ELSEVIER

Investigating the utility of Anchored Hybrid Enrichment data to investigate the relationships among the Killifishes (Actinopterygii: Cyprinodontiformes), a globally distributed group of fishes

Kyle R. Piller [a],[*], Elyse Parker [b], Alan R. Lemmon [c], Emily Moriarty Lemmon [d]

[a] *Department of Biological Science, Southeastern Louisiana University, Hammond, LA 70402, USA*
[b] *Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA*
[c] *Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4120, USA*
[d] *Department of Biological Science, Florida State University, Biomedical Research Facility, Tallahassee, FL 32306-4295, USA*

ARTICLE INFO

ABSTRACT

The Killifishes (Cyprinodontiformes) are a diverse and well-known group of fishes that contains sixteen families inclusive of Anablepidae, Aphaniidae Aplocheilidae, Cubanichthyidae, Cyprinodontidae, Fluviphylacidae, Fundulidae, Goodeidae, Nothobranchiidae, Orestiidae, Pantanodontidae, Poeciliidae, Procatopodidae, Profundulidae, Rivulidae, and Valenciidae and more than 1,200 species that are globally distributed in tropical and temperate, freshwater and estuarine habitats. The evolutionary relationships among the families within the group, based on different molecular and morphological data sets, have remained uncertain. Therefore, the objective of this study was to use a targeted approach, anchored hybrid enrichment, to investigate the phylogenetic relationships among the families within the Cyprindontiformes. This study included more than 100 individuals, representing all sixteen families within the Cyprinodontiformes, including many recently diagnosed families. We recovered an average of 244 loci per individual. These data were submitted to phylogenetic analyses (RaxML and ASTRAL) and although we recovered many of the same relationships as in previous studies of the group, several novel sets of relationships for other families also were recovered. In addition, two well-established clades (Suborders Cyprinodontoidei and Aplocheilodei) were recovered as monophyletic and are in agreement with most previous studies. We also assessed the degree of gene tree discordance in our dataset to evaluate support for alternative topological hypotheses for interfamilial relationships within the Cyprinodontiformes using a variety of different analyses. The results from this study will provide a robust, historical framework needed to investigate a plethora of biogeographic, taxonomic, ecological, and physiological questions for this group of fishes.

1. Introduction

A long-standing debate in systematic biology centers around the question of whether increased taxonomic sampling or increased character sampling will be of greater utility for resolving recalcitrant nodes in the Tree of Life (Hillis, 1998, Graybeal and Cannatella, 1998, Rannala et al. 1998, Hedtke et al. 2006, Heath et al. 2008, Philippe et al. 2011). The pendulum of discussion on these topics routinely swings back and forth (Rosenberg and Kumar 2001, Zwickl and Hillis, 2002, Hillis et al. 2003, Rosenberg and Kumar 2003). However, in response to the development of next-generation sequencing technologies, which enable the attainment of massive amounts of genomic data, much of the emphasis in phylogenetic resolution currently focuses on the inclusion of a greater number of characters. Restriction site associated DNA sequencing (ddRADseq) (Baird et al. 2008) and targeted sequence capture approaches, including Ultra-Conserved-Elements (UCEs, Faircloth et al. 2012) and Anchored Hybrid Enrichment (AHE, Lemmon et al. 2012), are powerful and affordable approaches for genomic data acquisition, particularly for non-model organisms. Although phylogenomics is still in its infancy relative to single gene or morphological approaches, phylogenomic studies, particularly targeted capture sequencing approaches, already have proven useful in the systematic resolution of many groups of organisms in the Tree of Life (Hamilton et al. 2016, Stout et al 2016, Dornburg et al. 2017, Chen et al. 2020).

---

* Corresponding author.
E-mail address: kyle.piller@selu.edu (K.R. Piller).

Phylogenomic datasets hold considerable promise for clarifying relationships for historically ambiguous nodes, but it is well-understood that simply applying more data to species tree inference is not guaranteed to provide resolution to long-standing controversies in systematics (Pyron et al., 2014; Brown and Thomson 2017; Chakrabarty et al. 2017; Reddy et al. 2017; Alda et al. 2019; Zhang et al. 2021). The evolutionary histories of individual genes across the genome may differ from each other and from the species tree as a result of mechanisms such as incomplete lineage sorting (ILS), introgression, low phylogenetic information content, or errors in gene tree estimation (Pamilo and Nei 1988; Maddison 1997). Discordance among gene trees has the potential to yield strongly conflicting hypotheses regarding relationships among species, thereby hindering our understanding of relationships for certain nodes of the Tree of Life (Rokas et al. 2003; Brown and Thomson 2017; Alda et al. 2019; Zhang et al. 2021). Thus, an emergent challenge to species tree inference in the phylogenomic era involves the need to account for the heterogeneity in genealogical histories across hundreds to thousands of sampled loci. Concatenation-based methods of phylogenetic inference, in which multiple genes are joined and analyzed as a single "supergene," may be misled when the most probable gene tree is not congruent with the species tree (i.e., when the evolutionary history of a clade of interest falls within the anomaly zone; Degnan and Rosenberg 2006; Kubatko and Degnan 2007). While analytical strategies such as the multispecies coalescent model (MSC) provide a means of accounting for gene tree discordance due to ILS, inference under the MSC may yield unreliable estimates of species trees under circumstances where ILS is low and gene tree discordance is due mostly to low phylogenetic information content or errors in gene tree estimation (Mirarab et al., 2014). Therefore, robust species tree inference should incorporate explicit investigation of genealogical discordance in phylogenomic datasets in order to identify potentially conflicting phylogenetic signals that may be obscured by strongly supported nodes in species trees generated using concatenation- or coalescent-based methods (Brown and Thomson 2017).

An enduring challenge in teleost fish systematics is represented by the order Cyprinodontiformes (Actinopterygii), a group that has been intensively studied over the last several decades, but for which the familial relationships remain inconsistent. The Cyprinodontiformes are a morphologically and ecologically diverse group of fishes that are globally distributed in tropical and temperate, freshwater and estuarine habitats. It contains more than 1,200 species, including important laboratory study species, such as the Mummichog (*Fundulus heteroclitus*) and popular aquarium fishes including swordtails (*Xiphophorus* spp.) and guppies/mollies (*Poecilia* spp.) (Parenti 1981, Nelson et al. 2016). Fishes in the order display an amazing variety of morphological and physiological adaptations that have long made them a group of interest to ecologists and evolutionary biologists alike (Parenti 1981; Culumber et al. 2012; Reznick et al. 2017, Brown et al. 2019).

The order has recently undergone substantial familial level taxonomic changes and currently includes sixteen families: Anablepidae, Aphaniidae, Aplocheilidae, Cubanichthyidae, Cyprinodontidae, Fluviphylacidae, Fundulidae, Goodeidae, Nothobranchiidae, Orestiidae, Pantanodontidae, Poeciliidae, Procatopodidae, Profundulidae, Rivulidae, and Valenciidae (Parenti 1981, Freyhof et al. 2017, Bragança and Costa, 2018). Phylogenetically, it is a monophyletic group diagnosed by a suite of morphological characters (Parenti 1981, Rosen and Parenti 1981, Costa 1998). Parenti's (1981) osteological study has served as the standard for cyprinodontiform taxonomy for over four decades. Subsequently, other phylogenetic studies of the group, inclusive of many of the families were based on osteology (Costa 1998, 2012a, Hertwig 2008), molecules (Meyer and Lydeard 1993, Pohl et al. 2015, Helmstetter et al. 2016), and combined molecular-morphological data sets (Parker 1997; Ghedotti and Davis 2013). Several of these studies, however, were not true tests of cyprindontiform relationships and monophyly since no non-cyprinodontiform taxa were included (Meyer and Lydeard 1993) or, at the time, they failed to include representatives

of several families in the order (Pohl et al. 2015), and therefore, precluded a comprehensive test of relationships of the families. This suggests that a comprehensive phylogenetic study of the Cyprinodontiformes is needed to better understand the evolutionary relationships among all families within the order, particularly in light of recent family level changes.

The objective of this study was to generate a data-rich phylogenetic hypothesis for the Cyprinodontiformes, inclusive of all sixteen families, using Anchored Hybrid Enrichment (AHE) data (Lemmon et al. 2012). This robust, targeted sequencing approach has produced resolved phylogenies for several groups of fishes (Eytan et al. 2015, Stout et al. 2016, Dornburg et al. 2017) and many other groups of vertebrates (Brandley et al. 2015, Ruane et al.2015, Prum et al. 2015). In addition, recent large-scale bony fish phylogenetic studies, which have used multi-locus approaches, have included several families of cyprinodontiform fishes, but have not included representatives of several of the families in the group (Near et al. 2012; Betancur et al. 2013), many of which have recently been classified as distinct families. A more comprehensive and robust understanding of relationships among the families is needed to provide researchers with a template to address a multitude of evolutionary questions, which will lead to a more accurate estimates of the timing of diversification of the families and allow for temporally based comparative analyses and biogeographic hypotheses to be conducted.

2. Methods

2.1. Taxon sampling

Representatives from each of the families in the order Cyprinodontiformes were included in the study (Table S1) and in most cases, multiple individuals per family were included. Our sampling approach included aquarium and wild-caught individuals and was aimed at including a broad representation of diversity for all sixteen families, rather than detailed survey of the diversity within families. In addition, other taxa within the Atheriniformes and Beloniformes were included to test the monophyly of the order and trees were rooted with *Trichromis salvini* (Cichlidae). Anchored Hybrid loci were mined using AHE probes developed using three genomes provided by C. Martin at UC Berkeley: *Orestias pentlandii* (Orestiidae), *Cubanichthys cubensis* (Cubanichthyidae), and *C. pengelleyi* (Cubanichthyidae).

2.2. Anchored hybrid enrichment data collection and processing

DNA was extracted using a DNeasy tissue kit (Qiagen, Inc) following the manufacturers recommendations. Locus selection, probe design, and data collection were carried out at the Center for Anchored Phylogeny (www.anchoredphylogeny.com) following the protocols outlined in Lemmon et al. (2012) and modified in Stout et al. (2016). We summarize the steps here. Extracted DNA was fragmented to a size range of ∼ 200–500 bp using a Covaris ultrasonicator. After blunt end repair, A-tailing, and adapter ligation, samples were indexed by PCR following Meyer and Kircher (2010). After quantification, libraries were pooled in equal concentration and enriched (in pools of 16 samples) using the teleost AHE probes developed and described by Stout et al. (2016). The quality of enriched library pools was verified using KAPPA qPCR prior pooling and sequencing on four lanes of Illumina HiSeq2500 with a paired-end, 150 bp protocol with 8 bp single-end indexing (129 Gb collected total, ~1.2 Gb collected per sample). All recovered data were deposited on Data Dryad (https://doi.org/10.5061/dryad.1g1jwstxp).

Sequencing reads were processed following methods detailed in Prum et al. (2015) and Hamilton et al. (2016). An outline of the steps follows. Reads passing the CASAVA high-chastity filter were merged (when overlapping) using the Bayesian merging script of Rokyta et al. (2012). Reads were then assembled using a quasi de novo assembler with *Maylandia zebra* and *Xiphophorus maculatus* as divergent references (see Hamilton et al. 2016 for details). Assembly clusters comprised of at

least 680 reads were used to generate consensus sequences. Orthology was assessed using alignment-free pairwise distances in a neighbor-joining algorithm. Orthologous sequences were aligned using Mafft (v7.023b, with -genafpair and -maxiterate 1000 flags), then trimmed and masked to reduce missing characters. Alignments were visually inspected to ensure quality.

2.3. Data accessibility

Raw AHE reads are available on NCBI GenBank Short Read Archive, BioProject ID PRJNA823926. All processed AHE data, assembled data matrices, and tree files are available on Data Dryad (https://doi.org/10.5061/dryad.1g1jwstxp).

2.4. Phylogenetic analyses

We estimated phylogenetic relationships using both ML concatenated (supertree), and coalescent-based approaches. The supertree analysis was conducted in RAxML using a GTR + Γ model partitioned by locus and 100 bootstrap replicates. After estimating locus-specific gene trees in a similar fashion, we used the resulting gene trees to estimate a species tree in ASTRAL II (Mirarab and Warnow, 2015) with 100 bootstrap replicates.

