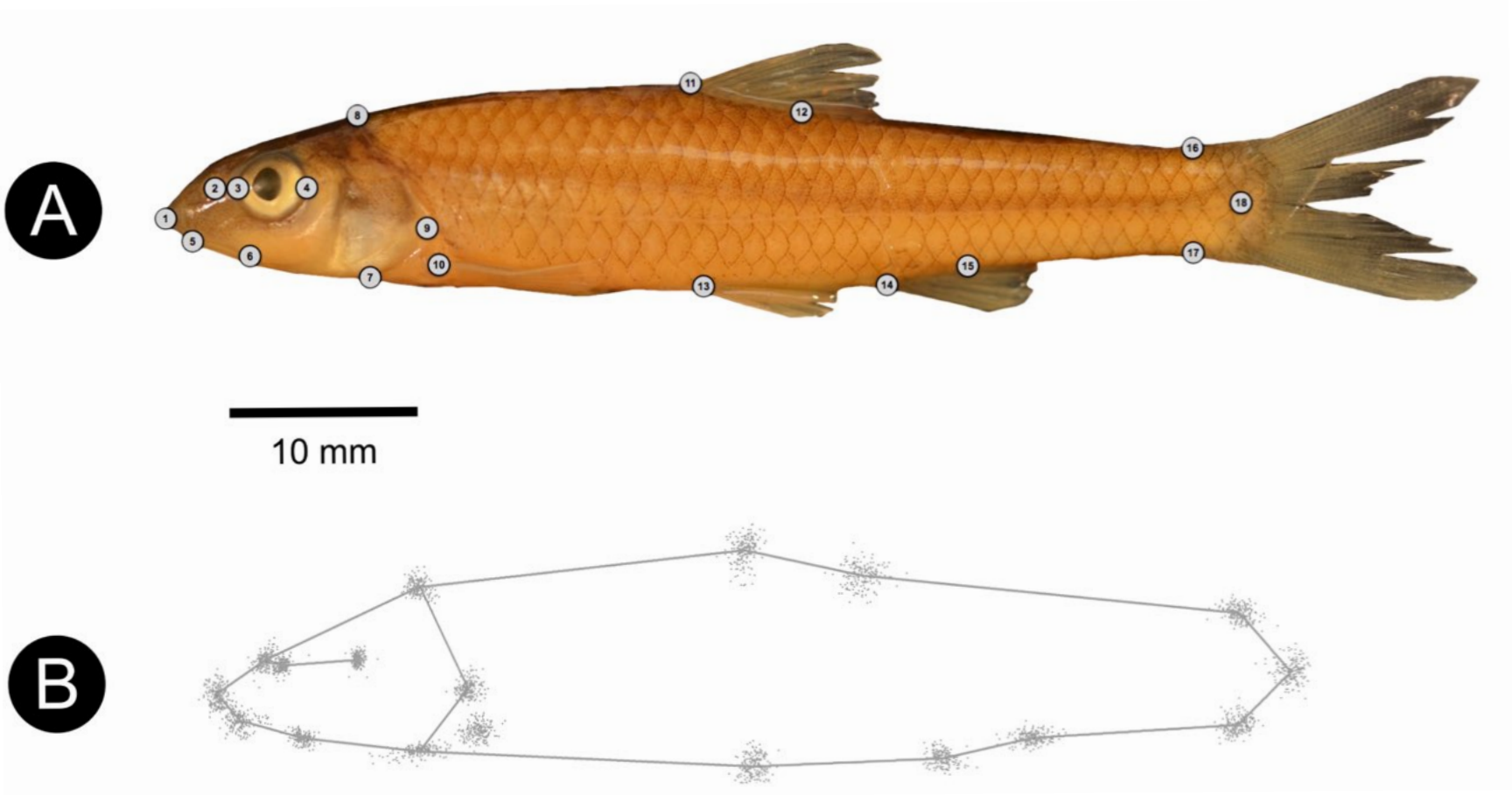

Figure 4.2. (A) Eighteen homologous landmarks used in this study. (B) Spread (variation) of landmarks across all specimens.

