Molecular Phylogenetics and Evolution 134 (2019) 282–290

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Assessing phylogenetic information to reveal uncertainty in historical data: An example using Goodeinae (Teleostei: Cyprinodontiformes: Goodeidae)

Elyse Parker[a,b,*], Alex Dornburg[c], Omar Domínguez-Domínguez[d], Kyle R. Piller[a]

[a] Southeastern Louisiana University, Dept. of Biological Sciences, Hammond, LA 70402, USA
[b] Yale University, Dept. of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA
[c] North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA
[d] Laboratorio de Biología Acuática, Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, Michoacán, Mexico

ARTICLE INFO

*Keywords:*
GC bias
Topological incongruence
Phylogenetic informativeness
Adaptive radiation
Middle American fishes

ABSTRACT

A major emerging challenge to resolution of a stable phylogenetic Tree of Life has been incongruent inference among studies. Given the increasing ubiquity of incongruent studies, analyzing the predicted phylogenetic utility and quantitative evidence regarding contributions toward resolution of commonly-used markers in historical studies over the last decade represents an important, yet neglected, component of phylogenetics. Here we examine the phylogenetic utility of two sets of commonly-used legacy markers for understanding the evolutionary relationships among goodeines, a group of viviparous freshwater fishes endemic to central Mexico. Our analyses reveal that the validity of existing inferences is compromised by both lack of information and substantially biased patterns of nucleotide substitution. Our analyses demonstrate that many of the evolutionary relationships of goodeines remain uncertain – despite over a century of work. Our results provide an updated baseline of critically needed areas of investigation for the group and underscore the importance of quantifying phylogenetic information content as a fundamental step towards eroding false confidence in results based on weak and biased evidence.

## 1. Introduction

The last several decades have yielded unparalleled progress towards the resolution of some of the most vexing problems across the Tree of Life (Qiu et al., 2006; Ebersberger et al., 2011; Romiguier et al., 2013; Misof et al., 2014; Prum et al., 2015; Alström et al., 2018; Streicher et al., 2018). Concomitantly, we are homing in on an increasingly stable framework for classifying and understanding the evolutionary relationships of earth's biota. However, despite massive innovations in both software and sequencing technology, some nodes continue to defy resolution (Regier et al., 2008; Dell'Ampio et al., 2013; Eytan et al., 2015; Brown and Thomson, 2016; King and Rokas, 2017). Furthermore, there is a trend of increasing numbers of studies reporting strongly supported relationships that are incongruent with previous phylogenetic hypotheses (Romiguier et al., 2013; Jarvis et al., 2014; Reddy et al., 2017; Simion et al., 2017). It has become readily apparent that heterogeneity in phylogenetic information content as well as other deviations from model assumptions (e.g. GC bias) often underlie incongruence among datasets (Betancur-R et al., 2013a; Cox et al., 2014; Dornburg et al., 2017a; Lammers et al., 2017; Reddy et al., 2017). The recognition that markers vary in their evolutionary dynamics, and therefore in their utility for specific phylogenetic problems, has driven the development of increasingly sophisticated approaches to phylogenetic experimental design (Philippe et al., 2011; Townsend et al., 2012; Su et al., 2014; Dornburg et al., 2017b; Shen et al., 2017). Although investigations of incongruence that account for expectations of phylogenetic experimental design have become common, scrutiny of the evidence for topological hypotheses in historical studies has been largely neglected.

Neglecting to scrutinize historical studies not only hinders our ability to evaluate hypothesis of diversification, but can inadvertently stymie the establishment of a stable taxonomic framework that reflects evolutionary history. For many clades, existing classification schemes are often based on markers that have not been profiled for phylogenetic information content, yet form the basis for studies integrating unsampled or extinct lineages into phylogenetic studies (Chatterjee et al., 2009; Smith et al., 2009; Jetz et al., 2012; Dornburg et al., 2017c; Economo et al., 2018). This lack of marker scrutiny creates the potential for false confidence in the accuracy of historical studies. Such false confidence can confound our understanding of phenotypic evolution

* Corresponding author at: Dept. of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA.
*E-mail address:* chantal.parker@yale.edu (E. Parker).

https://doi.org/10.1016/j.ympev.2019.01.025
Received 2 August 2018; Received in revised form 17 January 2019; Accepted 30 January 2019
Available online 04 February 2019
1055-7903/ 

and homology and fundamentally mislead our understanding of general features of macroevolution. Given the proliferation of methodologies that enable marker scrutiny (e.g.; Townsend, 2007; Townsend et al., 2012; Su et al., 2014), it is critical to evaluate the utility of existing datasets and quantify our confidence in evolutionary hypotheses and taxonomies derived from phylogenetic studies. Such efforts will allow us to disentangle historical inertia from evidence, thereby illuminating areas of confidence and uncertainty in the Tree of Life.

Inference of the evolutionary history of the Goodeidae (Teleostei: Cyprinodontiformes) remains a complex challenge in teleost systematics and therefore presents an opportunity to quantify the evidence supporting existing alternative phylogenetic hypotheses. Goodeids comprise 19 genera and 45 extant species of freshwater fishes currently classified into two subfamilies: Empetrichthyinae (Gilbert, 1893) and Goodeinae (Jordan, 1923). The Empetrichthyinae includes two genera and three species (Parenti, 1981; Minckley and Deacon, 1968; Grant and Riddle, 1995) which are restricted to pool and spring habitats in the Great Basin of the western United States (Minckley and Deacon, 1968; Soltz and Naiman, 1978). The remaining 17 genera and 42 species comprise the Goodeinae (Miller et al., 2005; Domínguez-Domínguez et al., 2008; Dominguez-Domínguez et al., 2016) and are endemic to the complex hydrological system of central Mexico (*sensu* Domínguez-Domínguez and Pérez-Ponce de León, 2009) and adjacent coastal basins (Meek, 1902, 1904; Hubbs, 1924, 1932; Miller et al., 2005; Domínguez-Domínguez et al., 2010).

A deep divergence between Goodeinae and Empetrichthyinae has been hypothesized based on aspects of reproductive biology (Parenti, 1981). Goodeine lineages exhibit viviparity, matrotrophy, and internal fertilization. In contrast, empetrichthyine lineages are oviparious, lecithotrophic, and external fertilizers. Species of Goodeinae have further been unified based on the following morphological features related to reproduction: (1) in males, a shortening and crowding of the first 6–7 anal rays, separated by a distinct notch from the rest of the anal fin; (2) the presence of an internal muscular organ in males which may function in sperm transfer; (3) a single median ovary in females; and (4) the presence of a trophotaeniae, a placenta-like rectal process that facilitates transfer of nutrients from the mother to the developing embryo (Meek, 1902, 1904; Hubbs & Turner, 1939; Parenti, 1981).

Relative to the Empetrichthyinae, goodeines are found in a diverse array of habitats, including fast-flowing streams, lakes, large rivers, warm and cool springs, and man-made ditches and canals. Corresponding with this shift in habitats, goodeines possess a wide range of body shapes, from the streamlined species of *Ilyodon*, which inhabit fast-flowing streams to the deep-bodied species of *Skiffia*, which are found primarily in lakes, ponds, and deep pools (Foster and Piller, 2018). Given their high ecomorphological disparity and their dominance of the ichthyofaunal diversity of central México, goodeines have been hypothesized to represent a central Mexican adaptive radiation (Meyer and Lydeard, 1993; Webb, 1998; Webb et al., 2004; Helmstetter et al., 2016). However, our understanding of the factors driving this radiation is challenged by the absence of a stable phylogenetic framework that accurately reflects the evolutionary history of the group. At present, the earliest divergences within the clade remain unresolved, and multiple morphological and molecular datasets support conflicting phylogenetic hypotheses (Fig. 1).