2.5. Gene tree discordance

We used several different approaches to interrogate the degree of gene tree discordance in our dataset and to evaluate support for alternative topological hypotheses (Fig. S1) for interfamily relationships within the Cyprinodontiformes. First, we used the 'RF.dist' function in the R package phangorn (Schliep 2011) to calculate Robinson-Foulds (RF) distances (the number of bipartitions that differ between two phylogenies; Robinson and Foulds 1981) among all individual gene trees as well as between each individual gene tree and the species trees generated using both concatenation- and coalescent-based approaches. To account for differences in the number of taxa between gene trees, we report RF distances as %RF, which represents the ratio of the RF distance between two given trees divided by the maximum RF distance, $2(n - 3)$, where n is the number of tips in the trees being compared (Kuhner and Yamato 2015). In order to determine whether distance between individual gene trees and the species tree may be related to properties such as evolutionary rate or phylogenetic information content of a given locus, we fit a generalized linear model using the function 'glm' in the R package Stats to test for correlations between percentage of variable sites (%VS, used as a proxy for evolutionary rate following Pfeiffer et al. 2019) and %RF or between percentage of parsimony informative sites (%PIS, used a proxy for phylogenetic information following Pie et al. 2019) and %RF. Furthermore, in order to evaluate the impact of increasing %VS or %PIS on inference of phylogenetic relationships, we sorted loci first according to increasing %VS and then according to increasing %PIS and generated two sets of ten bins of loci representing percentiles of increasing %VS and %PIS respectively. Because the number of loci (287) is not evenly divisible by ten, 7 of the bins include 29 loci and three bins include 28 loci. For each bin, we used the software IQ-TREE (Nguyen et al. 2015) to infer a maximum likelihood phylogeny from the concatenated loci after identifying the best-fit nucleotide substitution model for each dataset using ModelFinder (Kalyaanamoorthy et al. 2017). Node support was assessed using an ultrafast bootstrap approximation (UFboot) with 1000 replicates (Minh et al. 2013; Hoang et al. 2018). The trees generated from each of the %VS and % PIS bins were then compared to the RAxML tree by calculating % RF distances. Finally, we used the 'is.monophyletic' function in the R package ape (Paradis and Schliep 2019) to calculate the percentage of gene trees that recovered monophyly of each of the 12 out of 16 families of Cyprinodontiformes that were represented by two or more samples in this study as well as monophyly of several major clades inferred by RAxML

and ASTRAL.

Next, we used BUCKy v.1.4.4 (Ane et al., 2007; Larget et al. 2010) to evaluate gene tree discordance in our AHE datasets by calculating concordance factors (CFs), which represent the proportion of gene trees that recover a particular bipartition. We estimated CFs for two different taxonomic subsets. The first subset was limited to 30 taxa including one individual representing each of the four families Aplocheilidae, Pantanodontidae, Fluviphylacidae, and Orestiidae, and two representatives of each of the 12 remaining families of cyprinodontiform fishes. We additionally included one representative each of the closely related Beloniformes and Atheriniformes. Because there were no loci shared across all 30 samples, we included in our dataset 168 loci that were shared across at least 75% of the sampled taxa (this subset is hereafter referred to as "30tax168loc"). The second taxonomic subset included 28 samples and was identical to the first subset except that the individual representing Orestiidae (*Orestias pentlandii*) and one individual representing Cubanichthyidae (*Cubanichthys pengelleyi*) were both excluded due to high amounts of missing data. For this sample of 28 taxa, we analyzed two subsets of loci - one including the same 168 loci included in the 30tax168loc dataset (hereafter referred to as "28tax168loc") and another including only the 17 loci shared by all 28 individuals ("28tax17loc").

For each data subset, we used MrBayes v.3.2.7 (Ronquist et al. 2012) to infer a posterior distribution of gene trees assuming an HKY model of nucleotide substitution for each AHE locus. Two replicate runs of 10 million MCMC generations each were performed for each locus, with sampling every 1,000 generations. Convergence between independent MCMC runs was diagnosed by evaluating the average deviation of split frequencies and by visualizing trends in the log probability of the data. We then used an implementation of the *MBSum* program in BUCKy to summarize the distribution of gene trees for each locus, which yielded a posterior distribution of 7,500 trees after discarding the first 25% of trees as burn-in. For each data subset (30tax168loc, 28tax168 loc, and 28tax17loc), BUCKy was run for 110,000 generations on four chains, with the first 10,000 generations discarded as burn-in. Each analysis was repeated using different settings for the alpha parameter (α = 1, 5, and 10), which reflect increasing levels of *a priori* discordance among gene trees. We additionally applied a setting of α = ∞, which reflects the assumption that all gene trees are different from each other. Because estimated CFs and topologies were consistent across different alpha parameters for each data subset, we present the results only under a setting of α = 1. A primary concordance tree was then generated using the estimated CFs, and a population tree was inferred using a quartet-joining algorithm in which the quartets with the highest CF are retained.

Finally, we used the approximately unbiased (AU) test (Shimodaira, 2002) to compare the likelihood of alternative phylogenetic hypotheses to the ML phylogeny inferred from RAxML. Specifically, we tested several alternative hypotheses that have been proposed on the basis of previous morphological- and molecular-based systematic studies of the clade as well as alternative topologies supported by analyses in this study. Topological hypotheses include (Fig. S1): (1) alternative phylogenetic placements of Fluviphylacidae; (2) monophyly of Cubanichthyidae; (3) alternative phylogenetic placements of Cubanichthyidae; (4) alternative phylogenetic placements of a clade inclusive of Cubanichthyidae and Cyprinodontidae; (5) alternative phylogenetic placements of Orestiidae; and (6) monophyly of a clade inclusive of Procatopodidae and Poeciliidae. We used the constraint tree search option in IQ-TREE to estimate phylogenies consistent with these alternative topological hypotheses and compared each alternative phylogeny to the ML phylogeny inferred from RAxML using the method of resampling of estimated log-likelihoods (RELL; Kishino et al. 1990) with 10,000 re-samplings.

3. Results

3.1. Molecular data results

Anchored Hybrid loci were captured from 107 OTUs, including 89 species (92 individuals), 14 additional fish taxa within the Atheriniformes and Beloniformes, and a single outgroup taxon. The average locus length was 1,373 bp (101–3,562 bp). The total number of variable sites was 189,881 and the total number of informative characters was 154,079, with an average of 25.4% (1.01–95.9%) missing data. The number of included loci ranged from 12/295 for *Orestias pentlandii* (Orestiidae) to 292/295 for *Profundulus hildebrandi* (Profundulidae) with an average of 244.3 loci across all samples. The families Profundulidae, Goodeidae, and Poeciliidae had the highest number of recovered loci per species (Fig. 1). Samples of Orestiidae, *Cubanichthys cubensis,* and *Cubanichthys pengelleyi* were not directly sequenced in this study and anchored hybrid loci were harvested from previously sequenced genomes. The recovery of anchored hybrid loci from these samples was low (x = 92, 12–155/295 loci). Excluding these samples only increased the average number of loci captured to 253. Outside of the Cyprinodontiformes (i.e. orders Beloniformes and Atheriniformes, and Cichlidae), there was an average of 136.8 loci per sample (range = 134–189 loci).

3.2. Concatenated versus species tree

Comparisons between the concatenated and species trees were mostly similar for the relationships among the families within the Cyprinodontiformes, although support values were generally lower with the species tree analysis (Fig S2). Differences existed with the lack of monophyly for Cyprinodontidae, Valenciidae, and Cubanichthyidae for the species tree analysis. Other studies have shown poorer performance of coalescent species tree approaches in comparison to the concatenated approaches (Gatesy and Springer, 2014, Tonini et al. 2015). The concatenated tree will form the basis of the discussion that follows.

3.3. Phylogenetic resolution

For the ingroup comparisons, resolution was high. Nearly 78% (77.8%) of the nodes at or above the family level were recovered with bootstrap values of 90% or greater and 96.3% of the nodes possessed bootstrap values at 70% or greater (Fig. 2). The single exception was the relationship between the families Orestiidae (*Orestias pentlandii*) and Aphaniidae. All families included in this study were recovered as monophyletic with the exception of Cubanichthyidae. *Cubanichthys pengelleyi* was more closely related to species of Cyprinodontidae rather than to the other congeneric taxon in the data set, *C. cubensis*. Collectively, *C. pengelleyi,* Cyprinodontidae, and *C. cubensis* were strongly supported as a monophyletic group (BS = 96).

The results from this study also confirms the monophyly of the order Cyprinodontiformes, supporting the conclusions of previous studies (Parenti 1981, Rosen and Parenti 1981, Costa 1998, Hertwig 2008). Within the Cyprinodontiformes, two major clades were recovered that correspond to the suborders 1) Cyprinodontoidei (*sensu* Parenti 1981) inclusive of Anablepidae, Aphaniidae, Cubanichthyidae, Cyprinodontidae, Fluviphylacidae, Fundulidae, Goodeidae, Orestiidae, Pantanodontidae, Poeciliidae, Procatopodidae, Profundulidae, and Valenciidae and 2) the Aplocheilodei (*sensu* Parenti 1981) inclusive of Apolocheilidae, Notobranchidae, and Rivulidae.

3.4. Phylogenetic relationships within Cyprinodontoidei

The thirteen families within Cyprinodontoidei were recovered as a monophyletic group. Within this clade, Clade I consists of three families of New World fishes including Poeciliidae, Anablepidae, and Fluviphylacidae. Poeciliidae and Anablepidae were recovered as sister families, with the South American Fluviphylacidae as sister to Poeciliidae + Anablepidae with BS > 95 (Fig. 3). Within the Poeciliidae, 17 genera (*Alfaro, Belonesox, Brachyrhaphis, Carlhubbsia, Gambusia, Girardinus, Limia, Micropoecilia, Neoheterandria, Pamphorichthys, Poecilia, Poeciliopsis, Priapella, Pseudoxiphophorus, Scolichthys, Tomeurus,* and *Xiphophorus*) and 23 species were recovered as monophyletic with all genus level BS values >70. One genus within Poeciliidae, *Phallichthys,* was not recovered as monophyletic. The Anablepidae also was monophyletic and included representatives of all three genera, *Anableps, Jenynsia, Oxyzygonectes,* and three species. The Fluviphylacidae was represented by a single species, *Fluviphylax obscurus*.

Clade II contains four families, including a single New World family, Orestiidae, and three Old World families including Valenciidae, Aphaniidae, and Procatopodidae (Fig. 4). Aphaniidae (*Aphaniops sirhani* and *Anatolichthys anotoliae*), was recovered as sister to Orestiidae (*Orestias pentlandii*), however, support for this relationship was very poor (BS =
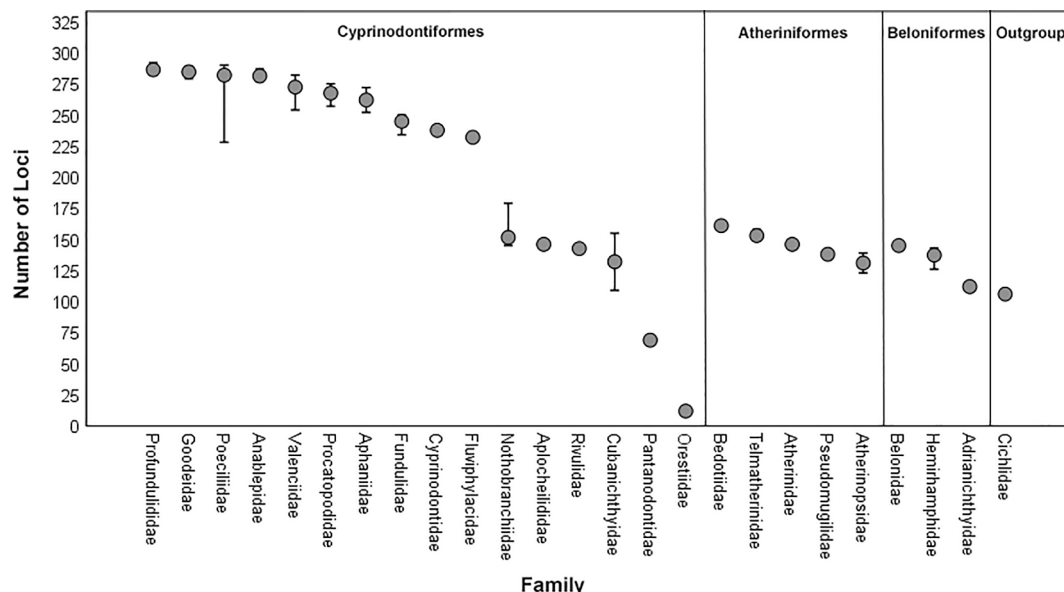

