Syst. Biol. 0(0):1–15, 2017

DOI:10.1093/sysbio/syx038

Phylogenomic Systematics of Ostariophysan Fishes: Ultraconserved Elements Support the Surprising Non-Monophyly of Characiformes

PROSANTA CHAKRABARTY[1,*], BRANT C. FAIRCLOTH[1,*], FERNANDO ALDA[1], WILLIAM B. LUDT[1], CALEB D. MCMAHAN[1,2], THOMAS J. NEAR[3], ALEX DORNBURG[4], JAMES S. ALBERT[5], JAIRO ARROYAVE[6], MELANIE L. J. STIASSNY[7], LAURIE SORENSON[1,8], AND MICHAEL E. ALFARO[8]

[1]*Museum of Natural Science and Department of Biological Sciences, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA;* [2]*The Field Museum of Natural History, 1400 S Lake Shore Dr, Chicago, IL 60605, USA;* [3]*Department of Ecology and Evolutionary Biology, and Peabody Museum of Natural History, Yale University, New Haven, CT 06520, USA;* [4]*North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA;* [5]*Department of Biology, University of Louisiana, Lafayette, LA 70504, USA;* [6]*Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, México;* [7]*Department of Ichthyology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA;* [8]*Department of Ecology and Evolutionary Biology, University of California Los Angeles, 610 Yound Drive South, Los Angeles, CA 90095, USA*
**Correspondence to be sent to: Museum of Natural Science and Department of Biological Sciences, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803 USA E-mail: prosanta@lsu.edu; brant@lsu.edu*

Prosanta Chakrabarty and Brant C. Faircloth contributed equally to this article.

Received 17 August 2015; reviews returned 11 September 2015; accepted 24 February 2016
Associate Editor: Mark Holder

Abstract.—Ostariophysi is a superorder of bony fishes including more than 10,300 species in 1100 genera and 70 families. This superorder is traditionally divided into five major groups (orders): Gonorynchiformes (milkfishes and sandfishes), Cypriniformes (carps and minnows), Characiformes (tetras and their allies), Siluriformes (catfishes), and Gymnotiformes (electric knifefishes). Unambiguous resolution of the relationships among these lineages remains elusive, with previous molecular and morphological analyses failing to produce a consensus phylogeny. In this study, we use over 350 ultraconserved element (UCEs) loci comprising 5 million base pairs collected across 35 representative ostariophysan species to compile one of the most data-rich phylogenies of fishes to date. We use these data to infer higher level (interordinal) relationships among ostariophysan fishes, focusing on the monophyly of the Characiformes—one the most contentiously debated groups in fish systematics. As with most previous molecular studies, we recover a non-monophyletic Characiformes with the two monophyletic suborders, Citharinoidei and Characoidei, more closely related to other ostariophysan clades than to each other. We also explore incongruence between results from different UCE data sets, issues of orthology, and the use of morphological characters in combination with our molecular data. [Conserved sequence; ichthyology; massively parallel sequencing; morphology; next-generation sequencing; UCEs.]

A number of large (multi-taxon) studies covering significant portions of the Tree of Life of fishes have been published during the last few years, including those focusing on diversification and morphological evolution (Rabosky et al. 2013), feeding innovations (Wainwright et al. 2015), fossils (Friedman and Sallan 2012), and acanthomorph (spiny-rayed fish) lineages (Wainwright et al. 2012; Near et al. 2012, 2013; Betancur-R et al. 2013; Thacker et al. 2015). Paralleling those analyses have been advances in technologies dealing with sequence data, including next-generation sequencing techniques (also called massively parallel sequencing) (Crawford et al. 2012; Faircloth et al. 2012, 2014; McCormack et al. 2012; Wagner et al. 2013; Smith et al. 2014), but few phylogenies using these data-rich approaches have focused on higher level relationships of fishes (Faircloth et al. 2013; Eytan et al. 2015; Harrington et al. 2016; Stout et al. 2016; Arcila et al. 2017). The essence of these new technologies is that they greatly reduce barriers to data collection that hamper traditional (Sanger) approaches and offer the potential to collect data from at least an order of magnitude more loci. In this study, we use a phylogenomic approach to examine one of the most difficult problems in systematic ichthyology: resolution of relationships among major ostariophysan lineages.