Hubbs and Turner (1939) proposed the first comprehensive classification of the goodeines, dividing the known diversity into four subfamilies, 18 genera, and 24 species based on anatomical variation in the ovary and the trophotaeniae. The subfamily Ataeniobiinae was considered to be the most 'primitive' of the goodeids and included only one species, *Ataeniobius toweri*, which was distinguished from all other species by an apparent lack of trophotaenial development in the embryonic form. Members of the subfamilies Girardinichthyinae and Goodeinae are distinguished from each other primarily by location of the ovigerous tissue in the ovary as well as the type of trophotaeniae exhibited by the embryonic forms. The subfamily Characodontinae was delimited to include only the genus *Characodon*, and was distinguished from the other genera and species by an ovary with characteristics intermediate between the Girardinichthyinae and Goodeinae types (Fig. 1). However, several subsequent studies over the next fifty years questioned the utility of characters related to reproduction in understanding phylogenetic relationships within the Goodeinae (Miller and Fitzsimons, 1971; Fitzsimons, 1972) citing that these characters have been shown to exhibit significant intraspecific variation (Mendoza, 1965). Smith (1980) reevaluated the classification of Hubbs and Turner (1939) based on osteological features of the mouth, dividing the goodeines into two major groups: the "*Goodea*" group, which was characterized by a protrusible premaxillae, and the "*Characodon*" group, which lacked a protrusible premaxillae (Fig. 1). The first attempt at classifying goodeine species diversity based on molecular data used a combined dataset of mitochondrial DNA (mtDNA) and allozymes and identified three major tribes within the Goodeinae: Chapalichthyni (comprising the genera *Alloophorus*, *Ameca*, *Zoogoneticus*, *Chapalichthys*, *Xenoophorus*, and *Xenotoca*), Ilyodontini (which united the genera *Ilyodon*, *Allodontichthys*, and *Xenotaenia*), and Girardinichthyini (initially including only *Girardinichthys*, *Allotoca*, and *Hubbsina*, but which later added the genus *Skiffia*) (Webb, 1998; Fig. 1). These, as well as two additional tribes (Characodontini, comprised of the genus *Characodon*, and Goodiini, uniting the species *Ataeniobius toweri* with the genus *Goodea*) were supported by several subsequent mtDNA-based phylogenetic analyses that included more comprehensive taxonomic sampling (Webb et al., 2004; Doadrio and Domínguez-Domínguez, 2004; Domínguez-Domínguez et al., 2010; Fig. 1). At present, this five-tribe scheme represents the most widely-accepted framework for classifying goodeine species diversity, but the predicted phylogenetic information of the markers used to generate this framework remains unclear. Does this classification accurately reflect goodeine evolutionary history?

Here we evaluate the evidence supporting the current phylogenetic framework that forms the basis for classifying and understanding the evolution of the Goodeinae. First, we evaluate the information content of previously used mitochondrial DNA sequence data. Next, we employ a multilocus nuclear DNA dataset, comprising six protein-coding genes and two introns, that has been fundamental to phylogenetic studies across the teleost Tree of Life (Near et al., 2012b; Betancur-R et al., 2013b; Near et al., 2013; Alfaro et al., 2018). We test for incongruence between these datasets using Bayesian and Maximum Likelihood-based methods and subsequently quantify the information content that gives rise to patterns of incongruence. Our analyses reveal a substantial lack of phylogenetic information for the problem of resolving the goodeines in both datasets, suggesting there is little evidence supporting the current classification of goodeine diversity. Expanding upon our results to investigate two other legacy datasets revealed similar pathologies. Our results suggest that further examination of historical studies on a case-by-case basis will be paramount to re-evaluating where we are confident in our resolution of the Actinopterygian Tree of Life.

## 2. Materials and methods

### 2.1. Taxon sampling

The majority of the specimens used in our analysis were field-collected across multiple basins spanning the Mesa Central of central Mexico between January 2005 and January 2015. Collection techniques included a combination of seining, dip-netting, and backpack electrofishing. The remaining specimens were obtained from captive stocks maintained by aquarists and hobbyists with known sources of origin. For all specimens, either muscle tissue or fin clip samples were taken and stored in 95% ethanol and were deposited in the Southeastern Louisiana University Tissue Collection (SLUTC). The taxonomic sample comprised a total of 51 individuals representing 35 species and all 18 genera of the subfamily. Each individual in this study

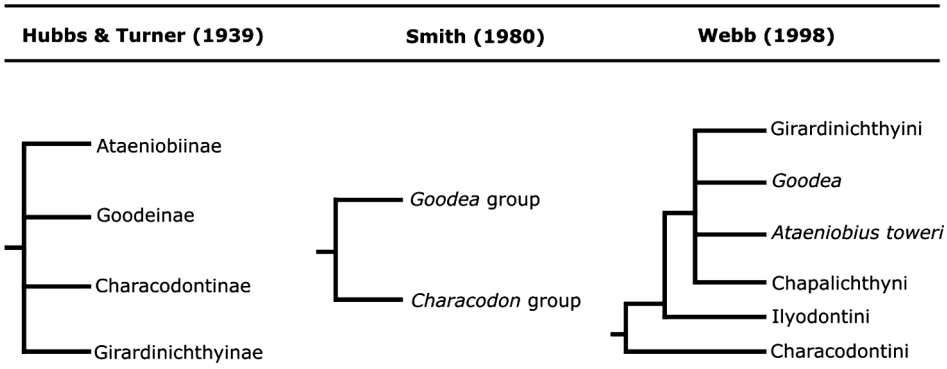


**Fig. 1.** Phylogenetic representation of alternative higher-level classification schemes for the subfamily Goodeinae. Note that the classification scheme labeled as Webb (1998) in this figure also reflects taxonomic revisions made after 1998 (see Webb et al., 2004; Doadrio and Domínguez-Domínguez, 2004).

represents a distinct evolutionarily significant unit (ESU; Moritz, 1994) that has been identified on the basis of morphological variation, genetic distinctiveness, or geographic separation for the purposes of conservation management (J. Lyons-Goodeid Working Group, pers. comm).

### 2.2. Molecular data collection

Whole genomic DNA was extracted from muscle biopsies and fin clips using the Qiagen DNeasy Tissue Extraction Kit following manufacturer protocol (Qiagen Inc., Valencia, CA). Amplification of the six nuclear protein-coding genes *ENC* subunit 1 (613 bp), *myh6* (679 bp), *plagl2* (653 bp), *SH3PX3* (661 bp), *sreb2* (905 bp), and *zic1* (645 bp), the two introns *S7* intron 1 (*S7-1*, 729 bp) and *PolB* (516 bp), and the mitochondrial gene cytochrome *b* (*cyt b*; 1094 bp) was carried out using polymerase chain reaction (PCR) in 25 μL reactions consisting of: 0.75 μL MgCl; 2.5 μL 10X Buffer; 0.5 μL dNTPs; 0.5 μL of each primer; 0.25 μL Taq; 1 μL DNA template; 19 μL water. Previously designed primers were used to amplify all loci (Li et al., 2007; Chow and Hazama, 1998; Kang et al., 2013). Amplification of each of the six nuclear loci was achieved using a nested PCR protocol consisting of two rounds of PCR. Detailed information about PCR cycling conditions for all loci are given in Supplementary Table 2. Because taxonomic sampling for the *cyt b* locus was relatively limited, the mitochondrial dataset was supplemented with *cyt b* data downloaded from GenBank (accession numbers AF510748 – AF510845), which were derived from Doadrio and Domínguez-Domínguez (2004). All PCR products were electrophoresed on 0.8% agarose gel to detect for presence, size, and quality of the amplified fragments, and all unpurified PCR products were sent to an external facility for sequencing (GENEWIZ, Cambridge, MA). Raw sequence data were edited and aligned using the MUSCLE algorithm (Edgar, 2004) as implemented in the program Geneious v. 7.0.6 (Kearse et al., 2012). All sequence data was deposited in the GenBank sequence repository (accession numbers MK522810-523234).

### 2.3. Phylogenetic analysis

Phylogenetic relationships within the subfamily Goodeinae were inferred using maximum likelihood and Bayesian inference methods. Prior to phylogenetic inference, the program PartitionFinder v1.1.1 (Lanfear et al., 2012) was used to determine the best-fit partition scheme and corresponding models of nucleotide substitution for each individual gene alignment and for the multilocus nuclear dataset (Table 1).