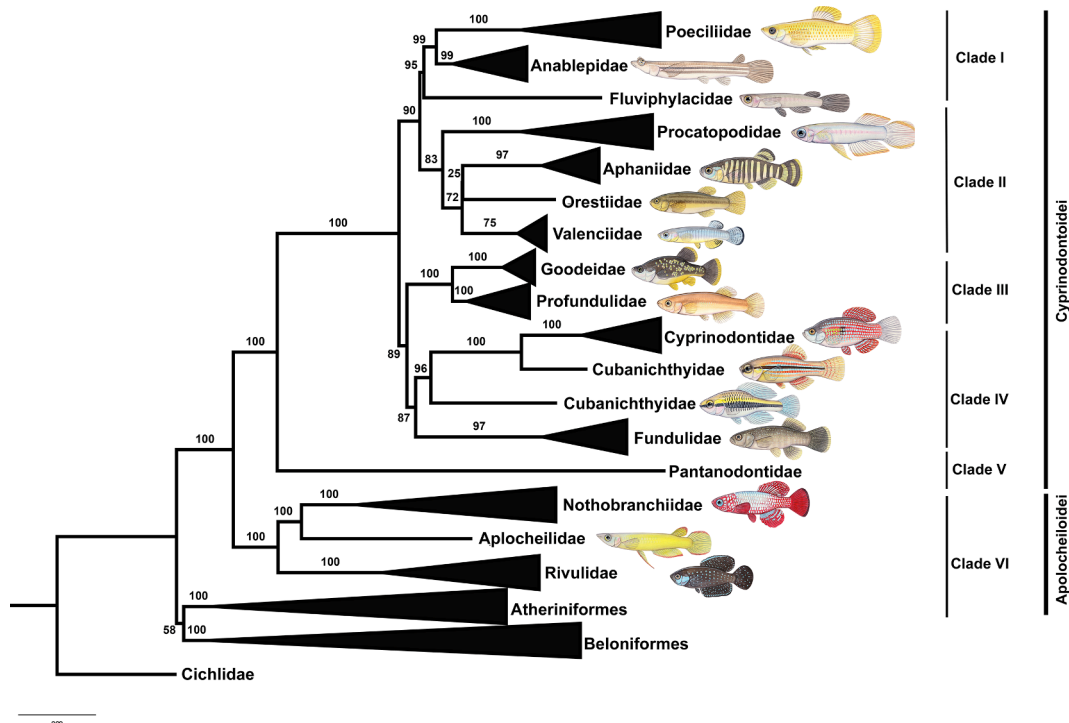**Fig. 1.** Number of recovered Anchored Hybrid Enrichment loci for Cyprinodontiformes and relatives.

Fig. 2. Concatenated Maximum Likelihood tree (RaXML) for the families within the Cyprinodontiformes. The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90 unless otherwise depicted. Species level diversity (Clades I-VI) will be depicted in subsequent phylogenies.

25). Valenciidae, inclusive of all three species (*V. hispanica*, *V. letourneuxi*, and *V. robertae*), was monophyletic and was the sister group to Aphaniidae + Orestiidae. Finally, the African fish family Procatopodidae, inclusive of five genera (*Procatopus, Lamprichthys, Poropanchax*, and *Rhexipanchax*) and eight species, was strongly supported as monophyletic (BS = 100) and sister to all families in Clade II. Only *Poropanchax* was not recovered as monophyletic.

Clade III included a sister group relationship between the Goodeidae and Profundulidae (Fig. 5). Representatives of twelve goodeid genera

(*Alloophorus, Xenotoca, Chaplichthys, Zoogoneticus, Goodea, Allotoca, Xenotaenia, Ilyodon, Girardinichthys, Skiffia, Crenichthys*, and *Empetrichthys*) were included and nodal support within the family was high (11/13 nodes, BS ≥ 100). All genera were monophyletic with the exception of *Xenotoca*. The Profundulidae, inclusive of both described genera (*Tlaloc* and *Profundulus*) and seven species, also was recovered as monophyletic and all nodes possessed bootstrap values of 100.

Clade IV included the other clade of New World fishes, and consisted of the Cyprinodontidae and two species of *Cubanichthys* (Fig. 6). A sister
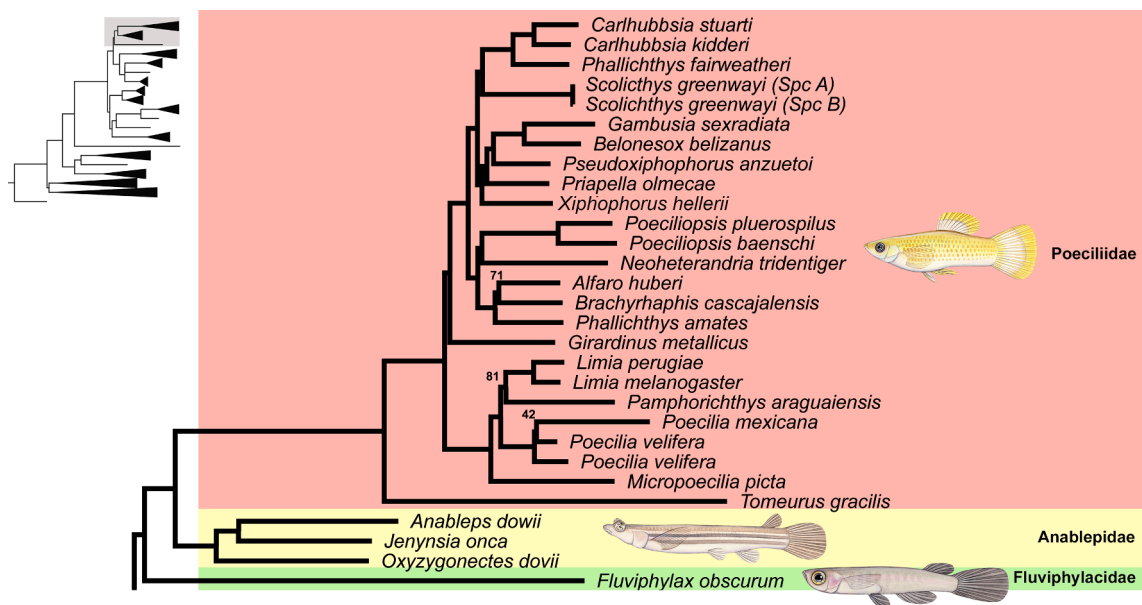

Fig. 3. Expanded Maximum Likelihood tree (RaxML) for Clade I, inclusive of Poeciliidae, Anablepidae, and Fluviphylacidae The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90, unless otherwise depicted.

Fig. 4. Expanded Maximum Likelihood tree (RaxML) for Clade II, inclusive of Protocatopodidae, Aphaniidae, Orestiidae, and Valenciidae. The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90, unless otherwise depicted.
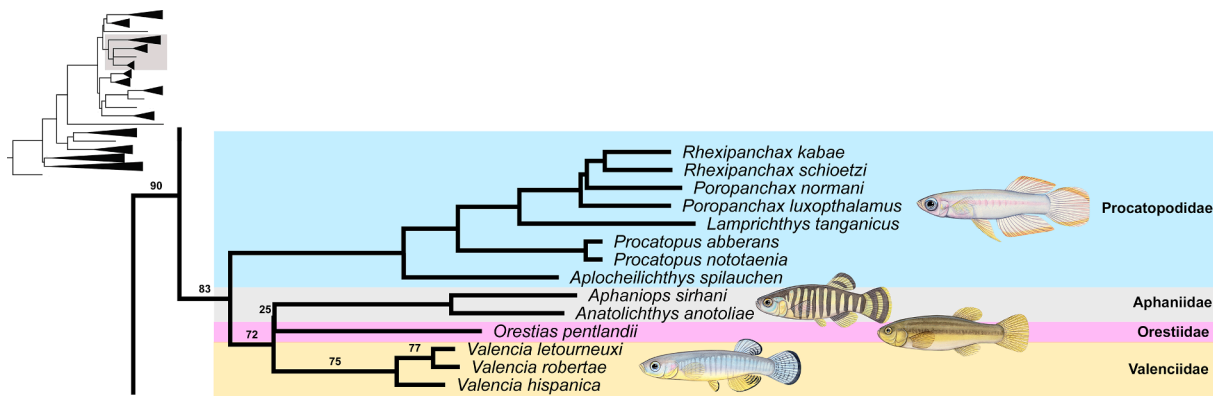

Fig. 5. Expanded Maximum Likelihood tree (RaxML) for Clade III, inclusive of Goodeidae and Profundulidae. The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90, unless otherwise depicted.
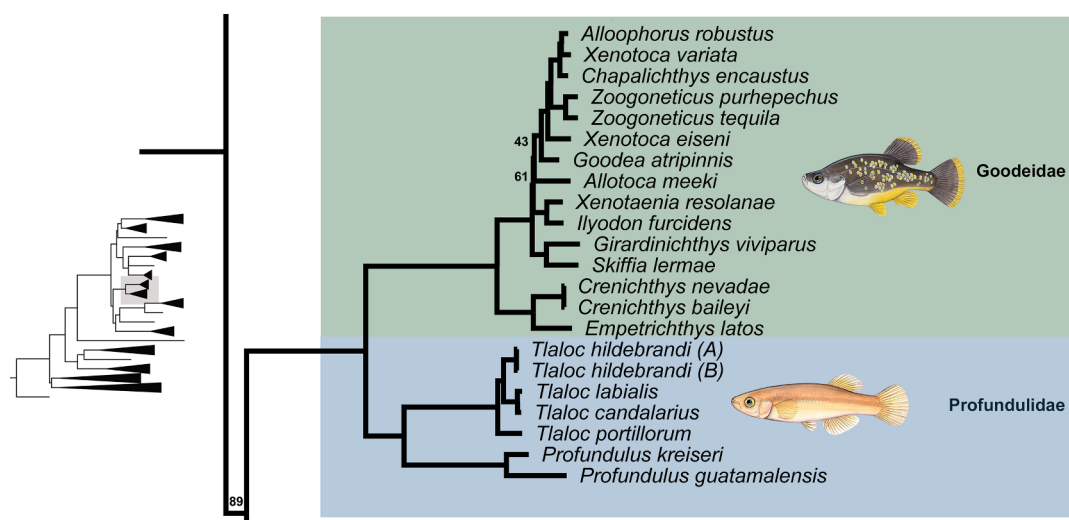

Fig. 6. Expanded Maximum Likelihood tree (RaxML) for Clade IV inclusive of Cyprinodontidae, Cubanichthyidae, and Fundulidae, and Clade V, which only includes Pantanodontidae. The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90, unless otherwise depicted.
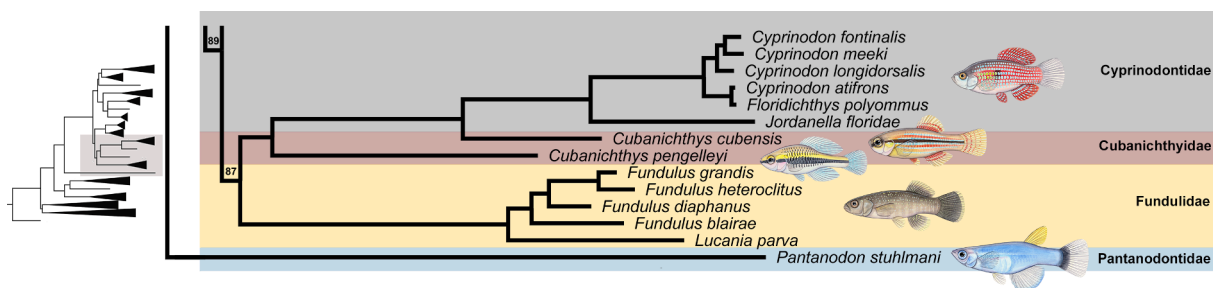
group relationship was inferred between *Cubanichthys cubensis* (Cubanichthyidae) and the Cyprinodontidae, inclusive of the genera *Cyprinodon*, *Floridichthys*, and *Jordanella*. The other species of *Cubanichthys* included in this study, *C. pengelleyi* (Cubanichthyidae), was sister to Cyprinodontidae + *C. cubensis*. All of these relationships were recovered with high bootstrap values (BS = 96–100). The family Fundulidae,

inclusive of species of *Fundulus* and *Lucania*, was sister to the Cyprinodontidae/*Cubanichthys* clade.