Ostariophysi consists of over 10,300 currently recognized species comprising roughly 30% of all known bony fish species, 75% of all freshwater fish species, and approximately one-sixth of all vertebrate species (Eschmeyer and Fong 2015; Nelson 2006; Van Der Laan et al. 2014). Ostariophysi includes some of the smallest known vertebrates, with adult sizes of about 1 mm (e.g., *Paedocypris progenetica*), and some of the largest known fishes including the enormous Mekong giant catfish (*Pangasianodon gigas*) 2.7 m in length and weighing 300kg and perhaps the even larger Siamese giant barb (*Catlocarpio siamensis*). The superorder also includes the electric catfishes (Malapteruridae) and electric knifefishes (Gymnotiformes), most famously the electric eel (*Electrophorus electricus*), and well-known apex predators such as piranhas (Serrasalmidae) in South America and tigerfishes (Alestidae) in Africa. Other ostariophysans include Cypriniformes such as the common goldfish (*Carassius auratus*), carp and koi (*Cyprinus carpio*), and the well-studied zebrafish (*Danio rerio*). The monophyly of this notably widespread, species rich, and morphologically diverse group has been supported by studies using morphological (Greenwood et al. 1966; Rosen and Greenwood 1970; Rosen 1973; Novacek and Marshall 1976; Fink and Fink 1981, 1996) and traditional DNA sequence

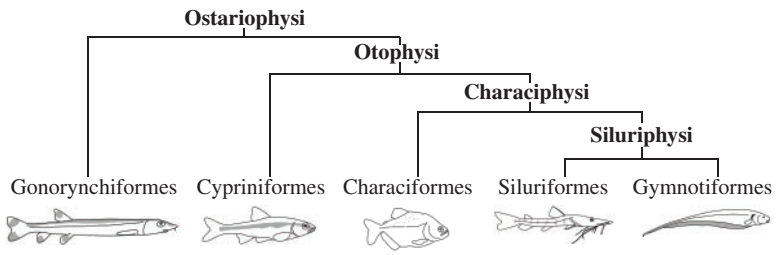

FIGURE 1. Traditional hypothesis of relationships of Ostariophysi, showing subgroups (viz. superorder, orders) from Fink and Fink (1981).

data (Ortí and Meyer 1996; Lavoué et al. 2005; Li et al. 2008; Poulsen et al. 2009; Broughton 2010; Nakatani et al. 2011; Near et al. 2012; Betancur-R et al. 2013; Broughton et al. 2013).

The incredible species richness of Ostariophysi, particularly in freshwaters, has been linked to a number of key traits. All ostariophysans share a specialized modification of the anterior vertebrae and surrounding tissues called the Weberian apparatus (Berg 1912; Rosen and Greenwood 1970). These structures transmit sound waves from the swim bladder to the inner ear, greatly enhancing sensitivity to acoustic information (Braun and Grande 2008) and increasing the hearing abilities of many species to levels on par with, or even exceeding, those of humans and other mammals (Yan et al. 2000). Most ostariophysans also emit an alarm substance, called schreckstoff, produced in specialized club cells in the skin (Frisch 1942; Chivers et al. 2007). This alarm substance is important in social communication and predator avoidance, and together with enhanced auditory acuity, has been considered an innovation underlying the successful radiation of ostariophysans (Helfman et al. 2009).

Ostariophysi is currently classified into five orders (Nelson 2006; Eschmeyer and Fong 2015): Gonorynchiformes (milkfishes and sandfishes; 37 species); Siluriformes (catfishes; ~3700 species); Characiformes (tetras, piranhas, and allies; ~2100 species); Gymnotiformes (electric eel and knifefishes; 225 species); and Cypriniformes (carps and minnows; ~4262 species). Unfortunately, there is no consensus of the evolutionary relationships between these major ostariophysan lineages, and different studies have recovered almost every possible arrangement of relationships among the Gymnotiformes, Siluriformes, and the two suborders of Characiformes; together the three orders are called the Characiphysi (Figs. 1 and 2).

For example, and perhaps most notably, many features of the passive electroreceptive system of Gymnotiformes and Siluriformes support their inclusion in a clade known as Siluriphysi (Fink and Fink 1996; Fig. 1). This relationship is morphologically supported by ultrastructural, chemical, functional, and embryological aspects of the low frequency-sensitive, ampullary-shaped electroreceptor organs in the skin; the uniquely laminated medullary electrosensory lateral line lobe; and other structures of the peripheral and central nervous systems (see reviews in Albert 2001; Albert et al. 1998; Liu et al. 2016). Newly discovered losses of SWS1 and SWS2 opsin genes and receptor proteins are also Siluriphysi synapomorphies (Liu et al. 2016). However, this clade has never been recovered in a published molecular phylogeny (Ortí and Meyer 1996, 1997; Nakatani et al. 2011; Chen et al. 2013, among others).