The candidate pool of partitions included every combination of mitochondrial and nuclear DNA codon positions, with the exception of the two introns *S7-1* and *PolB*, which were not partitioned. Individual gene trees for each locus were inferred under a maximum-likelihood framework in the program RAxML v7.2.6 (Stamatakis, 2014) using the appropriate BIC-selected models of nucleotide substitution. The concatenated nuclear dataset was analyzed in RAxML under the GTR + G substitution model. Topological support for the nodes of all individual gene trees and the concatenated tree were assessed using a thorough bootstrap analysis with 10 runs and 1000 replicates each.

**Table 1**
Bayesian Information Criterion (BIC)-selected nucleotide substitution models and partition schemes identified by PartitionFinder. For each of the dataset partitions, the number following the underscore represents the codon position of the gene that is included in the data partition.

| Dataset | Substitution Model | Dataset Partition |
|---|---|---|
| **Maximum likelihood** | | |
| Multilocus nuclear dataset | GTR + G | None |
| Mitochondrial dataset (*cyt b*) | GTR + G | None |
| **Bayesian inference** | | |
| *Multilocus nuclear dataset* | | |
| Subset 1 | K80 + G | *ENC*_1 |
| Subset 2 | GTR + I + G | *ENC*_2, *SH3PX3*_1, *myh6*_1, *plagl2*_1, *sreb2*_1, *zic1*_1 |
| Subset 3 | GTR + G | *ENC*_3, *SH3PX3*_2, *myh6*_2, *plagl2*_2 |
| Subset 4 | GTR + G | *SH3PX3*_3, *myh6*_3, *plagl2*_3, *sreb2*_3, *zic1*_3 |
| Subset 5 | HKY + I | *PolB* |
| Subset 6 | HKY + G | *S7* |
| Subset 7 | F81 + I | *sreb2*_2 |
| Subset 8 | JC | *zic1*_2 |
| *Mitochondrial dataset (cyt b)* | | |
| Subset 1 (first codon) | K80 + I + G | *cytb*_1 |
| Subset 2 (second codon) | HKY + G | *cytb*_2 |
| Subset 3 (third codon) | GTR + I + G | *cytb*_3 |

All of the above analyses were repeated in a Bayesian framework using MrBayes v3.2.6 (Ronquist et al., 2012). The joint posterior probabilities of tree topologies, branch lengths, and other parameters were estimated using two runs of Markov Chain Monte Carlo (MCMC) analysis, each with one cold chain and three heated chains. For each analysis, chains were run between 1 and 20 million generations, with sampling of the parameter states every 100 generations and with the first 25% of these samples discarded as burn-in. Visualization of the state likelihoods, potential scale reduction factors, and average deviation of split frequencies were used to diagnose convergence between independent MCMC runs. The post-burn-in distributions of the two runs were combined and used to construct a 50% majority rule consensus tree.

### 2.4. Quantifying phylogenetic information

For each locus, we quantified site-specific evolutionary rates using HyPhy (Pond and Muse, 2005) implemented in the PhyDesign web

interface (Lopez-Giraldez and Townsend, 2011). We estimated guide chronograms based on both the multilocus nuclear and mitochondrial Bayesian trees in conjunction with the nonparametric rate-smoothing algorithm in APE (Paradis et al., 2004). Recent studies have found that quantification of phylogenetic information content is robust to the choice of guide tree as the resulting site rate estimates are correlated under different tree topologies (Dornburg et al., 2017a). Quantifications of phylogenetic utility depend both on temporal depth of the tree and internode length, as such we used the quartet representing the most recent common ancestor of *Ataeniobius toweri* from other goodeines. This node represents a deep divergence in the goodeid Tree of Life that is an exemplar of the phylogenetic problems potentially underlying an unstable tree topology and corresponding taxonomy. As the nuclear and mitochondrial gene trees conflicted regarding the placement of this taxon, analyses were conducted under guide trees reflecting both topological hypotheses. We used the program PhyInformR (Dornburg et al., 2016) to quantify the predicted probabilities of correctly resolving (QIRP) or incorrectly resolving this quartet due to homoplasy (QIHP). Additionally, we tracked the predicted probability of loci contributing no phylogenetic information content for resolution thereby yielding a polytomy (QIPP). It is important to note that these approaches do not make any assumption regarding what the actual tree topology is for a focal group. These calculations are based on an s-state Poisson model that uses estimated evolutionary rates of character change and a character state space to provide a probabilistic prediction of whether convergence or parallelism will mislead resolution of a hypothetical phylogenetic quartet of depth T and internode t. Therefore, the calculation of what is "correct" or "incorrect" is agnostic to the empirical tree and simply reflects the expectations of any similar theoretical quartet. Information content was quantified for each gene's codon position independently as well as for the two introns. Additionally, as GC bias is also known to compromise inference, base compositional frequencies were also evaluated. In cases where elevated GC content was found, phylogenetic analyses were repeated using RY coding.

To test the generality of these results, we additionally downloaded sequence data from two independent datasets. These datasets were used in previous investigations of similar aged radiations of balistoid fishes (triggerfishes and filefishes) (Dornburg et al., 2011, 2008; Santini et al., 2013a) and Antarctic notothenioids (Dornburg et al., 2017d; Near et al., 2012a). Site rates and QIHP, QIPP, and QIRP values were estimated using the above protocols in combination with publicly available tree topologies that were used as guide trees.

## 3. Results

### 3.1. Phylogenetic analysis

PartitionFinder identified 8 partitions for the nuclear gene datasets and 3 partitions for cytb that correspond to differences between codon positions (Table 1). For each dataset, topologies recovered by Maximum Likelihood and Bayesian Inference methods were concordant (Fig. 2b and c). Bayesian inference of both the nuclear and the mitochondrial datasets strongly support monophyly of the family Goodeidae (Bayesian