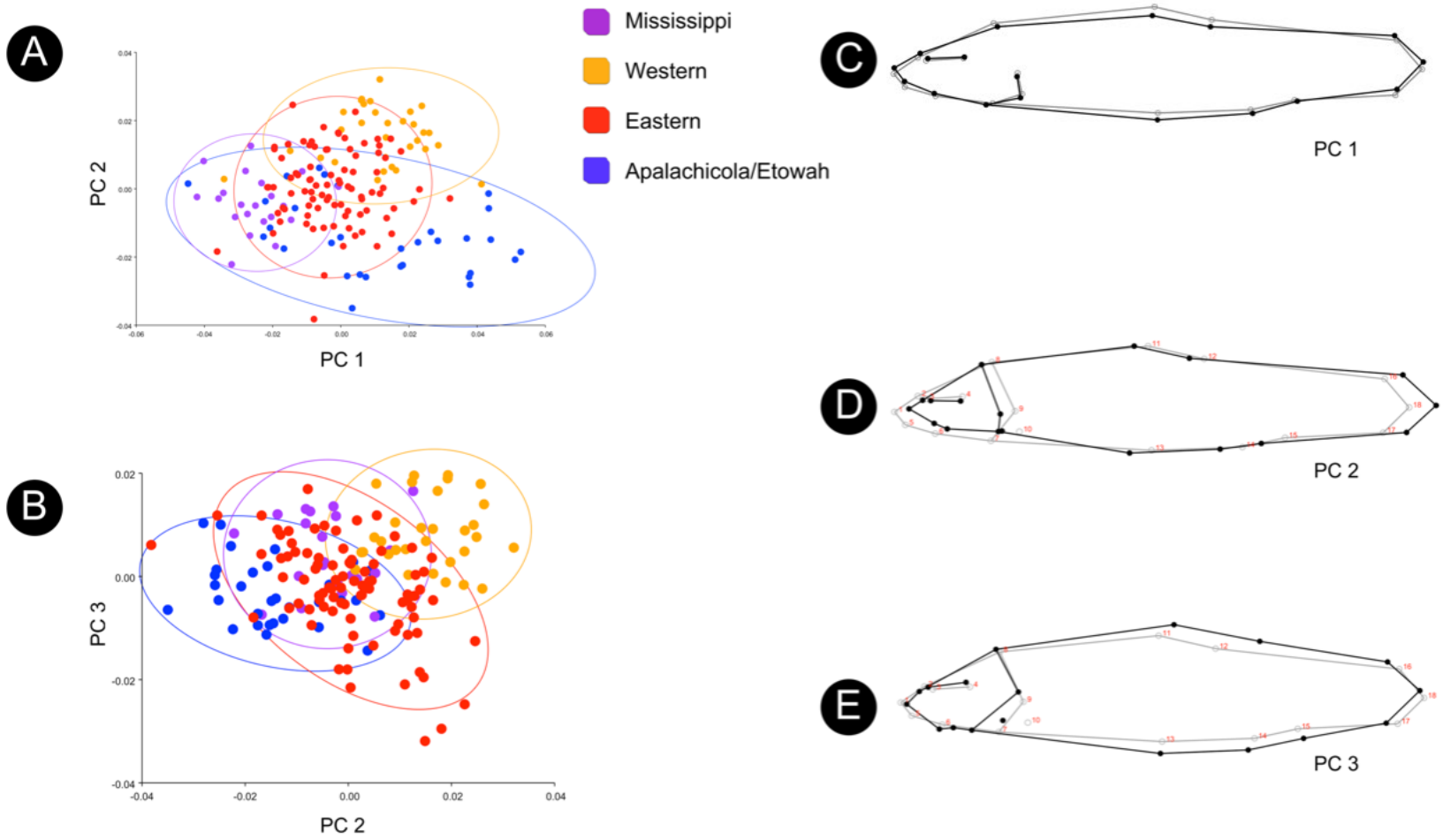

Figure 4.3. A) and (B) Results of PCA comparing PC1, PC2, and PC3. C-E) Comparisons of PC1-3 to the consensus (gray) wireframe.

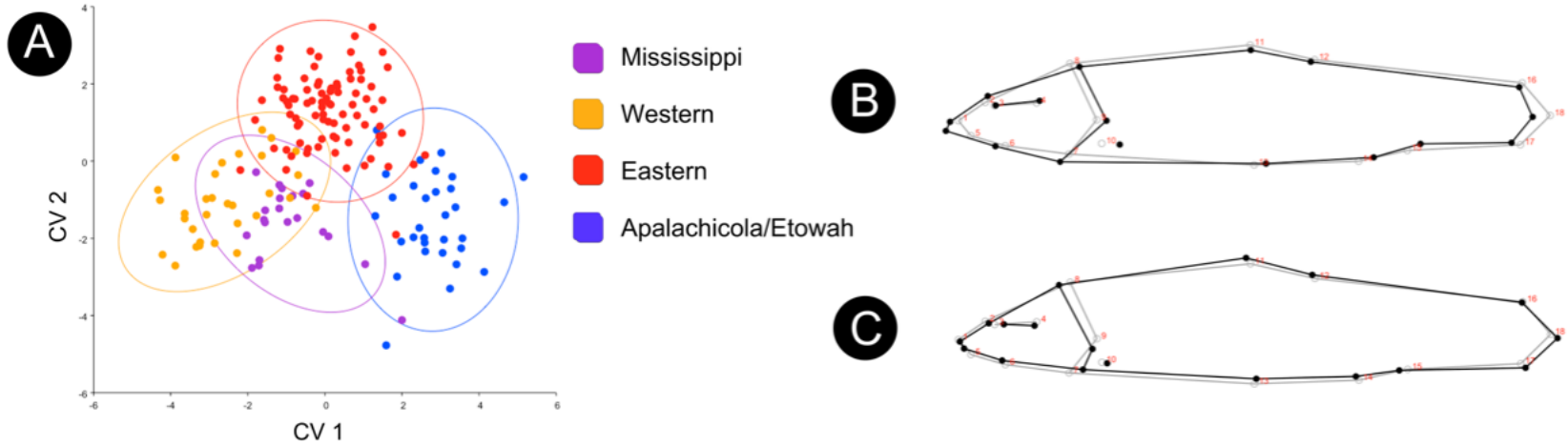

Figure 4.3. A) Results of CVA comparing CV1 and CV2. B-C) Comparisons of CV1 and CV2 to the consensus (gray) wireframe.