Finally, Clade V, inclusive of a single family, Pantanodontidae, represented by a single species (*Pantanodon stuhlmanni*), was recovered as the sister clade to all other families in Cyprinodontoidei with strong support (BS = 100) (Fig. 6).
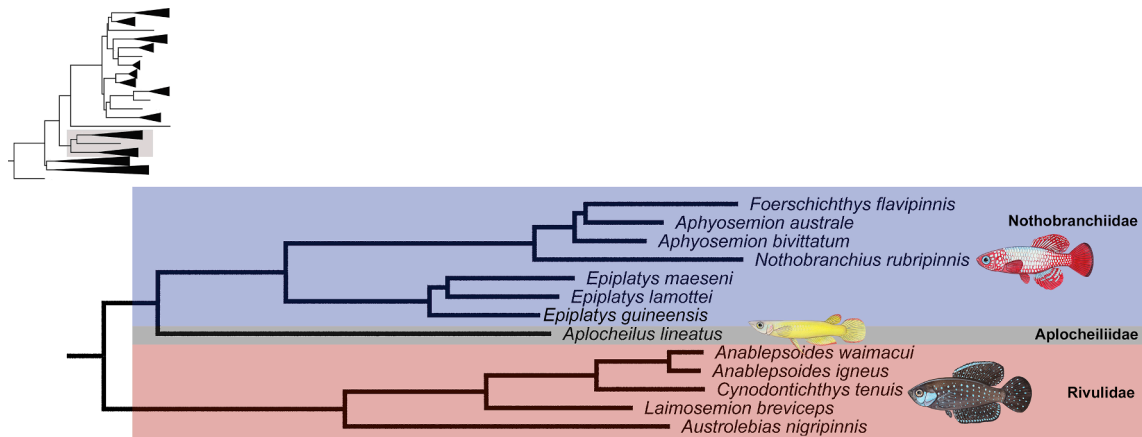
Fig. 7. Expanded Maximum Likelihood tree (RaxML) for Clade VI inclusive of the Nothobranchiidae, Aplocheiliidae, and Rivulide. The scale bar represents the number of substitutions per site and numbers represent bootstrap values. All bootstrap values were ≥ 90, unless otherwise depicted.
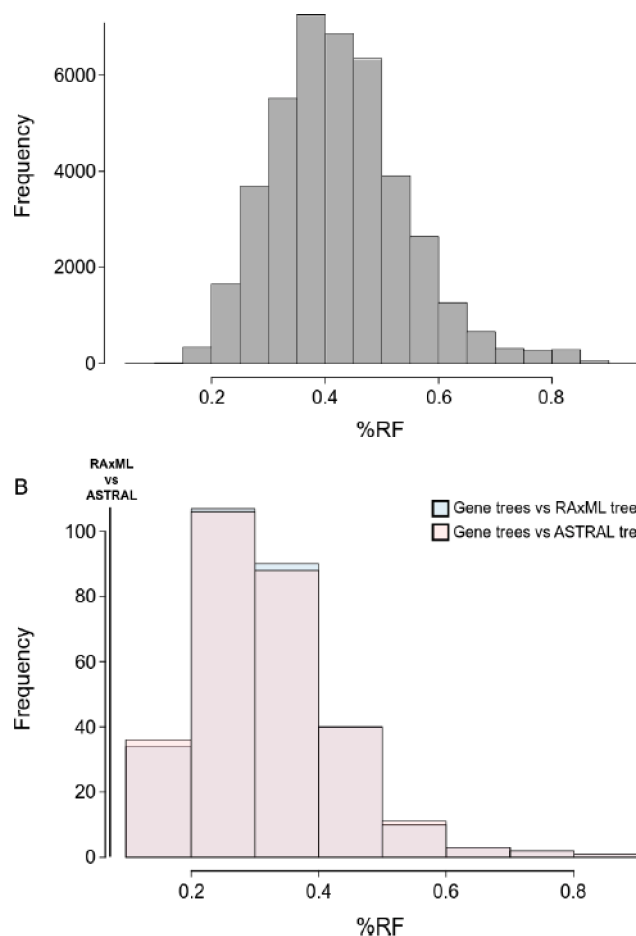

Fig. 8. Histogram depicting the distribution of % RF distances between (A) all pairwise comparisons of 287 gene trees and (B) each gene tree and either the Maximum Likelihood tree (blue bars) or the ASTRAL tree (pink bars). In panel (B) the single black line indicates the % RF distance between the Maximum Likelihood tree and the ASTRAL tree.

4. Phylogenetic relationships within the apolocheloidei

Within Apocheloidei, Clade VI was recovered as monophyletic (BS = 100) (Fig. 7). It contains three families of southern hemisphere fishes including Aplocheilidae, and Nothobranchiidae in the Old World, and Rivulidae in the New World. A strongly supported sister group relationship was recovered for Aplocheilidae and Nothobranchiidae.

Rivulidae, distributed in the Americas, formed the sister group to Aplocheilidae + Nothobranchiidae with strong support (BS = 100). All nodes within Nothobranchiidae (*Epiplatys, Nothobranchius, Aphyosemion*, and *Foerschichthys*) were strongly supported (BS = 100) and only *Aphyosemion* was not recovered as monophyletic. The lack of monophyly for *Aphyosemion* has previously been noted by Parenti (1981) and Murphy and Collier (1997). Within Rivulidae, representatives of four

Table 1
Percentage of gene trees recovering monophyly of each of twelve families of Cyprinodontiformes in addition to several clades recovered by the RAxML and ASTRAL phylogenetic analyses. Node support for monophyly of these clades in the RAxML and ASTRAL trees is also provided, with "n/a" displayed if a given clade is not recovered in the RAxML or ASTRAL tree.

Clade	Percent (%) gene trees recovering monophyly	Bootstrap support in RAxML tree	Bootstrap support in ASTRAL tree
Poeciliidae	69.68	100	100
Anablepidae	83.27	99	96
Procatopodidae	80.84	100	100
Aphanidae	94.08	97	97
Valenciidae	96.17	75	90
Fundulidae	85.02	97	95
Cubanichthyidae	48.43	n/a	n/a
Cyprinodontidae	58.89	100	n/a
Goodeidae	97.21	100	100
Profundulidae	77.70	100	98
Nothobranchidae	61.67	100	100
Rivulidae	53.65	100	100
Cyprinodontoidei	61.67	100	97
Clade I (Poeciliidae + Anablepidae + Fluviphylacidae)	32.40	95	93
Clade II (Orestiidae + Valenciidae + Aphaniidae + Procatopodidae)	63.41	83	80
Clade III (Goodeidae + Profundulidae)	81.88	100	98
Clade IV (Cubanichthyidae + Cyprinodontidae + Fundulidae)	34.84	87	n/a
Aplocheiloidei	57.14	100	100

genera and five species were included and all nodes were recovered with bootstrap values of 100. Aplocheilidae was represented by a single species, *Aplocheilus lineatus*, which was the sister group to Rivulidae.

4.1. Gene-tree and species-tree disagreement

We observed considerable topological discordance between individual gene trees, with a mean %RF = 42.45% (min %RF = 5.56%, max %RF = 90.7%; Fig. 8A). Although monophyly of many of the families within the Cyprinodontiformes (e.g. Fundulidae, Goodeidae, and Valenciidae) and major clades (e.g. Cyprindontoidei and Aplocheilodei)

were supported by a large fraction of individual gene trees (Table 1), there exists evidence for considerable disagreement among gene trees regarding interfamily relationships. For example, only 32.40% of gene trees recover a monophyletic group inclusive of Poeciliidae, Anablepidae, and Fluviphylacidae (Table 1), despite the strong bootstrap support for this clade in both the RAxML tree (BS = 95) and the ASTRAL tree (BS = 93). Similarly, only 34.84% of gene trees recover monophyly of a clade including Cubanichthyidae, Cyprinodontidae, and Fundulidae (Table 1).

Comparisons between individual gene trees and either the RAxML or ASTRAL species tree yielded nearly identical distributions of %RF distances, each with a mean %RF of 31.72% (Fig. 8B). This result is unsurprising given the high degree of similarity between the RAxML tree and the ASTRAL tree (%RF = 1.4%; Fig. 8B). In order to determine whether distance between individual gene trees and the species tree is a function of evolutionary rate or phylogenetic information content of a given locus, we sorted loci according to increasing percentage of variable sites (%VS) and according to increasing percentage of parsimony informative sites (%PIS) and plotted the %RF distances between each individual gene tree and the concatenated RAxML tree. We first investigated whether there was a correlation between %VS or %PIS of each locus and %RF between each locus and the RAxML tree by fitting a generalized linear model, and we find no significant relationship between either evolutionary rate ($p = 0.182$; Fig. 9A,B) or phylogenetic information ($p = 0.812$; Fig. 9C,D) and %RF. We additionally compared trees generated from each of the ten %VS bins and from each of the ten %PIS bins (representing increasing percentiles of %VS and %PIS respectively) to the concatenated Maximum Likelihood phylogeny by calculating %RF distances. In general, %RF distances between the concatenated Maximum Likelihood tree and each of the %VS and %PIS trees are small (mean %RF = 12.3% and 11.1% respectively). We expect that %RF distances will decrease with increasing %VS (Pfeiffer et al. 2019), but instead %RF appears to increase slightly with %VS represented by each bin (Fig. 9B). Similarly, we expect that %RF distances will decrease with increasing %PIS included in each bin, but instead we observe no clear trend of %RF distances with increasing %PIS (Fig. 9D).

The primary concordance tree inferred from BUCKy using a data subset including no missing data (the 28tax17loc dataset) recovers the same family level topology for the Cyprinodontiformes as that recovered by RAxML (Fig. 10). Overall, concordance factors (CF) for many of these relationships are high, indicating that most of the examined gene trees recover most of the same bipartitions among major lineages. The exceptions of high concordance factors are: (a) the sister relationship between Fluviphylacidae and Poeciliidae + Anablepidae (CF = 0.27); (b)
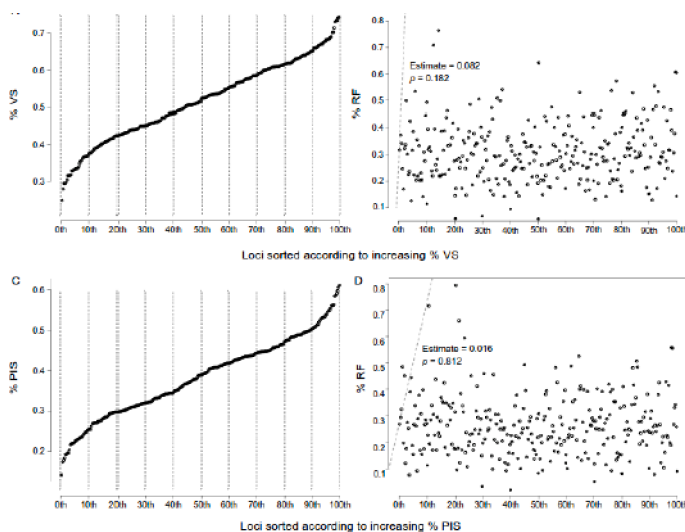

Fig. 9. (A-B) Distribution of percent variable sites (VS) of each locus and depiction of Robinson-Foulds (%RF) distances between each gene tree and the RAxML tree (open circles) as well as between trees estimated from each of the ten % VS bins and the RAxML tree (closed circles). (C-D) Distribution of percent parsimony informative sites (VS) of each locus and depiction of Robinson-Foulds (%RF) distances between each gene tree and the RAxML tree (open circles) as well as between trees estimated from each of the ten % PIS bins and the RAxML tree (closed circles).
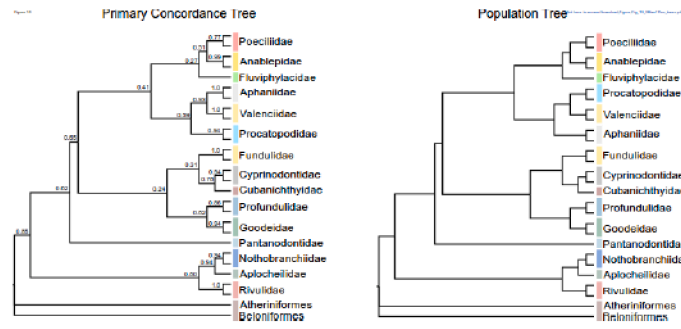
Fig. 10. Primary concordance tree and population tree inferred from BUCKy analysis of the 28tax17loc dataset. Nodes in the primary concordance tree are labeled with estimated concordance factors.
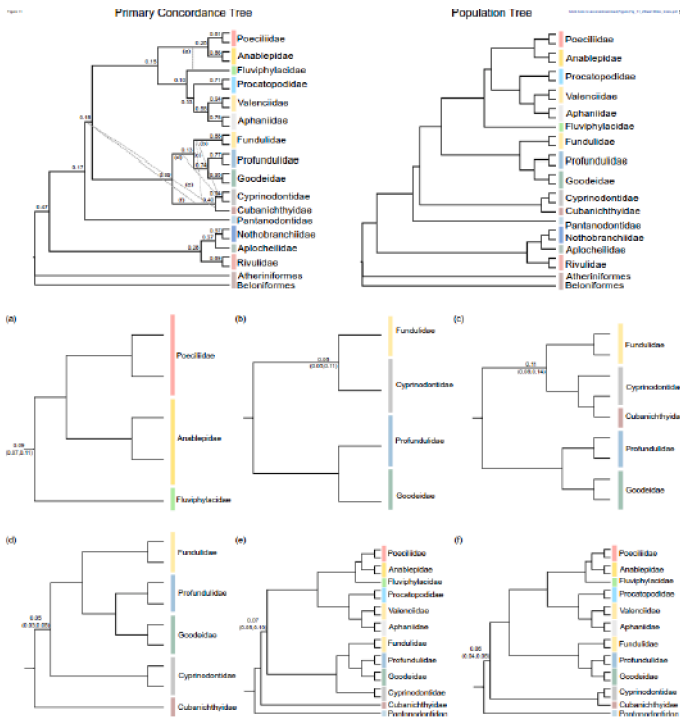