Resolving the relationship between the two suborders of Characiformes (viz., Citharinoidei, Characoidei) has been equally contentious. Characiformes is composed of Neotropical (14 families, 1400+ species) and African (4 families, 200+ spp.) lineages united by seven putatively unreversed osteological synapomorphies (Fink and Fink 1981, 1996). In addition to anatomical similarities, strong similarities in body shape and ecology have led most ichthyologists to the hypothesis that Characiformes is monophyletic—even among those who recovered Characiformes as non-monophyletic in their own studies (Nakatani et al. 2011; Ortí and Meyer 1997). Despite multiple molecular investigations, only three molecular phylogenies sampling at least two individuals from each suborder have recovered a monophyletic Characiformes: (i) Betancur-R et al. (2013) who included only two members of Citharinoidei; (ii) Oliveira et al. (2011) who did not sample any closely related ostariophysans (only cypriniform outgroups)—effectively forcing the monophyly of Characiformes; and (iii) Arcila et al. (2017) who recovered monophletic Characiformes using exons and a novel approach for data selection and signal parsing. A clear understanding of the evolutionary relationships of major characiform lineages is critical because Characiformes represents one of the few well-supported examples of Gondwanan fragmentation in fish biogeography; the origin of Characiformes has been hypothesized to precede the mid-Cretaceous fragmentation of Gondwana (Lundberg 1993; Buckup 1998; Malabarba and Malabarba 2010; Arroyave et al. 2013; Kappas et al. 2016), yet lack of phylogenetic resolution has hindered our understanding of how geological events shaped the early radiation of the group.

In this study, we analyze ultraconserved elements (UCEs) using concatenated maximum-likelihood (ML), concatenated Bayesian, and multispecies coalescent approaches to infer the major relationships among ostariophysans, with a particular focus on testing the hypothesis of characiform monophyly. We also use these data to explore how morphological and molecular data, in combination, can influence phylogenetic results in the era of phylogenomic approaches.

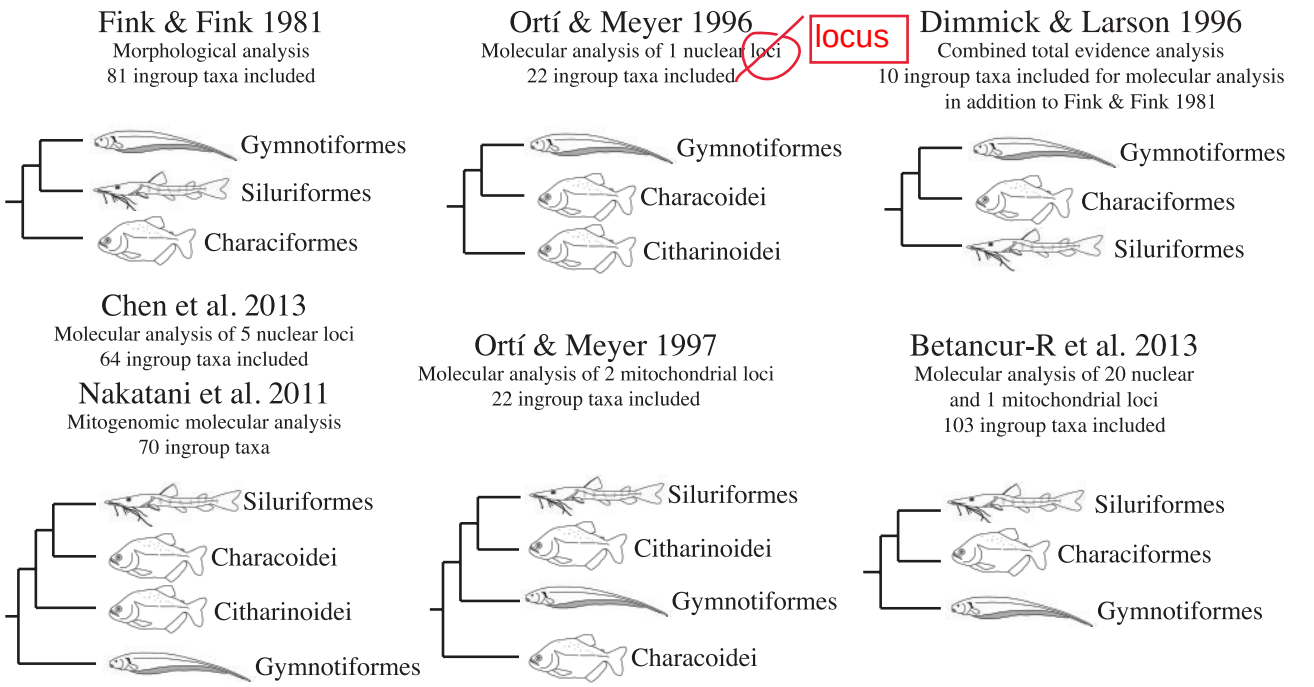

FIGURE 2. A review of previously recovered hypotheses of Characiphysian relationships. Note the Fink and Fink (1981) phylogeny was recently also recovered by Arcila et al. (2017) using a phylogeny of protein-coding genes (see text).