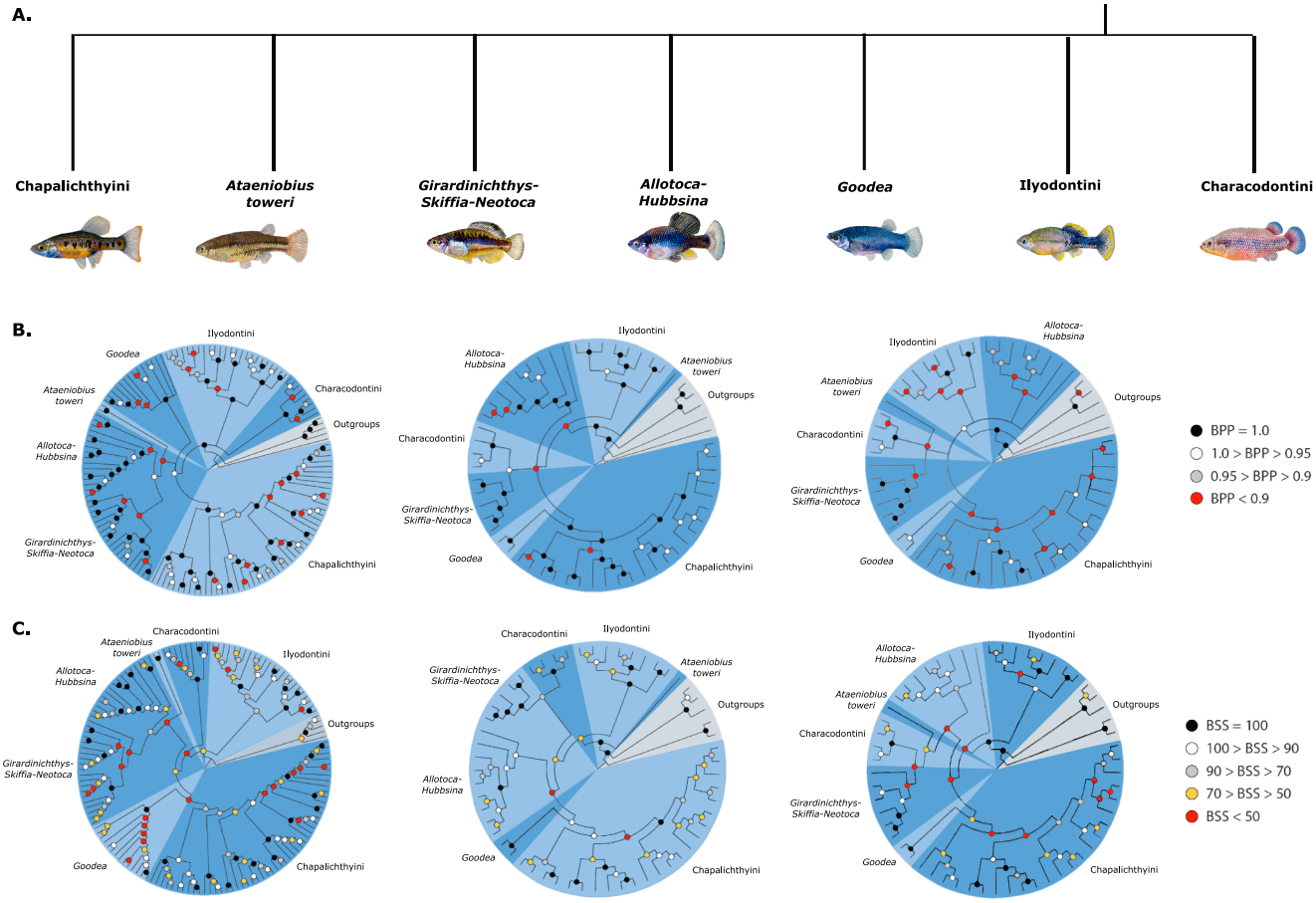


**Fig. 2.** Bayesian and maximum likelihood inferences of goodeine relationships. (A) A representation of uncertainty of higher level relationships indicating lack of resolution of major lineage relationships but congruence in identified clades between analyses based on (B) maximum likelihood or (C) Bayesian inference of (left) the mitochondrial locus *cytochrome b*, (middle) a multilocus nuclear dataset comprised of six exons and two introns, and (right) the multilocus nuclear dataset with third codon positions of all exons RY-coded. Colored slices correspond to major clades discussed in the text, and support values are given for each node. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Posterior probability [BPP] = 1.0) and of each of the currently-recognized genera (BPP ≥ 0.95 for each genus) with one exception: the genus *Xenotoca*. Analyses of both the nuclear and mitochondrial datasets strongly support monophyly of the tribes Ilyodontini (BPP = 1.0), Chapalichthyni (BPP = 1.0), and Characodontini (BPP = 1.0). Neither dataset recovers a monophyletic Goodiini, which includes *Goodea* and *Ataeniobius toweri*. The nuclear dataset strongly supports a sister relationship between *Goodea* and Chapalichthyni (BPP = 1.0), while *A. toweri* is recovered with weak support as the sister lineage to a clade containing all other goodeids (BPP = 0.869). Alternatively, the mitochondrial dataset recovers a weakly supported sister relationship between Girardinichthyini and *A. toweri* (BPP = 0.635), while the relationships among this clade, Chapalichthyni, and *Goodea* are not resolved.

The nuclear and mitochondrial phylogenies strongly conflict with regard to the monophyly of the tribe Girardinichthyini. The nuclear dataset supports monophyly of each of the constituent polytypic genera (*Allotoca*, *Girardinichthys*, and *Skiffia*) and recovers a sister relationship between *Girardinichthys* and *Skiffia* as well as a sister relationship between *Allotoca* and *Hubbsina turneri*. However, phylogenetic inference does not recover support for a sister relationship between the *Girardinichthys* + *Skiffia* + *Neotoca* clade and the *Allotoca* clade. Instead, the *Girardinichthys* + *Skiffia* + *Neotoca* clade is recovered with strong support as the sister group to Characodontini (BPP = 0.997). Relationships among the *Girardinichthys* + *Skiffia* + *Neotoca* + Characodontini clade, the *Allotoca* clade, and the Chapalichthyni + *Goodea* clade are not resolved. Phylogenetic inference of the mitochondrial dataset recovers monophyly of Girardinichthyini (BPP = 0.996), but monophyly of each of the clades *Allotoca* and *Girardinichthys* + *Skiffia* + *Neotoca* is not strongly supported (BPP = 0.897 and BPP = 0.811, respectively).

The relationships among the major goodeine lineages remain largely unresolved based on both mitochondrial and nuclear data. The nuclear-based phylogeny recovers *Ataeniobius toweri* as the sister lineage to a clade containing all other goodeine species, but monophyly of the clade exclusive of *A. toweri* is not strongly supported (BPP = 0.869). The next lineage-splitting event occurs between Ilyodontini and a clade containing all remaining goodeine lineages, but again monophyly of this clade is not strongly supported (BPP = 0.826). The nuclear data strongly support a sister relationship between *Goodea* and Chapalichthyni (BPP = 1.0) as well as a sister relationship between the *Girardinichthys* + *Skiffia* + *Neotoca* clade and Characodontini (genus *Characodon*) (BPP = 0.997), but the relationships among these clades and *Allotoca* are not resolved. The mitochondrial-based phylogeny recovers three major lineages within the Goodeinae: Characodontini, Ilyodontini, and a clade containing all remaining goodeids. The relationships among these three major lineages are represented by a polytomy in the phylogeny. The clade containing Chapalichthyni, a monophyletic Girardinichthyini with its sister lineage *Ataeniobius toweri* (BPP = 0.635), and *Goodea*, is strongly supported as monophyletic (BPP = 0.995), but the relationships among the constituent lineages are not resolved.

### 3.2. Quantifying phylogenetic information content

Results between both guide trees were equivalent, consistent with expectations that estimates of phylogenetic utility are robust to the choice of guide tree (Dornburg et al., 2017a). Quantification of phylogenetic utility revealed a substantial lack of information across all loci (Fig. 3). For each of the nuclear exons, both first and second codon positions were predicted to contribute to neither correct nor incorrect resolution of the quartet, instead possessing virtually no phylogenetic information (Fig. 3a). Perhaps more surprisingly, third positions were also information-deficient, exhibiting low probability of either correct or incorrect resolution of the quartet. Moreover, third positions were found to be heavily biased in GC content (Fig. 3b). Likewise, intronic regions were revealed to contain little phylogenetic information (Fig. 3a). The first and second positions of *cyt b* were predicted to