Fig. 11. Primary concordance tree and population tree inferred from BUCKy analysis of the 28tax168loc dataset. Nodes in the primary concordance tree are labeled with estimated concordance factors. Dotted lines connecting two branches indicate conflicting bipartitions (a-f). In each of the panels a-f, only the node representing the conflicting bipartition is labeled with the estimated CF and CF 95% confidence interval that overlaps with that of the bipartition present in the primary concordance tree.

the sister relationship between Fundulidae and Cyprinodontidae + Cubanichthyidae (CF = 0.31); and (c) the split between Fundulidae + Cyprinodontidae/Cubanichthyidae and Profundulidae + Goodeidae (CF = 0.24). However, analyses of the two datasets that each included 25% missing data result in primary concordance trees with topologies that slightly differ from that recovered by RAxML with regard to the phylogenetic placement of Fluviphylacidae, Cyprinodontidae, Cubanichthyidae, and Orestiidae. Specifically, while Fluviphylacidae is recovered as sister to a clade inclusive of Poeciliidae and Anablepidae in the RAxML phylogeny, it is instead recovered as sister to a clade including Procatopodidae, Aphaniidae, and Valenciidae in the primary concordance tree resulting from analysis of the 28tax168loc dataset in BUCKy. Additionally, in primary concordance trees inferred from both the 28tax168loc (Fig. 11) and 30tax168loc (Fig. 12) datasets, a clade inclusive of Cyprinodontidae and one Cubanichthyidae sample (*Cubanichthys cubensis)* is sister to a clade including Fundulidae, Goodeidae, and Profundulidae rather than as sister to Fundulidae as in the RAxML tree. In the BUCKy analysis of the 30tax168loc dataset, the sample

Cubanichthys pengelleyi is recovered as sister to all other families within Cyprinodontoidei exclusive of Pantanodontidae and Orestiidae rather than as nested within a clade including Cyprinodontidae and *Cubanichthys cubensis*, as in the RAxML tree. Finally, while the family Orestiidae is recovered as nested within a clade inclusive of Aphaniidae and Valenciidae in the RAxML tree, it is recovered as sister to all other families within Cyprinodontoidei in the primary concordance tree generated from analysis of the 30tax168loc dataset.

BUCKy additionally calculates CFs for alternative bipartitions not represented in the primary concordance tree. We interpret evidence for conflicting bipartitions by identifying any alternative bipartition that has a CF 95% credible interval that overlaps with that of a clade represented in the primary concordance tree (following MacGuigan and Near 2019). Under this criterion, analysis of the 28tax17loc dataset reveals no evidence for strongly conflicting bipartitions. However, we do identify conflicting bipartitions for the 30tax168loc and 28tax168loc datasets, primarily reflecting uncertainties in the phylogenetic placements of Fluviphylacidae, Cyprinodontidae, Cubanichthyidae, and Orestiidae (Figs. 11 and 12).
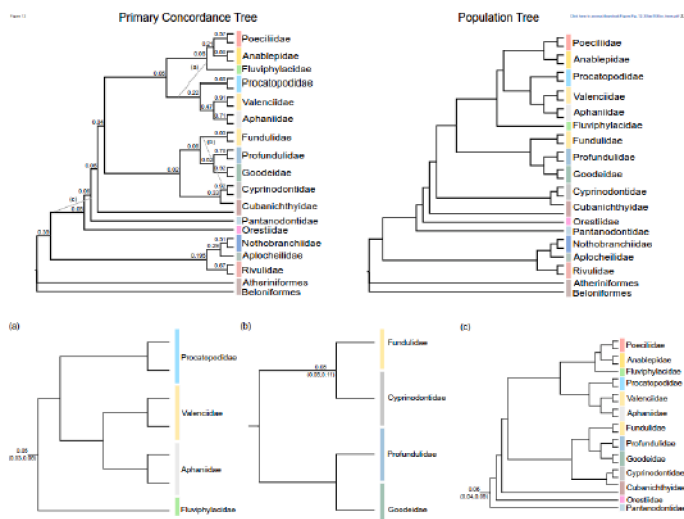
Fig. 12. Primary concordance tree and population tree inferred from BUCKy analysis of the 30tax168loc dataset. Nodes in the primary concordance tree are labeled with estimated concordance factors. Dotted lines connecting two branches indicate conflicting bipartitions (a-c). In each of the panels a-c, only the node representing the conflicting bipartition is labeled with the estimated CF and CF 95% confidence interval that overlaps with that of the bipartition present in the primary concordance tree.

The AU test rejected all alternative topological hypotheses regarding the phylogenetic placement of Fluviphylacidae except for a sister relationship between Fluviphylacidae and a clade inclusive of Poeciliidae and Anablepidae ($p = 0.726$; Table 2). Also rejected were hypotheses regarding the monophyly of Cubanichthyidae ($p = 0.000024$), the monophyly of all taxa within Cyprinodontoidei exclusive of Pantanodontidae and Cubanichthyidae ($p = 0.00$), and monophyly of a clade inclusive of all Procatopodidae and Poeciliidae taxa ($p = 0.000112$; Table 2). However, the AU test did not reject several alternative hypotheses regarding the phylogenetic placement of Orestiidae or of a clade inclusive of Cubanichthyidae and Cyprinodontidae, suggesting remaining uncertainty in the relationships of these taxa relative to other lineages within the Cyprinodontiformes. Specifically, the AU test did not reject the placement of Cubanichthyidae/Cyprinodontidae as sister to either Fundulidae ($p = 0.71$) or a larger clade inclusive of Fundulidae, Goodeidae, and Profundulidae ($p = 0.15$; Table 2). Similarly, only one of the six alternative hypotheses regarding the placement of Orestiidae was rejected (Table 2).

5. Discussion

Over the last several decades, the field of evolutionary biology has witnessed substantial analytical and methodological advances that have allowed us to address more complex and challenging questions in a variety of areas including the estimation of diversification rates, character evolution, and divergence time estimation to name a few (Drummond and Rambaut 2007, Garamszegi 2014, Mirarab et al., 2014, Rabosky 2014). All of these approaches, however, require high quality historical templates to address these types of questions. Fortunately, we have moved into an analytical arena where character limitation is no longer an issue when it comes to phylogenetic resolution. The ease of generating massive amounts of genomic data have facilitated advancement in modern phylogenetic studies and have allowed us to tackle some of the most difficult and comprehensive questions in evolutionary biology (Smith et al. 2011, Eytan et al. 2015, Prum et al. 2015, Hamilton et al. 2016), including resolving the relationships among the families within the Cyprinodontiformes.

This study represents the first phylogenomic analysis of the order Cyprinodontiformes, inclusive of all sixteen recognized families in the group, and confirms the monophyly of the superfamily, a group that spans more than 70 million years of evolutionary history (Helmstetter et al. 2016). Although the relationships among the families have been in conflict for some time (Parenti 1981, Costa 1998, Pohl et al. 2015), the utilization of anchored hybrid loci provides resolution and strong support for nearly all of the families within the Cyprinodontiformes. Furthermore, the inclusion of anchored hybrid loci also provides resolution for many of the shallow nodes in the tree, further supporting this approach as a valid technique for resolution across multiple hierarchical taxonomic levels, like it has for other groups (Wanke et al. 2017, Buys et al. 2019). Finally, the monophyly of both of the major lineages within the Cyprinodontiformes, also was confirmed, however, alternative sets of familial level relationships within the suborders were recovered and are discussed below.

5.1. Suborder Cyprinodontoidei

The inferred relationships of several of the families within the Cyprinodontoidei support many of the classic relationships recovered in earlier studies (Parenti 1981, Costa 1998, 2012a, Hertwig 2008). The recent and extensive taxonomic rearrangements for several families and the diagnosis of several previously unrecognized family level diversity (Freyhof et al. 2017, Bragança et al., 2018) has resulted in the discovery of several novel sets of relationships, many of which provide insight into the evolutionary history of the superfamily and families within.

Clade I: Until recently, the family Poeciliidae was comprised of both Old World and New World representatives, each with distinctive viviparous and oviparous life-histories. The New World poeciliids (Poeciliidae) are distributed throughout North, Central, and South America and include more than 270 species of livebearing fishes. The single exception is *Tomeurus gracilis*, an oviparous species (Parenti et al. 2010). The results from our study support the distinctiveness of New World poeciliids and the recent supposition of Bragança et al. (2018) that Poeciliidae should be confined to the New World livebearers. Other recent multilocus phylogenetic studies (Pohl et al. 2015, Helmstetter et al. 2016, Reznick et al. 2017), also recover Old World and New World species of poeciliid fishes as paraphyletic lineages, further supporting the confinement of Poeciliidae to the New World. Within Poeciliidae, *Phallichthys* was not recovered as monophyletic. The lack of monophyly for *Phallichthys* is not surprising as *P. fairweatheri* was previously aligned with *Carlhubbsia*, based on similarities in gonopodial morphology (Hubbs 1936, Rosen and Bailey, 1959).

The Anablepidae was recovered as the sister family to Poeciliidae and this result has also been recovered in many other studies (Meyer and Lydeard 1993, Parker 1997, Pohl et al. 2015, Helmstetter et al. 2016, Reznick et al. 2017, Bragança et al., 2018, Amorin and Costa 2018). Both Anablepidae and Poeciliidae are sympatrically and often syntopically distributed throughout the New World in southern Mexico, Central America, and northern South America. Nearly all species of Poeciliidae

Table 2
Results of the approximately unbiased (AU) tests comparing alternative phylogenetic hypotheses of interfamily relationships within Cyprinodontiformes to interfamily relationships inferred from the concatenated RaxML phylogeny.

Trees	logLn	ΔlogLn	bp-RELL	*p*-AU
Phylogenetic placement of Fluviphylacidae				
{(Poeciliidae, Anablepidae) | Fluviphylacidae}	−3963767	0	0.02	0.726
{(Procatopodidae, Aphaniidae, Valenciidae, Orestiidae) | Fluviphylacidae}	−3963890	123.06	0	**2.95E-10**
{(Procatopodidae, Aphaniidae, Valenciidae) | Fluviphylacidae}	−3963891	124.03	0	**5.37E-09**
{(Procatopodidae, Aphaniidae, Valenciidae, Orestiidae, Poeciliidae, Anabelpidae) | Fluviphylacidae}	−3963949	182.5	0	**0.000215**
{(Procatopodidae, Aphaniidae, Valenciidae, Poeciliidae, Anablepidae) | Fluviphylacidae}	−3963951	184.39	0	**3.52E-06**
{(Procatopodidae, Fluviphylacidae}) | …}	−3966306	2539.7	0	**0.000833**
Monophyly and phylogenetic placement of Cubanichthyidae				
{(Cubanichthyidae) | …}	−3964033	266.58	0	**2.40E-05**
{(Cubanichthyidae) | (Cyprinodontidae)}	−3964033	266.41	0	**8.87E-05**
{(Cubanichthyidae, Cyprinodontidae) | …}	−3963767	0.008917	0.163	0.655
{(Cyprinodontoidei) | Cubanichthyidae, Pantanodontidae}	−3965499	1732.3	0	**3.24E-06**
Phylogenetic placement of Cubanichthyidae / Cyprinodontidae				
{(Cyprinodontoidei) | Cubanichthyidae, Cyprinodontidae, Pantanodontidae}	−3964252	485.52	0	**1.20E-38**
{(Cubanichthyidae, Cyprinodontidae) | (Fundulidae)}	−3963767	0.000853	0.198	0.71
{(Cubanichthyidae, Cyprinodontidae) | (Fundulidae, Goodeidae, Profundulidae)}	−3964002	235.09	0	0.15
Phylogenetic placement of Orestiidae				
{(Cyprinodontoidei) | Orestiidae}	−3963774	7.6003	0.006	**0.00696**
{(Cyprinodontoidei) | Orestiidae, Pantanodontidae}	−3963771	4.7993	0.044	0.181
{(Procatopodidae, Aphaniidae, Valenciidae) | Orestiidae}	−3963767	0.35852	0.164	0.416
{(Procatopodidae, Aphaniidae, Valenciidae, Fluviphylacidae, Poeciliidae, Anabelpidae) | Orestiidae}	−3963768	1.1351	0.109	0.314
{(Aphaniidae) | Orestiidae}	−3963768	1.6604	0.106	0.169
{(Orestiidae, Aphaniidae, Valenciidae) | …}	−3963767	0.005142	0.069	0.725
Monophyly of Procatopodidae and Poeciliidae				
{(Procatopodidae, Poeciliidae) | …}	−3969848	6081.5	0	**0.000112**

and many species of Anablepidae (*Anableps* and *Jenynsia* spp.) are viviparous (Blackburn 2015). Although this sister group relationship suggests a single origin of viviparity, with several reversals within these families, Helmstetter et al. (2016) indicated that the ancestral state of this same clade was an oviparous ancestor. More detailed taxonomic sampling is needed to comprehensively address this question.