MATERIALS AND METHODS

Morphological Data

Binary-coded character states were transcribed for characters from the morphological data presented in Fink and Fink (1981). The original authors scored multiple characters from different anatomical systems that include the neurocranium (15), orbital region (3), suspensorium (17), oral jaws (9), gill arches (8), gas bladder (4), anterior vertebrae (37), pectoral girdle (5), pectoral fins (2), dorsal fin, anal fin, and fin supports (4), caudal fin and supports (8), fin spines (1). Fink and Fink (1981) classified 11 additional characters as miscellaneous, and these included states such as alarm substances, nuptial tubercles, and electroreception. Fink and Fink (1996) excluded one character in subsequent analyses, the condition of the posterior occipital margin (character number 14) based on an expanded survey of species, and we exclude this character as well. The final binary character matrix consisted of 127 characters, and full character descriptions are provided in Fink and Fink (1981, 1996).

A phylogeny was inferred from these morphological data using RAxML 8.0.19 (Stamatakis 2014) and the BINGAMMA model. We conducted 20 ML searches for the phylogenetic tree that best fit the data, and we assessed support for these relationships by generating and analyzing nonparametric bootstrap replicates of the input data using the autoMRE function of RAxML. Following convergence of the bootstrap replicates, we reconciled the best-fitting ML tree with the bootstrap replicates. The best-fit ML tree was rooted with the designated outgroup taxon scoring (i.e., all "0" outgroup), as presented in Fink and Fink (1981).

A Bayesian phylogenetic analyses was performed with the morphological data set using MrBayes 3.2.2 (Ronquist et al. 2012). Characters were treated as *standard* using the Mk_v model of character evolution. MrBayes was run for 1.0×10^6 generations (four chains per run; burn-in: 25%; thinning = 500). Convergence was assessed by visually inspecting the potential scale reduction factors and the average deviation of the split frequencies output from MrBayes. We also visually examined traces and Effective Sample Sizes (ESS) values for estimated parameters using Tracer v1.6 (Rambaut et al. 2014). We constructed majority-rule consensus trees and the posterior probabilities of nodes in the phylogeny using the sample of trees following burn-in and thinning.

UCEs Data Capture

A targeted sequencing approach was used to collect phylogenomic data from UCE loci (Faircloth et al. 2012, 2013) across a taxon sample spanning ostariophysans. Specifically, we sampled taxa widely across families to select those that best represented each of the orders while also having tissues available from vouchered museum specimens to avoid misidentifications (Table 1). In order to robustly test the monophyly of the Characiformes, our taxon sampling included representatives of several families from the two major characiform lineages (i.e., Citharinoidei and Characoidei), with thorough coverage of African families (e.g., Citharinidae, Alestidae,

TABLE 1. Scientific name, catalog number of specimen, and locality information for ostariophysan species enriched for UCEs.