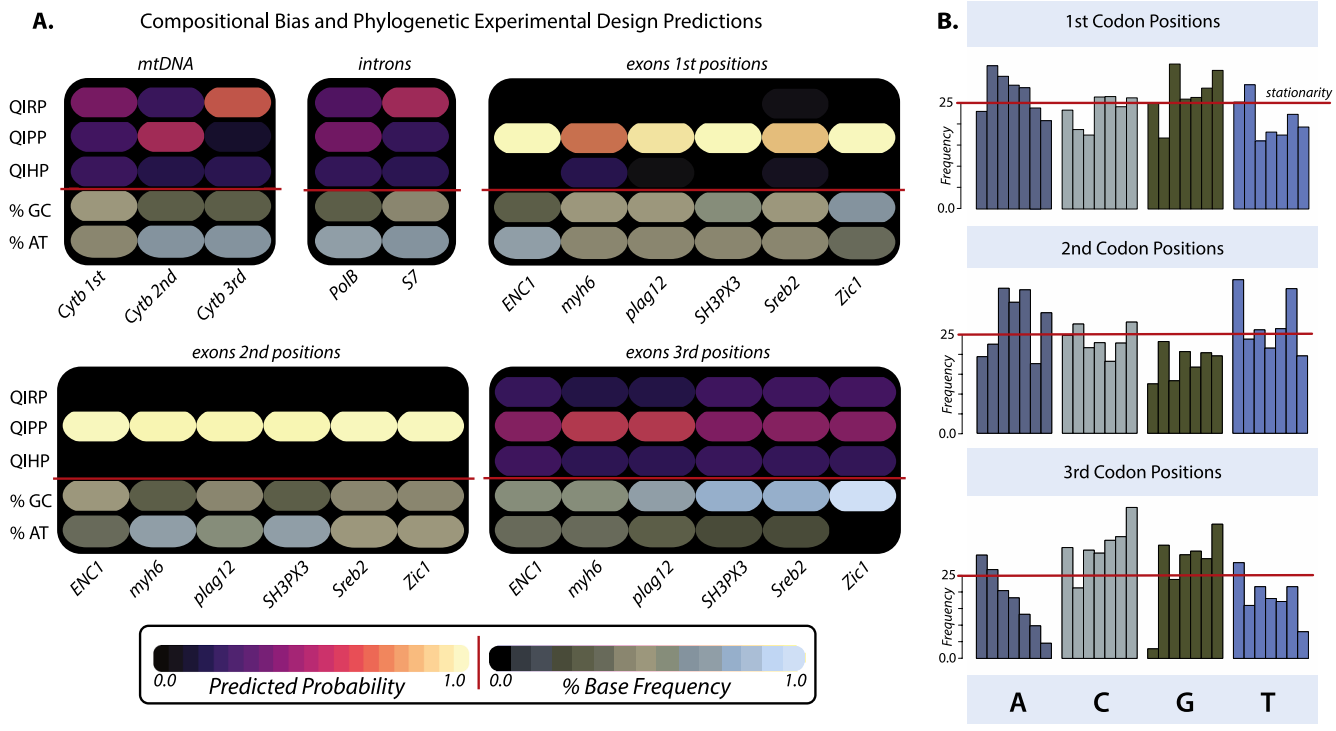


**Fig. 3.** Predicted phylogenetic utility and patterns of compositional bias across all markers. (A) Quartet internode resolution probabilities (QIRP), quartet internode polytomy probabilities (QIPP), and quartet internode homoplasy probabilities (QIHP) for each codon position of all protein coding genes (including six exons and the mitochondrial locus cytochrome *b*) as well as for both introns, contrasted with GC% and AT% for each locus. (B) Detailed base frequencies for each protein coding gene. Red line indicates stationarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

largely be uninformative (Fig. 3a). Third positions in the mitochondrial locus *cyt b* were predicted to contain the most information. However, in this locus – including third positions in codons – there was a bias toward loss of guanine content, indicating lack of stationarity of base composition. These results were consistent with our analyses of the two other legacy marker datasets. In both the cases of balistoids and notothenioids, our quantification of predicted phylogenetic information content revealed high values of QIPP or QIHP respectively (Supplementary Fig. 4). In total, all of these analyses suggest that these markers provide very little evidence for the current phylogenetic framework for goodeines.

## 4. Discussion

With their widespread distribution across central Mexico, high species diversity, and exceptional ecomorphological disparity, the goodeines are considered a model system for understanding both adaptive radiation and the biogeographic history of one of the most important faunal transitional zones in the world. However, despite over a century of study (Meek, 1902, 1904; Jordan, 1923; Hubbs and Turner, 1939; Turner, 1946; Miller and Fitzsimons, 1971; Parenti, 1981; Grudzien et al., 1992; Webb et al., 2004; Doadrio and Domínguez-Domínguez, 2004; Domínguez-Domínguez et al., 2006; Domínguez-Domínguez et al., 2010), the phylogenetic relationships among goodeine species remain uncertain. Our analyses demonstrate that this instability is a consequence of utilizing phylogenetic markers that contain little information for resolving the early divergences in the goodeine phylogeny, thereby obstructing the establishment of a robust phylogenetic framework that reflects the evolutionary history of the group. Given the widespread use of these markers for clades of similar ages (Dornburg et al., 2015; Friedman et al., 2013; Near et al., 2014, 2013; Santini et al., 2013b), our results suggest that evaluating whether there is evidence for competing topological hypotheses is a fundamental step for consistent phylogenetic inference.

### 4.1. Opening cold cases to end incongruence

The advent of molecular phylogenies has catalyzed a systematic renaissance. For decades, increasingly sophisticated sequencing and analysis methods have fundamentally rearranged our understanding of earth's biodiversity. Although many branches of the Tree of Life have long stabilized, some branches have remained neglected, while others have been characterized by competing topological hypotheses (Chakrabarty et al., 2017; Oscar et al., 2017; Tang et al., 2018; Bangs et al., 2018). For these problematic branches, evaluating the evidence that underlies existing hypotheses represents a fundamental step towards establishing how much support and confidence we have in our inferences. Explicit data scrutiny has the potential to reveal sources of error, including systematic bias and convergences in character that do not reflect evolutionary history yet can promote spurious inference (Jeffroy et al., 2006; Philippe et al., 2011; Klopfstein et al., 2017; Dornburg et al., 2018; Gilbert et al., 2018). Our analyses reveal that in the case of goodeines, there is little phylogenetic information contributing to the resolution of the earliest divergences within the group.

Principles of phylogenetic experimental design have been increasingly adopted in recent phylogenetic studies, with investigators seeking genes with strong phylogenetic signal and low levels of nucleotide saturation (Rokas and Chatzimanolis, 2008; Shen et al., 2017; Dornburg et al., 2017a). However, for studies based on legacy datasets, this principle of phylogenetic experimental design was not commonly adopted with the same markers being applied across studies spanning population-level divergences to the interrelationships of major vertebrate clades (Alfaro et al., 2009; Near et al., 2013, 2015; Dornburg et al., 2015). This practice may have been largely driven by primers for legacy markers being passed between laboratory groups, a practice once criticized as preferring 'empirical folklore' over experimental design principles (Goldman, 1998). Our findings underscore the need to consider historical data in light of experimental design predictions. Markers of high utility for resolving deep nodes in the teleost tree of Life may not necessarily be of high utility for resolving a recent rapid radiation. Instead, we find that first and second exonic positions contribute virtually no phylogenetic information, while only a handful of sites characterized by biases in base composition are contributing to resolution (Fig. 3a and b).

Biases in nucleotide acquisition have repeatedly been demonstrated to mislead phylogenetic inference, in the worst cases lending strong support and false confidence to incorrect inference (Betancur-R et al., 2013a; Cox et al., 2014; Dornburg et al., 2017a; Reddy et al., 2017). Our analyses reveal a tendency toward GC acquisition bias in the third codon positions of the nuclear exons we investigated. Given that there is little information present in the first and second positions, this suggests that exon positions characterized by bias are disproportionately influencing topological resolution based on the nuclear genes. Analyses of an RY coded nDNA exonic dataset provide support for this hypothesis, eroding support values for many previously supported relationships based on these exons (Fig. 2b and c). Additionally, scrutiny of the mitochondrial DNA revealed a loss of guanine (Fig. 3a and b), highlighting an additional nucleotide acquisition bias known to occur in protein-coding mitochondrial genes as a result of single-stranded replication (Hassanin et al., 2005). The bias we report here may not be restricted to goodeines as the loci evaluated here represent some of the most common markers that have been used for teleost systematics. Although resolution of many of the deep divergences in the teleost Tree of Life based on these markers have been found to be robust to the problems exhibited here (Alfaro et al., 2007; Santini et al., 2009; Near et al., 2012a; Price et al., 2014), this is not the case for all clades (Betancur-R et al., 2013a; Dornburg et al., 2017a). Additionally, there are numerous other commonly used markers for clades that span the Tree of Life that we do not evaluate, but that have also repeatedly been shown to exhibit clade-specific nucleotide acquisition biases (Romiguier et al., 2013; Reddy et al., 2017; Galtier et al., 2018). Given the ubiquity of bias across broad sets of markers, this suggests that careful scrutiny of historic studies represents an underappreciated axis from which to evaluate the support for incongruent inferences.