The recently elevated family Fluviphylacidae (Bragança and Costa, 2018), an oviparous Amazonian group comprised of five miniature species (Costa 1996, Costa and Bail, 1999), was recovered as the sister taxon to Poeciliidae + Anablepidae. Previously, the taxonomic status and phylogenetic position of the fluviphylacids was uncertain, as morphological data sets aligned the genus with the group currently recognized as the Procatopodidae (Parenti 1981, Costa 1996, Ghedotti 2000). More recently, molecular data sets identified fluviphylacids as an independent lineage closely related to Poeciliidae and Anablepidae (Pollux et al. 2014; Pohl et al. 2015; Helmstetter et al. 2016; Reznick et al. 2017; Bragança and Costa, 2018). From a biogeographic perspective, the close relationship among Poeciliidae, Anablepidae, and Fluviphylacidae is geographically logical and indicates a potential South American origin for this clade as has previously been suggested by others (Reznick et al. 2017, Amorim and Costa, 2018).

Clade II: Procatopodidae was recently elevated to family rank (Bragança et al., 2018) and includes all Old World species (African Lampeyes) and genera formerly assigned to the Poeciliidae (Parenti 1981). The results from our study confirm this taxonomic recommendation. All of the nodes within the family were strongly supported, but it is difficult to say anything definitive about the genus level relationships due to limited taxon sampling within the family. Clearly, a comprehensive phylogenomic study targeted toward understanding the relationships among the genera and species within Procatopodidae is needed, and will likely result in substantial genus level taxonomic changes within the family as has been observed in recent molecular phylogenies (Bragança and Costa 2019) and taxonomic descriptions within the family (Van Der Zee et al. 2019, Bragança et al. 2020).

The phylogenetic position of Orestiidae has remained uncertain for some time. Both Parenti (1981) and Parenti (1984) suggested a close relationship between Orestiidae and Aphaniidae. Specifically, Parenti (1984) stated that orestiids were more closely related to Old World groups of cyprinodontiforms than to South America cyprinodontiforms. The supposition was supported by Parker and Kornfield (1995) who recovered a sister group relationship between genera of orestiids and aphaniids based on mtDNA sequences. Other studies that have included members of Orestiidae, however, have not supported these results (Hertwig 2008, Costa et al. 2012b, Helmstetter et al. 2016). Our results clearly support that of Parenti (1981) and Parenti (1984) and these results have significant zoogeographic ramifications for understanding the diversification of cyprinodontiform fishes.

Finally, the phylogenetic placement of Valenciidae has been enigmatic. Several morphologically based studies recovered Valenciidae is various phylogenetic positions based on morphology and these studies are not in mutual agreement (Parenti 1981, Hertwig 2008, Ghedotti and Davis 2013). Other studies based on genetic data have recovered Valenciidae in close association with Mediterranean cyprinidontiformes (Pohl et al. 2015, Helmstetter et al. 2016, Amorim and Costa, 2018, Bragança and Costa 2019), but most of these studies failed to include representatives of Orestiidae. The results from our AHE support Valenciidae as being closely aligned with Orestiidae and Aphaniidae.

Clade III: The sister group relationship between Goodeidae and Profundulidae has been recovered by Myer and Lydeard (1993), Costa (1998), and Helmstetter et al. (2016), but not by Parenti (1981). A plethora of new biodiversity has been discovered within the profundulids and goodeids over the last several years (Ornelas-García et al., 2015, Domínguez-domínguez et al., 2016, Matamoros et al. 2018) and additional new species are likely to be named as these families are currently under study by several researchers. The geographic distributions of the two families are adjacent to one another with the Goodeidae

distributed north of and within the Trans-Mexican Volcanic Belt (TMVB) in central Mexico and the Great Basin of the United States, and the Profundulidae occurring south of the TMVB in southern Mexico, Guatemala, El Salvador, and Honduras (Miller et al., 2006, Ornelas-García et al., 2015). The current distribution of these families suggests that the formation of the TMVB, which began 10–20 mya (Ferrari et al. 2012), likely played a substantial role of in the isolation and subsequent diversification of the two families. This geologic date overlaps with the cytochrome *b* molecular clock estimate of 10.5–16.5 mya for the separation of Goodeidae and Profundulidae (Doadrio and Domínguez-Domínguez, 2004), but younger than the date (33.11 mya) from Reznick et al. (2017).

Clade IV: Clade IV includes representatives of three families, but only two of which, were recovered as monophyletic with the AHE data. Cyprinodontidae, inclusive of six species and three genera in this study was recovered as the sister group to *Cubanichthys cubensis* (Cubanichthyidae), whereas another congeneric, *C. pengelleyi*, was recovered as sister to a Cyprinodontidae + *Cubanichthys cubensis* clade, rendering the family Cubanichthyidae as paraphyletic. *Cubanichthys* has rarely been included in phylogenetic analyses (Parenti 1981, Helmstetter et al. 2016) so its phylogenetic position has long been been enigmatic. Parenti (1981) recognized *Cubanichthys* as a genus within the subfamily Cubanichthyinae within Cyprinodontidae and sister to three other genera. A similar result was recovered by Pohl et al. (2015). Helmstetter et al. (2016) recovered *Cubanichthys* as the sister group to a large clade inclusive of Orestiidae, Poeciliidae, Procatopodidae, Aphanidae, Fluviphylacidae, Fundulidae, Profundulidae, Goodeidae, and Cyprinodontidae. Our results strongly place both species of *Cubanichthys* with Cyprinodontidae.

Although a large number of loci were recovered for the *Cubanichthys* (155 and 109, respectively) species in this study, this was far fewer than most other taxa in this study (x = 244.4 loci, 12–292). Despite this, these results do not support the recent recognition of Cubanichthyidae as a family distinct from Cyprinodontidae (*sensu* Freyhof et al. 2017) due to the paraphyly of *Cubanichthys*. This is supported by the results of the AU test, which rejects the alternative topological hypothesis of monophyly of Cubanichthyidae (*p* = 0.000024; Table 2). Indeed, less than half (48.43%) of gene trees recover monophyly of Cubanichthyidae (Table 2). Furthermore, Costa (2015) indicated that *Cyprinodon martae* (Cyprinodontidae), a species from Colombia that is only known from the holotype, is a taxonomically problematic species, whose phylogenetic affinities are unclear. It possesses jaw dentition features that unite it with others members of the Cyprinodontidae, but it also possess an enlarged supraoccipital crest, which suggests that it is likely a member of *Cubanichthys*. The results from our study, as well as Costa (2015), suggests that a re-examination of the taxonomic validity of Cubanichthyidae and *Cubanichthys* is warranted.

The *Cubanichthys*-Cyprinodontidae clade was sister to a well-supported and monophyletic Fundulidae. The sister group relationship between Fundulidae and Cyprinodontidae (excluding *Cubanichthys*) has been supported in other studies (Helmstetter et al. 2016), but not by others (Parenti 1981, Meyer and Lydeard 1993, Costa 1998, 2012b, Parker 1997, Pohl et al. 2015). Many of these studies recover a close relationship among fundulids, goodeids, and profundulids. In fact, the results of the AU topology tests highlight remaining uncertainty in the placement of the *Cubanichthys*-Cyprinodontidae clade as sister to either Fundulidae or to a larger clade inclusive of Fundulidae, Goodeidae, and Profundulidae. In addition, our study recovered *Fundulus* as monophyletic and sister to *Lucania*, a result previously supported by multiple studies (Ghedotti and Davis 2013, Rodgers et al., 2018), but not by others (Wiley 1986, Whitehead 2010).

Clade V: Multiple studies have recognized the distinctiveness and uniqueness of *Pantanodon* including Whitehead (1962), Pohl et al. (2015) and Betancur et al. (2017). Bragança et al. (2018) formally recognized Pantanodontidae as a family comprised of a single genus (*Pantanodon*) and two species (*Pantanodon stuhlmanni* and *P. madagascariensis*) based on a suite of eleven morphological apomorphies and its phylogenetic placement as the sister taxon to Cyprinodontidei. *Pantanodon* was previously aligned with the Aplocheilichthyinae, formerly of the family Poeciliidae (*sensu lato*) (Parenti 1981). Representatives of Pantanodontidae infrequently have been included in molecular phylogenetic studies as *P. stuhlmani* is rare in nature and *P. madagascariensis* is believed to be extinct (Sparks 2016). Pohl et al. (2015) used two mtDNA and three nDNA loci and also recovered *Pantanodon* as sister to the remainder of Cyprinodontodei, but with low posterior probabilities (PP = 65). Bragança et al. (2018) attributed the low topological support in Pohl et al. (2015), in part, to the incomplete data matrix for the species. Bragança et al. (2018), using five nuclear loci and one mtDNA locus, also recovered *Pantanodon* as sister to the remaining families in Cyprinodontoidei, but with strong likelihood (BS = 91) and Bayesian support (PP = 99). Alternatively, long-branch attraction cannot be ruled out as an explanation for its phylogenetic position. Morphologically, *Pantanodon* is a unique genus with many autapomorphies. Resolution of this issue is challenging as no other extant species of Pantanodon are known to exist. In our study, *Pantanodon* was also firmly positioned as the sister family to all families within Cyprinodontoidei.

5.2. Suborder Aplocheiloidei

Clade VI: Aplocheilidae, Nothobranchiidae, and Rivulidae were recovered as a monophyletic group and other studies also have supported the close relationship among them (Parenti 1981, Pohl et al. 2015, Helmstetter et al. 2016). The AHE data recovered a sister group relationship for the Old World families Aplocheilidae and Nothobranchiidae with strong support, with Rivulidae sister to the Aplocheilidae + Nothobranchiidae clade. This topology also has been recovered by other molecular and morphological data sets (Parenti 1981, Pohl et al. 2015, Helmstetter et al. 2016), however, other studies have provided support for a Nothobranchiidae and Rivulidae sister group relationship based on mtDNA (Murphy and Collier 1997), and morphology (Costa 2013).

5.3. Gene tree discordance and topological conflict

While genomic-scale datasets are expected to yield unprecedented resolution of historically challenging nodes in the Tree of Life, heterogeneity in the genealogical histories of hundreds to thousands of sampled loci may contribute evidence for conflicting topological hypotheses. In this study, investigations of genealogical discordance among the AHE loci sampled in our dataset generally reveal strong concordance among gene trees regarding the monophyly of most families within Cyprinodontoidea as well as many interfamily relationships that have been hypothesized on the basis of both morphological and molecular data (Figs. 10, 11, and 12; Table 1). However, estimation of concordance factors (CFs) and tests of alternative topological hypotheses uncover remaining conflict regarding the phylogenetic positions of *Cubanichthys*-Cyprinodontidae and Orestiidae (Figs. 10, 11, and 12; Table 2). Specifically, our analyses reveal uncertainty regarding the placement of *Cubanichthys*-Cyprinodontidae as sister to Fundulidae or to a larger clade inclusive of Fundulidae, Profundulidae, and Goodeidae (Figs. 1, 10, 11, and 12; Table 2). While CF estimation in BUCKy also identifies gene trees that recover *Cubanichthys*-Cyprinodontidae as sister to all other lineages within Cyprindontoidei lineages exclusive of Pantanodontidae (Figs. 11 and 12), this topological hypothesis is rejected by the AU test (*p* = 1.20E-38; Table 2). On the other hand, the AU test does not reject the hypothesized placement of Orestiidae as sister to all other lineages with Cyprinodontoidei exclusive of Pantanodontidae (*p* = 0.18; Table2), nor does it reject five other topological hypotheses that place Orestiidae as nested within or sister to a larger clade inclusive of Procatopodidae, Aphaniidae, Valenciidae, Poecilidae, Anablepidae, and Fluviphylacidae (Table 2).