Scientific name	Group represented	Voucher catalog number	Locality
Brycinus macrolepidotus	Characiformes (Alestidae)	AMNH 259000 (AMCC 236014)	Guinea
Bryconaethiops yseuxi	Characiformes (Alestidae)	AMNH 253583 (Tissue # t-085-8485)	Democratic Republic of Congo
Astyanax aeneus	Characiformes (Characidae)	LSUMZ 14521 (Tissue # 1179)	Honduras
Citharinus gibbosus	Characiformes (Citharinidae)	AMNH 243512 (Tissue # t-030-2974)	Democratic Republic of Congo
Ctenolucius beani	Characiformes (Ctenoluciidae)	LSUMZ 14819 (Tissue # 1793)	Panama
Distichodus hypostomatus	Characiformes (Distichodontidae)	AMNH 253909 (Tissue # t-088-8738)	Republic of Congo
Distichodus maculatus	Characiformes (Distichodontidae)	AMNH 252806 (Tissue # t-080-7944)	Democratic Republic of Congo
Nannaethiops unitaeniatus	Characiformes (Distichodontidae)	AMNH 249555 (Tissue # t-063-6266)	Cameroon
Xenocharax crassus	Characiformes (Distichodontidae)	AMNH 242511 (Tissue # t-034-3330)	Democratic Republic of Congo
Hoplias malabaricus	Characiformes (Erythrinidae)	LSUMZ 14831 (Tissue # 1853)	Panama
Hepsetus lineata	Characiformes (Hepsetidae)	AMNH 253986 (AMCC 255985)	Democratic Republic of Congo
Chirocentrus dorab[a]	Clupeiformes (Chirocentridae)	LSUMZ 13950 (Tissue # 620)	Vietnam
Dorosoma petenense	Clupeiformes (Clupeidae)	LSUMZ 13694 (Tissue # 394)	Louisiana
Thryssa hamiltonii[a]	Clupeiformes (Engraulidae)	LSUMZ 13314 (Tissue # 17)	Taiwan
Moxostoma poecilurum	Cypriniformes (Catosomidae)	LSUMZ 14297 (Tissue # 989)	Louisiana
Cyprinella venusta	Cypriniformes (Cyprinidae)	LSUMZ 14299 (Tissue # 991)	Louisiana
Puntioplites falcifer	Cypriniformes (Cyprinidae)	LSUMZ 14182 (Tissue # 872)	Vietnam
Chanos chanos	Gonorhynchiformes (Chanidae)	LSUMZ 16723 (Tissue # 5274)	Singapore
Gonorynchus cf. abbreviatus	Gonorhynchiformes (Gonorhynchidae)	LSUMZ 13531 (Tissue # 245)	Taiwan
Parakneria abbreviata	Gonorhynchiformes (Kneriidae)	AMNH 253961 (AMCC 236010)	Congo
Phractolaemus ansorgii	Gonorhynchiformes (Phractolaemidae)	AMNH 250985 (Tissue # t-069-6895)	Democratic Republic of Congo
Apteronotus albifrons	Gymnotiformes (Apteronotidae)	MUSM 36939	Peru
Gymnotus carapo	Gymnotiformes (Gymnotidae)	MUSM 36951	Peru
Gymnotus cylindricus	Gymnotiformes (Gymnotidae)	LSUMZ 14535 (Tissue # 1201)	Honduras
Brachyhypopomus occidentalis	Gymnotiformes (Hypopomidae)	LSUMZ 14836 (Tissue # 1849)	Panama
Sternopygus macrurus	Gymnotiformes (Sternopygidae)	MUSM 39502	Peru
Sciades felis	Siluriformes (Ariidae)	LSUMZ 13605 (Tissue # 326)	Gulf of Mexico
Auchenoglanis occidentalis	Siluriformes (Claroteidae)	AMNH 256412 (AMCC 221170)	Guinea
Ameiurus natalis	Siluriformes (Ictaluridae)	LSUMZ 14282 (Tissue # 974)	Louisiana
Hypostomus aspidolepis	Siluriformes (Loricariidae)	LSUMZ 14823 (Tissue # 1809)	Panama
Malapterurus stiassnyae	Siluriformes (Malapteruridae)	AMNH 255559 (AMCC 220546)	Guinea
Synodontis filamentosus	Siluriformes (Mochokidae)	AMNH 256414 (AMCC 223046)	Guinea
Pangasius cf. pangasius	Siluriformes (Pangasiidae)	LSUMZ 14187 (Tissue # 880)	Vietnam
Plotosus lineatus	Siluriformes (Plotosidae)	LSUMZ 13391 (Tissue # 94)	Taiwan

Notes: Links to GBIF (Global Biodiversity Information Facility) or other web records are included (when available) with catalog numbers when clicked in the "Voucher Catalog Number" column. Genetic data from these organisms are posted on NCBI GenBank (PRJNA363064) and we list voucher information there. AMCC = Ambrose Monell Cryo Collection; AMNH = American Museum of Natural History (New York, USA); LSUMZ = Louisiana State University Museum of Natural Science (Baton Rouge, USA); MUSM = Museu Universidade San Marcos (Lima, Peru).
[a]Outgroup taxa.

Hepsetidae, and Distichodontidae), which were underrepresented in all previous studies. Our focus was on having deep sampling of characiphysans. Among the Characoidei, 6 of 21 families were sampled, including both African and Neotropical members representing all the major clades, and also including those sampled in previous studies by Chen et al. (2013) and Nakatani et al. (2011). Both families of Citharinoidei were sampled for five species, all of African origin; sampling was guided by the relationships recovered in Arroyave et al. (2013). Among Siluriformes, 8 of 36 families were sampled, and 4 of 5 families of Gymnotiformes were sampled; our sampling of major lineages was guided by previous studies (see references in Fig. 2). All identifications were conducted by experts and all specimens were vouchered in collections corresponding to gen-seq 4 (vouchered non-type specimens) following the reliability ranking system of sequence data (Chakrabarty et al. 2013).