### 4.2. Evidence, uncertainty, and the evolutionary history of goodeines

Critical evaluation of the evidence supporting alternative phylogenetic frameworks for the Goodeinae provides a useful case study demonstrating the importance of data scrutiny. Prior to quantification of the information contained within each dataset, the mitochondrial and nuclear DNA-based topologies alone call into question the validity of two of the five currently-recognized tribes for organizing goodeid diversity (Fig. 2). Both datasets fail to recover a monophyletic Goodiini (Fig. 2), which was once proposed to unite *Goodea* with *Ataeniobius toweri* (Doadrio and Domínguez-Domínguez, 2004). Recent biogeographic analyses of the Goodeinae based on *cytochrome b* also failed to recover monophyly of Goodiini and questioned the designation of this taxonomic grouping as a naturally occurring clade (Domínguez-Domínguez et al., 2010). Given the lack of support for this clade in both our analyses and previous studies, the validity of Goodiini as a natural group remains uncertain. Likewise, validity of the tribe Girardinichthyini remains controversial, as the mitochondrial and nuclear topologies presented here strongly conflict regarding the phylogenetic placement of the *Girardinichthys* + *Skiffia* + *Neotoca* clade (Fig. 2). The remaining three tribes – Characodontini, Ilyodontini, and Chapalichthyni – are each strongly supported as monophyletic groups by phylogenetic analyses of both the mitochondrial and the nuclear datasets (Fig. 2).

However, quantification of the phylogenetic information content of both datasets reveals that the sites that exhibit the highest predictive probabilities for resolution are also those that exhibit biases in base

composition. Our analyses further demonstrate that these biases lend false confidence to resolution of the early evolutionary history of the Goodeinae (Fig. 2b and c) and, by extension, mislead the development of a taxonomic framework that accurately reflects the group's phylogeny. The effect of this bias becomes striking when we use RY coding to account for elevated GC content in the third codon position of the nuclear loci. In the phylogeny generated from this dataset, only two of the five previously proposed tribes are strongly supported: Ilyodontini, which comprises three genera restricted to the Pacific coastal drainages of central Mexico (Webb, 1998; Webb et al., 2004, Doadrio and Domínguez-Domínguez, 2004), and Characodontini, which is represented only by the genus *Characodon* (Doadrio and Domínguez-Domínguez, 2004). This result is in line with a growing number of studies finding biases in nucleotide frequency to mislead inference of various clades across the Tree of Life (Dornburg et al., 2017a; Bossert et al., 2017; Reddy et al., 2017), thereby undermining support for the hypothesis that goodeines are comprised of five major clades. This is not to say that these clades are not valid, however more work is needed to evaluate the validity of existing taxonomic designations.

Our finding that lack of evidence and biases in nucleotide composition undermine support for the prevailing phylogenetic hypotheses of Goodeinae has important implications for our understanding of the evolutionary history of the group. The high species richness, ecomorphological disparity, and trophic diversity exhibited by the Goodeinae relative to their sister lineage, the Empetrichthyinae, has been promoted as evidence that the group has undergone an adaptive radiation in central Mexico (ca. 9–14 Ma; Webb et al., 2004; Doadrio and Domínguez-Domínguez, 2004; Domínguez-Domínguez et al., 2010; Pérez-Rodríguez et al., 2015; Foster and Piller, 2018). It has been hypothesized that this radiation was driven in large part by repeated cycles of dispersal and subsequent allopatric speciation as a result of frequent and widespread volcanic and tectonic activity since at least the early Miocene (Doadrio and Domínguez-Domínguez, 2004; Webb et al., 2004; Domínguez-Domínguez et al., 2006; Domínguez-Domínguez et al., 2010). Furthermore, recent study has demonstrated that goodeines experienced a relatively high rate of body shape evolution, particularly in the trunk region, suggesting that diversification of the clade into a wide range of novel habitats in central Mexico represents a significant driver of the Goodeinae radiation (Foster and Piller, 2018). However, a lack of topological confidence precludes testing the extent to which various ecological or geological factors have impacted diversification rates, thereby obstructing our ability to harness the power of this group as a basis for furthering adaptive radiation theory and testing general principles of macroevolution.

### 4.3. Conclusion

Inference of the evolutionary history of the Goodeinae has represented a complex challenge in teleost systematics for over a century. Here, the results of molecular phylogenetic analysis provide strong support for recognition of seven major goodeine lineages, but the relationships among them remain uncertain. Investigation of phylogenetic information content reveals that this uncertainty is a result of utilizing markers that contain little information for resolving the earliest divergences within the Goodeinae. This finding undermines support for the current Goodeinae taxonomic framework and, more broadly, underscores the importance of carefully scrutinizing historical studies and legacy markers. Although numerous topological hypotheses have withstood the test of time, our results illustrate that investigating the information content of markers used in previous studies can illuminate areas of false confidence in the Tree of Life and therefore prevent future investigations from being misled or from spending too much time and money on sequencing efforts with little return. Evaluating how confident we should be in past inferences represents an important, but all too often neglected, axis fundamental to understanding the evolution of life on our planet.

## Acknowledgements

We would like to acknowledge John Lyons, Norman Mercado-Silva, Pablo Gesundheidt, Martina Medina, Diego Elias, Devin Bloom, Caleb McMahan, Malorie Hayes, Aaron Geheber, and the Goodeid Working Group for invaluable assistance with collection of samples used in the analyses presented here. We would also like to thank JP Townsend for helpful comments on a prior version of this manuscript. Finally, we would like to express our appreciation to the crew of the Cabo de Hornos for allowing us to use the bridge as an office to complete an early version of this manuscript. Funding for this project was provided by an NSF grant (DEB 1354930) to Kyle Piller.

## Ethics statement

The IACUC protocol was approved by the Southeastern Institutional Animal Care and Use Committee, with approval number SLU#0002. Our sampling protocol followed the AFS guidelines for Use of Fishes in Research (which can be accessed here: https://fisheries.org/docs/wp/Guidelines-for-Use-of-Fishes.pdf). Specifically, fish were euthanized with a lethal dose of MS-222, tissue samples were obtained and preserved in 95% ethanol, and whole specimens were fixed in formalin for permanent archiving.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.01.025.

## References

Alfaro, M.E., Santini, F., Brock, C.D., 2007. Do reefs drive diversification in marine teleosts? Evidence from the pufferfish and their allies (Order Tetraodontiformes). Evolut.: Int. J. Org. Evolut. 61 (9), 2104–2126.
Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G., Harmon, L.J., 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Nat. Acad. Sci. 106 (32), 13410–13414.
Alström, P., Cibois, A., Irestedt, M., Zuccon, D., Gelang, M., Fjeldså, J., Andersen, M.J., Moyle, R.G., Pasquet, E., Olsson, U., 2018. Comprehensive molecular phylogeny of the grassbirds and allies (Locustellidae) reveals extensive non-monophyly of traditional genera, and a proposal for a new classification. Mol. Phylogenet. Evolut. 127, 367–375.
Alfaro, M.E., Faircloth, B.C., Harrington, R.C., Sorenson, L., Friedman, M., Thacker, C.E., Oliveros, C.H., Černý, D., Near, T.J., 2018. Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. Nature Ecol. Evolut. 2, 688–696.
Bangs, M.R., Douglas, M.R., Mussmann, S.M., Douglas, M.E., 2018. Unraveling historical introgression and resolving phylogenetic discord within Catostomus (Osteichthys: Catostomidae). BMC Evol. Biol. 18 (1), 86.
Betancur-R, R., Li, C., Munroe, T.A., Ballesteros, J.A., Ortí, G., 2013a. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62 (5), 763–785.
Betancur-R, R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Cureton II, J.C., 2013b. The tree of life and a new classification of bony fishes. PLoS Curr. 5.
Bossert, S., Murray, E.A., Blaimer, B.B., Danforth, B.N., 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Mol. Phylogenet. Evolut. 111, 149–157.
Brown, J.M., Thomson, R.C., 2016. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. Syst. Biol. 66 (4), 517–530.
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., Sorenson, L., 2017. Phylogenomic systematics of ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of characiformes. Syst. Biol. 66 (6), 881–895.
Chatterjee, H.J., Ho, S.Y., Barnes, I., Groves, C., 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC Evol. Biol. 9 (1), 259.
Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mol. Ecol. 7 (9), 1255–1256.
Cox, C.J., Li, B., Foster, P.G., Embley, T.M., Civán, P., 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63 (2), 272–279.
Dell'Ampio, E., Meusemann, K., Szucsich, N.U., Peters, R.S., Meyer, B., Borner, J., Petersen, M., Aberer, A.J., Stamatakis, A., Walzl, M.G., Minh, B.Q., 2013. Decisive data sets in phylogenomics: lessons from studies on the phylogenetic relationships of primarily wingless insects. Mol. Biol. Evolut. 31 (1), 239–249.