Discordance among genealogical histories may be driven by mechanisms including incomplete lineage sorting (ILS), introgressive hybridization, low phylogenetic content of sampled loci, error in gene tree estimation, and the relative contributions of these different mechanisms may impact the reliability of species tree inference under concatenation or coalescent-based methods. In the case of Orestiidae, we feel that confident phylogenetic placement is precluded by a high amount of missing data for the single individual that represents this family in our dataset (Fig. 1). This is due, in part, to the fact that *Orestias* was not included in the probe design itself, but instead was mined for the taxon set. The *Orestias* sequences included in the taxon set are those sequences mined from the genome suggesting that the reason for the missing data is that the genome sequences did not contain many AHE loci resulting from poor quality genome assembly, or gene loss. Relative to the other two species for which AHE were mined from the genome *Orestias* has the lowest sequence recovery. This indicates that there were many orthosets missing entirely, suggesting that a poor genome quality is the primary cause of the missing data for *Orestias*.

The mechanism underlying discordance in the placement of *Cubanichthys*-Cyprinodontidae is unclear. If gene tree discordance is a primarily result of ILS, we expect that the random sorting of allelic variation into descendent lineages will produce roughly equal frequencies of gene trees that recover alternative sister relationships between *Cubanichthys*-Cyprinodontidae and other taxa of Cyprinodontoidei. Estimation of CFs reveals that gene trees recover *Cubanichthys*-Cyprinodontidae as sister to Fundulidae, to a clade inclusive of Fundulidae, Profundulidae, and Goodeidae, or to a large clade inclusive of all other taxa within Cyprinodontoidei exclusive of Pantanodontidae in relatively equal frequencies. However, gene trees do not recover sister relationships between *Cubanichthys*-Cyprinodontidae and any other Cyprinodontoidei, calling into question the role played by ILS in driving the observed discordance. We therefore cannot rule out the possibility that additional factors such as low phylogenetic information content also contribute to uncertainty in the phylogenetic placement of *Cubanichthys*-Cyprinodontidae.

Our findings underscore the fact that strong statistical support for nodes recovered in concatenation- or coalescent-based species trees may conceal extensive genealogical conflict in phylogenomic datasets (Brown and Thomson 2017). Interrogation of gene tree heterogeneity therefore represents an essential step in robust phylogenetic inference. We uncover phylogenetic uncertainty in the placements of *Cubanichthys*-Cyprinodontidae and Orestiidae that will likely impact downstream macroevolutionary studies of the Cyprinodontiformes. However, explicit investigation of genealogical heterogeneity in our dataset enables us to identify alternative hypotheses for phylogenetic relationships of these taxa, which can then be applied to test alternative scenarios of biogeographic history or patterns of trait evolution in future work. Furthermore, our investigation builds confidence in hypotheses concerning the monophyly of the families within the Cyprinodontiformes as well as concerning many interfamily relationships within the superfamily. As such, this work presents the most robust study of phylogenetic relationships within the Cyprinodontiformes to date, and will provide essential phylogenetic information for future studies of evolution and diversification of the clade.

6. Conclusions

Developing a robust historical template is an important first step in comparative biology. In the case of the Cyprinodontiformes, the Anchored Hybrid Enrichment approach (Lemmon et al. 2012) utilized in this study provided a high level of resolution at multiple hierarchical levels and shed light on the phylogenetic relationships of several recently diagnosed families in the superfamily. The results of this study are relevant because the study includes representatives of all of the currently known family level taxonomic diversity within the Cyprinodontiformes and a plethora of next-generation sequence data making

it the most robust molecular study of the group to date. In the future, an improved taxon-sampling scheme, particularly at the genus level within families, is needed to better understand the evolutionary history of this globally distributed group of fishes and allow future researchers to address large-scale macroevolutionary questions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107482.

References

Alda, F., Tagliacollo, V.A., Bernt, M.J., Waltz, B.T., Ludt, W.B., Faircloth, B.C., Alfaro, M. E., Albert, J.S., Chakrabarty, P., Near, T., 2019. Resolving deep nodes in an ancient radiation of Neotropical fishes in the presence of conflicting signals from incomplete lineage sorting. Syst. Biol. 68 (4), 573–593.

Amorim, P.F., Costa, W.J.E.M., 2018. Multigene phylogeny supports diversification of four-eyed fishes and one-sided livebearers (Cyprinodontiformes:Anablepidae) related to major South American geological events. PLoSONE 13 (6). https://doi.org/10.1371/journal.pone.0199201.

Ane, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24 (2), 412–426.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., Fay, J.C., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3 (10), e3376. https://doi.org/10.1371/journal.pone.0003376.

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., Ortí, G., 2013. The tree of life and a new classification of bony fishes. PLOS Currents Tree of Life. 2013 Apr 18 [last modified: 2013 Jun 3]. Edition 1. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., Ortí, G., 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17 (1) https://doi.org/10.1186/s12862-017-0958-3.

Blackburn, D.G., 2015. Evolution of vertebrate viviparity and specializations for fetal nutrition: A quantitative and qualitative analysis. J. Morphol. 276 (8), 961–990.

Bragança, P.H.N., Costa, W.J.E.M., 2018a. Time-calibrated molecular phylogeny reveals a Miocene-Pliocene diversification in the Amazon miniature killifish genus *Fluviphylax* (Cyprinodontiformes: Cyprinodontoidei). Org. Divers. Evol. 18 (3), 345–353.

Bragança, P.H.N., Amorim, P.F., Costa, W.J.E.M., 2018. Pantanodontidae (Teleostei, Cyprinodontiformes), the sister group to all other cyprinodontoid killifishes as inferred by molecular data. Zoosyst. Evol. 94, 137–145.

Bragança, P.H.N., Costa, W.J.E.M., 2019. Multigene fossil-calibrated analysis of the African lampeyes (Cyprinodontoidei: Procatopodidae) reveals an early Oligocene origin and Neogene diversification driven by palaeogeographic and palaeoclimatic events. Org. Divers. Evol. 19, 303–320.

Bragança, P.H.N., van Zeeventer, R.M., Bills, R., Tweddle, D., Chakona, A., 2020. Diversity of the southern Africa *Lacustricola* Myers, 1924 and redescription of

Lacustricola johnstoni (Günther, 1894) and *Lacustricola myaposae* (Boulenger, 1908) (Cyprinodontiformes, Procatopodidae). Zookeys 923, 91–113.

Brandley, M.C., Bragg, J.G., Singhal, S., Chapple, D.G., Jennings, C.K., Lemmon, A.R., Moriarty Lemmon, E., Thompson, M.B., Moritz, C., 2015. Evaluating the performance of anchored hybrid enrichment at the tips of the tree of life: a phylogenetic analysis of Australian *Eugongylus* group scincid lizards. BMC Evol. Biol. 15, 62.

Brown, A.P., McGowan, K.L., Schwarzkopf, E.J., Greenway, R., Rodriguez, L.A., Tobler, M., Kelley, J.L., 2019. Local ancestry analysis reveals genomic convergence in extremophile fishes. Philos. Trans. Roy. Soc. B 374 (1777), 20180240. https://doi.org/10.1098/rstb.2018.0240.

Brown, J.M., Thomson, R.C., 2017. Bayes Factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst. Biol. 66, 517–530.

Buys, M.H., Winkworth, R.C., de Lange, P.J., Wilson, P.G., Mitchell, N., Lemmon, A.R., Moriarty Lemmon, E., Holland, S., Cherry, J.R., Klápště, J., 2019. The phylogenomics of diversification on an island: applying anchored hybrid enrichment to New Zealand *Leptospermum scoparium* (Myrtaceae). Bot. J. Linn. Soc. 191 (1), 1–17.

Chakrabarty, P., Faircloth, B.C., Alda, F., Ludt, W.B., Mcmahan, C.D., Near, T.J., Dornburg, A., Albert, J.S., Arroyave, J., Stiassny, M.L.J., Sorenson, L., Alfaro, M.E., 2017. Phylogenomic systematics of Ostariophysan fishes: Ultraconserved Elements support the surprising non-monophyly of Characiformes. Syst. Biol. 66 (6), 881–895.

Chen, J.-M., Prendini, E., Wu, Y.-H., Zhang, B.-L., Suwannapoom, C., Chen, H.-M., Jin, J.-Q., Lemmon, E.M., Lemmon, A.R., Stuart, B.L., Raxworthy, C.J., Murphy, R.W., Yuan, Z.-Y., Che, J., 2020. An integrative phylogenomic approach illuminates the evolutionary history of Old World tree frogs (Anura: Rhacophoridae). Mol. Phylogenet. Evol. 145, 106724. https://doi.org/10.1016/j.ympev.2019.106724.

Costa, W.J.E.M., 1996. Relationships, monophyly and three new species of the Neotropical miniature poeciliid genus *Fluviphylax* (Cyprinodontiformes: Cyprinodontoidei). Ichthyol. Exploration Freshwars 7, 11–130.

Costa, W.J.E.M., 1998. Phylogeny and classification of the Cyprinodontiformes (Euteleostei, Atherinomorpha): A reappraisal. In: Malabarba, L.R., Reis, R.E., Vari, R. P., Lucena, Z.M., Lucena, C.A.S. (Eds.), Phylogeny and Classification of Neotropical Fishes. Porto Alegre, Edipucrs, pp. 537–560.

Costa, W.J.E.M., Bail, P.-Y., 1999. *Fluviphylax palikur:* A new poeciliid from the Rio Oiapoque Basin, northern Brazil (Cyprinodontiformes: Cyprinodontoidei), with comments on miniaturization in *Fluviphylax* and other Neotropical freshwater fishes. Copeia 1999 (4), 1027. https://doi.org/10.2307/1447977.

Costa, W.J.E.M., 2012a. The caudal skeleton of extant and fossil cyprinodontiform fishes (Teleostei: Atherinomorpha): comparative morphology and delimitation of phylogenetic characters. Vertebrate Zool. 52, 161–180.

Costa, W.J.E.M., 2012b. Oligocene killifishes (Teleostei:Cyprinodontiformes) from southern France: relationships, taxonomic position, and evidence of internal fertilization. Vertebrate Zool. 62, 371–386.

Costa, W.J.E.M., 2013. Historical biogeography of aplocheiloid killifishes (Teleostei: Cyprinodontiformes). Vertebrate Zool. 63, 139–154.

Costa, W.J.E.M., 2015. Phylogenetic position and tentative generic placement for *Cyprinodon martae* Steindachner, 1875 (Teleostei: Cyprinodontiformes), a killifish from northern Colombia. Vertebrate Zool. 65, 27–30.

Culumber, Z.W., Shepard, D.B., Coleman, S.W., Rosenthal, G.G., Tobler, M., 2012. Physiological adaptation along environmental gradients and replicated hybrid zone structure in swordtails (Teleostei: *Xiphophorus*). J. Evol. Biol. 25, 1800–1814.

Degnan, J.H., Rosenberg, N.A., Wakeley, J., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2 (5), e68. https://doi.org/10.1371/journal.pgen.0020068.

Doadrio, I., Domínguez-Domínguez, O., 2004. Phylogenetic relationship within the fish family Goodeidae based on cytochrome b sequence data. Mol. Phylogenet. Evol. 31, 416–430.

Domínguez-domínguez, OMAR, Bernal-zuñiga, D.MARÍA., Piller, K.R., 2016. Two new species of the genus *Xenotoca* Hubbs and Turner 1939 (Teleostei: Goodeidae) from central-western Mexico. Zootaxa 4189 (1), 81. https://doi.org/10.11646/zootaxa.4189.110.11646/zootaxa.4189.1.3.

Dornburg, A., Townsend, J.P., Brooks, W., Spriggs, E., Eytan, R.I., Moore, J.A., Wainwright, P.C., Lemmon, A., Lemmon, E.M., Near, T.J., 2017. New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. Mol. Phylogenet. Evol. 110, 27–38.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7 (1), 214. https://doi.org/10.1186/1471-2148-7-214.