DNA was extracted using DNeasy extraction kits (Qiagen, Inc.) or phenol–chloroform extraction procedures (Maniatis et al. 1982). Following extraction, we quantified DNA extracts using a Qubit fluorometer (Life Technologies, Inc.) and randomly sheared DNA to a target size of approximately 500 bp (range 400–800 bp) by sonication (Diagenode BioRuptor). We input 100–1000-ng sheared DNA to a modified genomic DNA library preparation protocol (v1.10; http://ultraconserved.org) that incorporated "with-bead" cleanup steps (Fisher et al. 2011) using a generic SPRI substitute (hereafter SPRI; Rohland and Reich 2012). This protocol is similar to the Kapa Biosystems protocol that uses commercial SPRI chemistry for cleanup, except that the Fisher modification removes and replaces a 25-mM NaCl + PEG solution, leaving the beads in-solution throughout the library preparation steps until their removal prior to PCR amplification. During adapter ligation, we also substituted custom-designed, sequence-tagged adapters to the ligation reaction (Faircloth and Glenn 2012).

Following adapter ligation, 50% of the resulting library volume (~15μL; 50–500 ng) was PCR amplified using a reaction mix of 25μL HiFi HotStart polymerase (Kapa Biosystems), 5μL of Illumina TruSeq primer mix (5μM

each), and 5μL ddH_20 using the following thermal protocol: 98° C for 45 s; 10–16 cycles of 98° C for 15 s, 60° C for 30 s, 72° C for 60 s; and a final extension of 72° C for 5 min. Resulting reactions were purified using 1X SPRI, and we rehydrated libraries in 33μL ddH_2O. We quantified 2μL of each library using a Qubit fluorometer. Groups of eight libraries at equimolar ratios were combined into enrichment pools having a final concentration of 147 ng/μL.

Using a custom target enrichment bait kit from MYcroarray Inc. (Faircloth et al. 2012; McGee et al. 2016), a standard UCE enrichment protocol was followed (v1.4; http://ultraconserved.org). Hybridization reactions were run for 24 h at 65° C. Following hybridization we bound all pools to streptavidin beads (MyOne C1, Life Technologies) and washed bound libraries according to the protocol. We prepared DNA for post-enrichment recovery PCR following the standard (v2.4; http://ultraconserved.org) approach of dissociating enriched DNA from RNA baits bound to streptavidin-coated beads with 0.1 N NaOH, followed by a 5-min neutralization of NaOH using an equal volume of 1 M Tris-HCl, a 1X SPRI cleanup, and elution of the SPRI-purified sample in 30μL of ddH_2O. We combined 15μL of clean, post-enrichment template DNA with 25μL HiFi HotStart Taq (Kapa Biosystems), 5μL of Illumina TruSeq primer mix (5μ M each), and 5μL of ddH_2O. We ran PCR recovery of each library using the following thermal profile: 98° C for 45 s; 16–18 cycles of 98° C for 15 s, 60° C for 30 s, 72° C for 60 s; and a final extension of 72° C for 5 min. Resulting reactions were purified using 1.8X SPRI, and enriched pools were rehydrated in 33μL ddH_2O. Using a Qubit fluorometer, 2μL of each enriched pool was quantified and diluted to 2.5 ng/μL for qPCR library quantification. Using the diluted DNA, we used a qPCR quantification kit (Kapa Biosystems) to quantify libraries, assuming an average library fragment length of 500 bp during our calculations. Based on the size-adjusted concentrations estimated by qPCR, we created an equimolar pool of libraries at 10 nM concentration, and we sequenced 9–10 pmol of each pool-of-pooled libraries using either (Supplementary Table S2, available on Dryad at http://dx.doi.org/10.5061/dryad.n9g60) paired-end, 250 bp sequencing on an Illumina MiSeq (v2; UCLA Genotyping Core Facility) or paired-end 100 bp sequencing on an Illumina HiSeq (v3; UCLA Neuroscience Genomics Core).