Domínguez-Domínguez, O., Pérez-Rodríguez, R., Doadrio, I., 2008. Morphological and genetic comparative analyses of populations of Zoogoneticus quitzeoensis (Cyprinodontiformes: Goodeidae) from Central Mexico, with description of a new species. Revista Mexicana de Biodiversidad 79 (2), 373–383.
Doadrio, I., Domínguez-Domínguez, O., 2004. Phylogenetic relationships within the fish family Goodeidae based on cytochrome b sequence data. Mol. Phylogenet. Evolut. 31 (2), 416–430.
Domínguez-Domínguez, O., Doadrio, I., Pérez-Ponce de León, G., 2006. Historical biogeography of some river basins in central Mexico evidenced by their goodeine freshwater fishes: a preliminary hypothesis using secondary Brooks parsimony analysis. J. Biogeogr. 33 (8), 1437–1447.
Domínguez-Domínguez, O., Pérez-Ponce de León, G., 2009. ¿ La mesa central de México es una provincia biogeográfica? Análisis descriptivo basado en componentes bióticos dulceacuícolas. Revista mexicana de biodiversidad 80 (3), 835–852.
Domínguez-Domínguez, O., Pedraza-lara, C., Gurrola-Sánchez, N., Perea, S., Pérez-Rodríguez, R., Israde-Alcántara, I., Garduno-Monroy, V.H., Doadrio, I., Pérez-Ponce de León, G., Brooks, D.R., 2010. Historical biogeography of the Goodeinae (Cyprinodontiformes). In: Uribe, M.C., Grier, H.J. (Eds.), Viviparous Fishes II, pp. 19–61.
Dominguez-Dominguez, O., Bernal-Zuniga, D.M., Piller, K.R., 2016. Two new species of the genus *Xenotoca* Hubbs and Turner, 1939 (Teleostei, Goodeidae) from central-western Mexico. Zootaxa 4189 (1), 81–98.
Dornburg, A., Santini, F., Alfaro, M.E., 2008. The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). Syst. Biol. 57 (6), 905–919.
Dornburg, A., Sidlauskas, B., Santini, F., Sorenson, L., Near, T.J., Alfaro, M.E., 2011. The influence of an innovative locomotor strategy on the phenotypic diversification of triggerfish (family: Balistidae). Evolut.: Int. J. Org. Evolut. 65 (7), 1912–1926.
Dornburg, A., Moore, J., Beaulieu, J.M., Eytan, R.I., Near, T.J., 2015. The impact of shifts in marine biodiversity hotspots on patterns of range evolution: evidence from the Holocentridae (squirrelfishes and soldierfishes). Evolut.: Int. J. Org. Evolut. 69 (1), 146–161.
Dornburg, A., Fisk, J.N., Tamagnon, J., Townsend, J.P., 2016. PhyinformR: phylogenetic experimantal design and Phylogenomic data exploration in R. BMC Evolut. Biol. 16 (1), 262.
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., 2017a. New insights on the sister lineage of percomorph fishes with an anchored hybrid enrichment dataset. Mol. Phylogenet. Evolut. 110, 27–38.
Dornburg, A., Townsend, J.P., Wang, Z., 2017b. Maximizing power in phylogenetics and phylogenomics: a perspective illuminated by fungal big data. In: Advances in Genetics, vol. 100. Academic Press, pp. 1–47.
Dornburg, A., Forrestel, E.J., Moore, J.A., Iglesias, T.L., Jones, A., Rao, L., Warren, D.L., 2017c. An assessment of sampling biases across studies of diel activity patterns in marine ray-finned fishes (Actinopterygii). Bull. Mar. Sci. 93 (2), 611–639.
Dornburg, A., Federman, S., Lamb, A.D., Jones, C.D., Near, T.J., 2017d. Cradles and museums of Antarctic teleost biodiversity. Nat. Ecol. Evolut. 1 (9), 1379–1384.
Dornburg, A., Su, Z., Townsend, J.P., 2018. Optimal rates for phylogenetic inference and experimental design in the era of genome-scale datasets. Syst. Biol.
Economo, E.P., Narula, N., Friedman, N.R., Weiser, M.D., Guénard, B., 2018. Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nat. Commun. 9 (1), 1778.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acids Res. 32 (5), 1792–1797.
Ebersberger, I., de Matos Simoes, R., Kupczok, A., Gube, M., Kothe, E., Voigt, K., von Haeseler, A., 2011. A consistent phylogenetic backbone for the fungi. Mol. Biol. Evolut. 29 (5), 1319–1334.
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 acanthomorph teleost phylogeny using Anchored Hybrid Enrichment. BMC Evolut. Biol. 15 (1), 113.
Fitzsimons, J.M., 1972. A revision of two genera of goodeid fishes (Cyprinodontiformes, Osteichthyes) from the Mexican Plateau. Copeia 1972 (4), 728–756.
Foster, K.L., Piller, K.R., 2018. Disentangling the drivers of diversification in an imperiled group of freshwater fishes (Cyprinodontiformes: Goodeidae). BMC Evolut. Biol. 18 (1), 116.
Friedman, M., Keck, B.P., Dornburg, A., Eytan, R.I., Martin, C.H., Hulsey, C.D., Wainwright, P.C., Near, T.J., 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. Roy. Soc. London B: Biol. Sci. 280 (1770), 20131733.
Galtier, N., Roux, C., Rousselle, M., Romiguier, J., Figuet, E., Glémin, S., Bierne, N., Duret, L., 2018. Codon usage bias in animals: disentangling the effects of natural selection, effective population size, and GC-biased gene conversion. Mol. Biol. Evolut. 35 (5), 1092–1103.
Gilbert, C.H., 1893. Report on the fishes of the Death Valley expedition collected in southern California and Nevada in 1891, with descriptions of new species. North Am. Fauna 229–234.
Gilbert, P.S., Wu, J., Simon, M.W., Sinsheimer, J.S., Alfaro, M.E., 2018. Filtering nucleotide sites by phylogenetic signal to noise ratio increases confidence in the Neoaves phylogeny generated from ultraconserved elements. Mol. Phylogenet. Evolut. 126, 116–128.
Goldman, N., 1998. Phylogenetic information and experimental design in molecular systematics. Proc. Roy. Soc. B: Biol. Sci. 265 (1407), 1779–1786.
Grant, E.C., Riddle, B.R., 1995. Are the endangered springfish (*Crenichthys* Hubbs) and poolfish (*Empetrichthys* Gilbert) fundulines or goodeids?: A mitochondrial DNA assessment. Copeia 1995 (1), 209–212.
Grudzien, T.A., White, M.M., Turner, B.J., 1992. Biochemical systematics of the viviparous fish family Goodeidae. J. Fish Biol. 40 (5), 801–814.
Hassanin, A., Leger, N.E., Deutsch, J., 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst. Biol. 54 (2), 277–298.
Helmstetter, A.J., Papadopulos, A.S., Igea, J., Van Dooren, T.J., Leroi, A.M., Savolainen, V., 2016. Viviparity stimulates diversification in an order of fish. Nat. Commun. 7 (11271), 1–7.
Hubbs, C.L., 1924. Studies of the fishes of the order Cyprinodontes. V. Notes on species of Goodea and Skiffia. Occasional Papers of the Museum of Zoology, University of Michigan No. 148, pp. 1–8.
Hubbs, C.L., 1932. Studies of the fishes of the Order Cyprinodontes. XII. A new genus related to Empetrichthys. Occasional Papers of the Museum of Zoology, University of Michigan No. 252, pp. 1–5.
Hubbs, C.L., Turner, C.L., 1939. Studies of the fishes of the order Cyprinodontes. XVI. A revision of the Goodeidae. Miscellaneous Publications of the Museum of Zoology, University of Michigan No. 42, pp. 1–80.
Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y., Faircloth, B.C., Nabholz, B., Howard, J.T., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346 (6215), 1320–1331.
Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22 (4), 225–231.
Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. Nature 491 (7424), 444.
Jordan, David S., 1923. A classification of fishes. Stanford University Publications, University Series, Biological Sciences 3 (2), 77–243.
Kang, J.H., Schartl, M., Walter, R.B., Meyer, A., 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. BMC Evolut. Biol. 13 (1), 25.
Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649.
King, N., Rokas, A., 2017. Embracing uncertainty in reconstructing early animal evolution. Curr. Biol. 27 (19), R1081–R1088.
Klopfstein, S., Massingham, T., Goldman, N., 2017. More on the best evolutionary rate for phylogenetic analysis. Syst. Biol. 66 (5), 769–785.
Lammers, F., Gallus, S., Janke, A., Nilsson, M.A., 2017. Phylogenetic conflict in bears identified by automated discovery of transposable element insertions in low-coverage genomes. Genome Biol. Evolut. 9 (10), 2862–2878.
Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evolut. 29 (6), 1695–1701.
Li, C., Ortí, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evolut. Biol. 7 (1), 44.
Lopez-Giraldez, F., Townsend, J.P., 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evolut. Biol. 11 (1), 152.
Meek, S.E., 1902. A contribution to the ichthyology of Mexico (Vol. 3, No. 6). Field Columbian Museum Publication, Zoological Series 3, 63–128.
Meek, S.E., 1904. The fresh-water fishes of Mexico north of the Isthmus of Tehuantepec. Field Columbian Museum Publication 5 (93), 136–144.
Mendoza, G., 1965. The ovary and anal processes of "*Characodon*" *eiseni*, a viviparous cyprinodont teleost from Mexico. Biol. Bull. 129 (2), 303–315.
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. London B: Biol. Sci. 254 (1340), 153–162.
Miller, R.R., Fitzsimons, J.M., 1971. *Ameca splendens*, a new genus and species of Goodeid fish from western México, with remarks on the classification of the Goodeidae. Copeia 1–13.
Miller, R.R., Minckley, W.L., Norris, S.M., 2005. Freshwater Fishes of Mexico. University of Chicago Press, Chicago, IL.
Minckley, W.L., Deacon, J.E., 1968. Southwestern fishes and the enigma of "endangered species". Science 159 (3822), 1424–1432.
Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., et al., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346 (6210), 763–767.
Moritz, C., 1994. Defining "Evolutionarily Significant Units" for conservation. Tree 9 (10), 373–375.
Near, T.J., Dornburg, A., Kuhn, K.L., Eastman, J.T., Pennington, J.N., Patarnello, T., Zane, L., Fernández, D.A., Jones, C.D., 2012a. Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. Proc. Nat. Acad. Sci. 109 (9), 3434–3439.
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., 2012b. Resolution of ray-finned fish phylogeny and timing of diversification. Proc. Nat. Acad. Sci. 109 (34), 13698–13703.
Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Nat. Acad. Sci. 110 (31), 12738–12743.
Near, T.J., Dornburg, A., Friedman, M., 2014. Phylogenetic relationships and timing of diversification in gonorynchiform fishes inferred using nuclear gene DNA sequences (Teleostei: Ostariophysi). Mol. Phylogenet. Evolut. 80, 297–307.
Near, T.J., Dornburg, A., Harrington, R.C., Oliveira, C., Pietsch, T.W., Thacker, C.E.,