Eytan, R.I., Evans, B.R., Dornburg, A., Lemmon, A.R., Moriarty Lemmon, E., Wainwright, P.C., Near, T.J., 2015. Are 100 enough? Inferring acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. BMC Evol. Biol. 15, 113.

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. Syst. Biol. 61 (5), 717–726.

Ferrari, L., Orozco-Esquivel, T., Manea, V., Manea, M., 2012. The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. Tectonophysics 522–523, 122–149. https://doi.org/10.1038/ncomms11271.

Freyhof, J., Ozulug, M., Gülşah, S., 2017. Neotype designation of *Aphanius iconii*, first reviser action to stabilise the usage of *A. fontinalis* and *A. meridionalis* and comments on the family group names of fishes placed in Cyprinodontidae (Teleostei: Cyprinodontiformes). Zootaxa 4294, 573–1558.

Garamszegi, L.Z., 2014. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer, Heidelberg, London.

Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol. Phylogenet. Evol. 80, 231–266.

Ghedotti, M.J., 2000. Phylogenetic analysis and taxonomy of the poecilioid fishes (Teleostei: Cyprinodontiformes). Zool. J. Linn. Soc. 130, 1–53.

Ghedotti, M.J., Davis, M.P., 2013. Phylogeny, classification, and evolution of salinity tolerance of the North American Topminnows and Killifishes, family Fundulidae (Teleostei: Cyprinodontiformes). Fieldiana Life Earth Sci. 7, 1–65.

Graybeal, A., Cannatella, D., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47 (1), 9–17.

Hamilton, C.A., Lemmon, A.R., Lemmon, E.M., Bond, J.M., 2016. Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. BMC Evol Biol 16, 212. https://doi.org/10.1186/s12862-016-0769-y.

Heath, T.A., Hedtke, S.M., Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. System. Evolut. 46, 239–257.

Hedtke, S.M., Townsend, T.M., Hillis, D.M., Collins, T., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55 (3), 522–529.

Helmstetter, A.J., Papadopulos, A.S.T., Igea, J., Van Dooren, T.J.M., Leroi, A.M., Savolainen, V., 2016. Viviparity stimulates diversification in an order of fish. Nat. Commun. 7, 11271.

Hertwig, S.T., 2008. Phylogeny of the Cyprinodontiformes (Teleostei: Atherinimorpha): The contribution of cranial soft tissue characters. Zoolog. Scr. 37, 141–174.

Hillis, D.M., 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47 (1), 3–8.

Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52 (1), 124–126.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35 (2), 518–522.

Hubbs, C.L., 1936. XVII. Fishes of the Yucatan Peninsula. Carnegie Inst. Wash. Publ. 457, 157–287.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589.

Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31 (2), 151–160.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. System. Biol. 56, 17–24.

Kuhner, M.K., Yamato, J., 2015. Practical performance of tree comparison metrics. Systematic Biology 64:205–214.Langerhans, R.B., M. E. Gifford, and E. O. Joseph. 2007. Ecological speciation in *Gambusia* fishes. Evolution 64 (2), 205–214.

Larget, B.R., Kotha, S.K., Dewey, C.N., Ané, C., 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910–2911.

Lemmon, A.R., Emme, S.A., Moriarty-Lemmon, E., 2012. Anchored hybrid enrichment for massively high-throughput phylogenies. Syst. Biol. 61, 727–744.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46 (3), 523–536.

Matamoros, W.A., Domínguez-Cisneros, S.E., Velázquez-Velázquez, E., McMahan, C.D., 2018. Description of a New Species of Killifish of the Genus *Profundulus* (Atherinomorpha: Profundulidae) from the Mexican State of Oaxaca. Copeia 106 (2), 239–246.

Meyer, A., Lydeard, C., 1993. The evolution of copulatory organs, internal fertilization, placentae and viviparity in killifishes (Cyprinodontiformes) inferred from a DNA phylogeny of the tyrosine kinase gene X-src. Proc. Roy. Soc. Lond. B 254, 153–162.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010. doi: 10.1101/pdb.prot5448.

Miller, R.R., Minckley, W.L., Norris, S.M., 2006. Freshwater Fishes of Mexico. University of Chicago Press, Chicago, IL.

Minh, B.Q., Minh, A., Nguyen, T., von Haeseler, A., 2013. Ultrafast Approximation for Phylogenetic Bootstrap. Mol. Biol. Evol. 30, 1188–1195.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30 (17), i541–i548.

Mirarab, S., Warnow, T., 2015. ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31 (12), i44–i52.

Murphy, W.J., Collier, G.E., 1997. A molecular phylogeny for aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of annualism. Mol. Biol. Evol. 14 (8), 790–799.

Near, T.J., Eytan, R.I., Dornburg, A., Kuhn, K.L., Moore, J.A., Davis, M.P., Wainwright, P. C., Friedman, M., Smith, W.L., 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Natl. Acad. Sci. USA 109 (34), 13698–13703.

Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. Fishes of the World, 5th Ed. John Wiley and Sons, Hoboken, NJ.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268–274.

Ornelas-García, C.P., Martínez-Ramírez, E., Doadrio, I., 2015. A new species of killifish of the family Profundulidae from the highlands of the Mixteca region, Mexico. Revista Mexicana de Biodiversidad 86 (4), 926–933.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Paradis, E., Schliep, K., 2019. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528.

Parenti, L.R., 1981. A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). Bull. Am. Museum Nat. History 168, 335–557.

Parenti, L.R., 1984. A taxonomic revision of the Andean killifish genus *Orestias* (Cyprinodontiformes, Cyprinodontidae). Bull. Am. Museum Nat. History 178, 107–214.

Parenti, L.R., LoNostro, F.L., Grier, H.J., 2010. Reproductive histology of *Tomeurus gracilis* Eigenmann, 1909 (Teleostei: Atherinomorpha: Poeciliidae) with comments on evolution of viviparity in atherinomorph fishes. J. Morphol. 271 (11), 1399–1406.

Parker, A., 1997. Combining molecular and morphological data in fish systematics: Examples from the Cyprinodontiformes. In: Kocher, T.D., Stepien, C.A. (Eds.), Molecular Systematics of Fishes. Academic Press, San Diego, CA, pp. 163–188.

Parker, A., Kornfield, I., 1995. Molecular perspective on evolution and zoogeography of cyprinodontid killifishes (Teleostei: Atherinomorpha). Copeia 1995 (1), 8. https://doi.org/10.2307/1446795.

Pfeiffer, J.M., Breinholt, J.W., Page, L.M., 2019. Unioverse: a phylogenetic resource for reconstructing the evolution of freshwater mussels (Bivalvia, Unionoida). Mol. Phylogenet. Evol. 137, 114–126.

Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., Penny, D., 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. PLoS Biol 9 (3), e1000602. https://doi.org/10.1371/journal.pbio.1000602.

Pie, M.R., Bornschein, M.R., Ribeiro, L.F., Faircloth, B.C., McCormack, J.E., 2019. Phylogenomic species delimitation in microendemic frogs of the Brazilian Atlantic Forest. Mol. Phylogenet. Evol. 141, 106627 https://doi.org/10.1016/j.ympev.2019.106627. Epub 2019 Sep 17 PMID: 31539606.

Pohl, M., Milvertz, F.C., Meyer, A., Vences, M., 2015. Multigene phylogeny of cyprinodontiform fishes suggests continental radiations and a rogue taxon position of *Pantanodon* Vertebrate. Zoology 65, 37–44.

Pollux, B.J.A., Meredith, R.W., Springer, M.S., Garland, T., Reznick, D.N., 2014. The evolution of the placenta drives a shift in sexual selection in livebearing fish. Nature 513 (7517), 233–236.

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 (7574), 569–573.

Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R., Burbrink, F.T., 2014. Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). Mol. Phylogenet. Evol. 81, 221–231.

Rabosky, D.L., 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS ONE 9, e89543.

Rannala, B., Huelsenbeck, J.P., Yang, Z., Nielsen, R., Cannatella, D., 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47 (4), 702–710.

Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L., Harshman, J., Huddleston, C.J., Kingston, S., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Witt, C.C., Yuri, T., Braun, E.L., 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the Avian Tree of Life more than taxon sampling. Syst. Biol. 66 (5), 857–879.

Reznick, D.N., Furness, A.I., Meredith, R.W., Springer, M., 2017. The origin and biogeographic diversification of fishes in the family Poecliidae. PlosOne 12, e0172546. https://doi.org/10.1371/journal.pone.0172546.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53 (1-2), 131–147.

Rodgers, R., Roach, J.L., Reid, N.M., Whitehead, A., Duvernell, D.D., 2018. Phylogenomic analysis of Fundulidae (Teleostei: Cyprinodotiformes) using RNA-sequencing data. Mol. Phylogenet. Evol. 121, 150–157.

Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature 425 (6960), 798–804.

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 (1), 312. https://doi.org/10.1186/1471-2164-13-312.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, 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. Syst. Biol. 61, 539–542.

Rosen, D.E., Bailey, R.M., 1959. Middle-American poeciliid fishes of the genera *Carlhubbsia* and *Phallichthys* with descriptions of two new species. Zoologica 44, 1–44.

Rosen, D.E., Parenti, L.R., 1981. Relationships of *Oryzias*, and the groups of Atherinomorph fishes. Am. Mus. Novit. 2719, 1–25.

Rosenberg, M.S., Kumar, S., 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. 98 (19), 10751–10756.

Rosenberg, M.S., Kumar, S., 2003. Taxon sampling, bioinformatics and phylogenomics. Syst. Biol. 52 (1), 119–124.

Ruane, S., Raxworthy, C.J., Lemmon, A.R., Moriarty Lemmon, E., Burbrink, F.T., 2015. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15, 221.

Schliep, K.P., 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27 (4), 592–593.

Shimodaira, H., Goldman, N., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51 (3), 492–508.

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W., Giribet, G., Dunn, C.W., 2011. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature 480 (7377), 364–367.

Sparks, J.S., 2016. Pantanodon madagascariensis. IUCN Red List of Threatened Species. 2016: e.T15949A58297104. doi:10.2305/IUCN.UK.2016-3.RLTS.T15949A58297104.en.

Stout, C.C., Tan, M., Lemmon, A.R., Moriarty Lemmon, E., Armbruster, J.W., 2016. Resolving Cypriniformes relationships using an anchored enrichment approach. BMC Evol. Biol. 16, 244. https://doi.org/10.1186/s12862-016-0819-5.

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. 2015 https://doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be. Tree of Life 1.

Van Der Zee, J.R., Bernotas, K., Bragança, P.H.N., Stiassny, M.L.J., 2019. An unexpected new *Poropanchax* (Cyprinodontiformes, Procatopodidae) from the Kongo Central Province, Democratic Republic of Congo. Am. Mus. Novit. 2019, 1–12.

Wanke, S., Granados Mendoza, C., Müller, S., Paizanni Guillén, A., Neinhuis, C., Lemmon, A.R., Lemmon, E.M., Samain, M.-S., 2017. Recalcitrant deep and shallow nodes in *Aristolochia* (Aristolochiaceae) illuminated using anchored hybrid enrichment. Mol. Phylogenet. Evol. 117, 111–123.

Whitehead, P.J.P., 1962. The Pantanodontinae, edentulous toothcarps from East Africa. Bull. Br. Museum Nat. History 9, 105–137.

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

Wiley, E.O., 1986. A study of the evolutionary relationships of *Fundulus* topminnows (Teleostei: Fundulidae). Am. Zool. 26 (1), 121–130.

Zhang, D., Rheindt, F.E., She, H., Cheng, Y., Song, G., Jia, C., Qu, Y., Alström, P., Lei, F., Beaulieu, J., 2021. Most genomic loci misrepresent the phylogeny of an Avian radiation because of ancient gene flow. Syst. Biol. 70 (5), 961–975. https://doi.org/10.1093/sysbio/syab024.

Zwickl, D.J., Hillis, D.M., Crandall, K., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51 (4), 588–598.

Further reading

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.

Machado, B.E., Podrabsky, J.E., 2007. Salinity tolerance in diapausing embryos of the annual killifsh *Austrofundulus limnaeus* is supported by exceptionally low water and ion permeability. J. Comp. Physiol. B. 177, 809–820.

MacGuigan, Daniel, Near, Thomas, 2019. Phylogenomic signatures of ancient introgression in a rogue lineage of darters (Teleostei: Percidae). Systematic Biology 329–346. https://doi.org/10.1093/sysbio/syy074.