Analysis of Captured UCE Data

FASTQ data from Illumina BaseSpace were trimmed for adapter contamination and low-quality bases using a parallel wrapper (https://github.com/faircloth-lab/illumiprocessor) around Trimmomatic (Bolger et al. 2014). Following read trimming, summary statistics were computed on the data using get_fastq_stats.py from the PHYLUCE package. To assemble the cleaned reads, we generated separate data sets using parallel wrappers around Trinity [trinityrnaseq-r2013-02-25; assemblo_trinity.py) (Marçais and Kingsford 2011; Grabherr et al. 2011)]. We computed coverage across assembled contigs using a program (get_trinity_coverage.py) that realigns the trimmed sequence reads to each set of assembled contigs using BWA-MEM (Li 2013), cleans the resulting BAM files using PICARD (1.99; http://picard.sourceforge.net/), adds read-group (RG) information to each library using PICARD, indexes the resulting BAM file using SAMTOOLS (Li et al. 2009), and calculates coverage at each base of each assembled contig using GATK (2.7.2; McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013).

To identify assembled contigs representing enriched UCE loci from each species, we aligned species-specific contig assemblies to a FASTA file of all enrichment baits using match_contigs_to_loci.py. We created a file containing the names of 34 enriched taxa from which we collected data, as well as *D. rerio*, a genome-enabled fish, and we input this list to an additional program (get_match_counts.py) that queries the relational database created by matching baits to assembled contigs, as well as a relational database containing UCE match data for genome-enabled taxa (http://github.com/faircloth-lab/uce-data-sets), to generate a list of UCE loci shared among all taxa. We then used this list with an additional program (get_fastas_from_match_counts.py) to create a monolithic FASTA file containing all UCE sequence data for all taxa. We exploded the data in this file by locus (explode_get_fastas_file.py), and we uploaded the individual files to a cluster computer for alignment using SATé (2.2.7; Katoh et al. 2002; Wheeler and Kececioglu 2007; Liu et al. 2009, 2012: Sukumaran and Holder 2010; Katoh and Standley 2013; Yu et al. 2013). Following SATé alignment, we trimmed the resulting alignments using a parallel wrapper around GBLOCKS (get_gblocks_trimmed_alignments_from_untrimmed.py). From these trimmed alignments, we created two separate data sets (get_only_loci_with_min_taxa.py) representing different subsets of alignments: one subset that was 75% complete (each alignment contained data from a minimum of 26 taxa), and another that was 50% complete (each alignment contained data from a minimum of 17 taxa). For each data set, we generated alignment statistics and computed the number of informative sites across all alignments using get_align_summary_data.py and get_informative_sites.py. We also computed the number of alignment patterns in each concatenated alignment using RAxML.

The UCE data sets were concatenated into separate PHYLIP-formatted supermatrices (format_nexus_files_for_raxml.py), and we determined the best-fitting partitioning scheme and nucleotide substitution models for the concatenated loci using PartitionFinder (Lanfear et al. 2012) with the hcluster (Lanfear et al. 2014) search scheme. We evaluated models for RAxML (only) and selected among candidate partitioning

schemes using BIC. We used the hcluster partitioning method due to the number of loci in the data matrix, which renders traditional partitioning approaches computationally intractable (Lanfear et al. 2014).

A phylogeny was inferred from each UCE data subset (both 75% and 50% complete) using RAxML, the partitioning scheme output by PartitionFinder, and the GTRGAMMA model. We conducted 20 ML searches for the phylogenetic tree that best fit the data, and we assessed support for these relationships by generating and analyzing nonparametric bootstrap replicates of the input data using the autoMRE function of RAxML. Following convergence of the bootstrap replicates, we reconciled the best-fitting ML tree with the bootstrap replicates.

A Bayesian phylogenetic tree was also inferred for each UCE data subset using MrBayes. The PHYLIP data were converted to NEXUS format, and we used the scheme estimated by PartitionFinder to partition the data and apply a GTRGAMMA substitution model to each partition. MrBayes was run for 2.5 to 5.0×10^6 generations (four chains per run; burn-in: 50%; thinning = 500). Convergence was assessed using visualizations of the potential scale reduction factors and the average deviation of the split frequencies output by MrBayes. We also visually examined traces and ESS values for parameters using Tracer, and we computed the parameter estimates and the standard deviation (SD) of split frequencies, post hoc, using the postProcParam and sdsf programs from the ~~exabayes~~ (v1.4.1) package (Aberer et al. 2014). We constructed majority-rule consensus trees and the posterior probabilities of nodes in the phylogeny using the sample of trees following burn-in and thinning

UCE Multispecies Coalescent Approach

To account for heterogeneous gene histories that may influence accurate resolution of phylogenetic relationships (Degnan and Rosenberg 2006; Edwards 2009), we inferred a species tree from individual gene trees. We used RAxML to estimate gene trees for each UCE locus by conducting 20 ML searches. We assessed support for relationships in each individual gene tree by generating 200 nonparametric bootstrap replicates of the input data. The species tree was inferred from the best-fitting ML gene trees and bootstrap replicates using ASTRAL-II v4.10.5 (Mirarab ~~et al.~~ 2015).