Satoh, T.P., Katayama, E., Wainwright, P.C., Eastman, J.T., 2015. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. BMC Evolut. Biol. 15 (1), 109.
Oscar, M., Edgardo, M., Beryl, B., 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent highâ Andean diversification (Asteraceae: Astereae: Diplostephium). New Phytol. 214 (4), 1736–1750.
Parenti, L.R., 1981. A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). Am. Mus. Nat. History 168 (4), 335–557.
Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20 (2), 289–290.
Pérez-Rodríguez, R., Domínguez-Domínguez, O., Doadrio, I., Cuevas-García, E., Pérez-Ponce de León, G., 2015. Comparative historical biogeography of three groups of Nearctic freshwater fishes across central Mexico. J. Fish Biol. 86 (3), 993–1015.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9 (3), e1000602.
Pond, S.L.K., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. In: Statistical Methods in Molecular Evolution. Springer, New York, NY, pp. 125–181.
Price, S.A., Schmitz, L., Oufiero, C.E., Eytan, R.I., Dornburg, A., Smith, W.L., Friedman, M., Near, T.J., Wainwright, P.C., 2014. Two waves of colonization straddling the K-Pg boundary formed the modern reef fish fauna. Proc. Roy. Soc. London B: Biol. Sci. 281 (1783), 20140321.
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.
Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Nat. Acad. Sci. 103 (42), 15511–15516.
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., 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.
Regier, J.C., Shultz, J.W., Ganley, A.R., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J.E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. 57 (6), 920–938.
Rokas, A., Chatzimanolis, S., 2008. From gene-scale to genome-scale phylogenetics: the data flood in, but the challenges remain. In: Phylogenomics. Humana Press, pp. 1–12.
Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., Douzery, E.J.P., 2013. Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Mol. Biol. Evolut. 30 (9), 2134–2144.
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 (3), 539–542.
Santini, F., Harmon, L.J., Carnevale, G., Alfaro, M.E., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evolut. Biol. 9 (1), 194.
Santini, F., Sorenson, L., Alfaro, M.E., 2013a. A new multi-locus timescale reveals the evolutionary basis of diversity patterns in triggerfishes and filefishes (Balistidae, Monacanthidae; Tetraodontiformes). Mol. Phylogenet. Evolut. 69 (1), 165–176.
Santini, F., Sorenson, L., Marcroft, T., Dornburg, A., Alfaro, M.E., 2013b. A multilocus molecular phylogeny of boxfishes (Aracanidae, Ostraciidae; Tetraodontiformes). Mol. Phylogenet. Evolut. 66, 153–160.
Streicher, J.W., Miller, E.C., Guerrero, P.C., Correa, C., Ortiz, J.C., Crawford, A.J., Pie, M.R., Wiens, J.J., 2018. Evaluating methods for phylogenomic analyses, and a new phylogeny for a major frog clade (Hyloidea) based on 2214 loci. Mol. Phylogenet. Evolut. 119, 128–143.
Shen, X.-X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evolut. 1 (5), 0126.
Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D.J., Di Franco, A., Roure, B., Satoh, N., Quéinnec, É., Ereskovsky, A., Lapébie, P., 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr. Biol. 27 (7), 958–967.
Smith, M.L., 1980. The Evolutionary and Ecological History of the Fish Fauna of the Rio Lerma Basin, Mexico. Doctoral dissertation, Ph. D. Thesis. University of Michigan, pp. 201.
Smith, S.A., Beaulieu, J.M., Donoghue, M.J., 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evolut. Biol. 9 (1), 37.
Soltz, D.L., Naiman, R.J., 1978. The natural history of native fishes in the Death Valley system. Natural History of Los Angeles County, Los Angeles, CA, USA. Science Series 30.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313.
Su, Z., Wang, Z., López-Giráldez, F., Townsend, J.P., 2014. The impact of incorporating molecular evolutionary model into predictions of phylogenetic signal and noise. Front. Ecol. Evolut. 2, 11.
Tang, Q., Edwards, S.V., Rheindt, F.E., 2018. Rapid diversification and hybridization have shaped the dynamic history of the genus Elaenia. Mol. Phylogenet. Evolut. 127, 52–533.
Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56 (2), 222–231.
Townsend, J.P., Su, Z., Tekle, Y.I., 2012. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. Syst. Biol. 61 (5), 835–849.
Turner, C.L., 1946. A contribution of the taxonomy and zoogeography of the goodeid fishes. University of Michigan. Occasional Papers of the Museum of Zoology, University of Michigan No. 495, pp. 1–13.
Webb, S.A., 1998. A phylogenetic analysis of the Goodeidae (Teleostei: Cyprinodontiformes). Ph.D. Dissertation. University of Michigan, Ann Arbor, MI.
Webb, S.A., Graves, J.A., Macias-garcia, C., Ritchie, M.G., Magurran, A.E., Diarmaid, O., 2004. Molecular phylogeny of the livebearing Goodeidae (Cyprinodontiformes). Mol. Phylogenet. Evolut. 30 (3), 527–544.