We used the topologies and branch lengths obtained from ASTRAL-II to identify regions of the species tree that are potentially in the anomaly zone, and could be responsible for phylogenetic conflict among differing analytical approaches. The anomaly zone occurs when a gene tree topology is more common than the true species tree and can result as a consequence of rapid diversification in combination with large effective population sizes (Degnan and Rosenberg 2006). In such cases, concatenated phylogenetic inferences will likely support the anomalous gene tree topology (Kubatko and Degnan 2007, Liu and Edwards 2009).

Following the unifying principle of the anomaly zone (Rosenberg 2013) and the procedure described in Linkem et al. (2016), we broke up our species tree topology into sets of four-taxon trees, and used them individually to calculate, for each pair of parent–child internodes, the limit of the anomaly zone $a(x)$ using the following equation (Linkem et al. 2016):

$$a(x) = \log\left[\frac{2}{3} + \frac{3e^{2x} - 2}{18(e^{3x} - e^{2x})}\right]$$

Here, x is the length of the branch in the species tree that has a descendant internal branch y. If the length of the descendant internal branch y is shorter than $a(x)$, then the species tree is in the anomaly zone. According to this relationship, the probability of having an anomalous gene tree increases exponentially as the branch length x gets shorter. This calculation was conducted on the median values of internal branch lengths estimated by ASTRAL-II in coalescent units.

UCEs + Morphology

The UCE data described above was combined with the morphological data from Fink and Fink (1981, 1996) to create a combined UCE+morphology data set. A phylogeny was inferred from these combined data using RAxML, the partitioning scheme output by PartitionFinder and the GTRGAMMA model for the UCE data, and the BINGAMMA model for the morphological data. Twenty ML searches were conducted for the phylogenetic tree that best fit the data, and we assessed support for these relationships by generating and analyzing nonparametric bootstrap replicates of the input data using the autoMRE function of RAxML. Following convergence of the bootstrap replicates, we reconciled the best-fitting ML tree with the bootstrap replicates.

A Bayesian phylogenetic ~~analyses~~ of these combined data was conducted using MrBayes. The partitioning scheme described above was used for the UCE data, the character data were treated as *standard*, and we used the Mk_v model of character evolution, as we did for the morphology-only data analysis described above. MrBayes was run for 5.0×10^6 generations (four chains per run; burn-in: 50%; thinning = 500). Convergence was assessed using visualizations of the potential scale reduction factors and the average deviation of the split frequencies output by MrBayes. We also visually examined traces and ESS values for parameters using Tracer, and we computed the parameter estimates and the SD of split frequencies, post hoc, using the postProcParam and sdsf programs from the ~~exabayes~~ (v1.4.1) package. We constructed majority-rule consensus trees and the posterior probabilities of nodes in the phylogeny using the sample of trees following burn-in and thinning.

Shimodaira–Hasegawa (SH) tests (likelihood-based statistical tests of competing evolutionary hypotheses, Shimodaira and Hasegawa 1999) were run to examine statistical differences in tree topology (related to Characiform monophyly) for both partitioned and concatenated data sets with and without morphology using the –f h option in ~~RAXML.~~

Results

For an unrelated project, we used a fraction of our HiSeq run to generate more reads than normal from *Gymnotus cylindricus* (24 million reads). Excluding the large number of reads dedicated to *G. cylindricus*, sequencing (Supplementary Table S1, available on Dryad) produced an average of 755,927 reads per sample (±95 CI = 148,320) with 737,327 (±95 CI = 179,893) reads per sample on the MiSeq and 784,542 (±95 CI = 263,920) reads per sample on the HiSeq (Supplementary Table S1, available on Dryad). After trimming adapter contamination and low-quality bases, reads averaged 93.3 bp (±95 CI = 0.8) for PE100 and 188.2 bp (±95 CI = 4.6) for PE250. These reads assembled into an average of 31,231 contigs (±95 CI = 14,784), and the number of contigs differed by platform/read length (Supplementary Table S2, available on Dryad). Among those samples from which we enriched UCE loci, we identified an average of 510.0 contigs (±95 CI =35.6) representing unique UCE loci (Supplementary Table S3, available on Dryad). On average, these UCE contigs were 536.4 bp (±95 CI =49.0; range: 201–1726 bp) in length, the average coverage per UCE locus was 23.7× (±95 CI=4.3×), and the average number of reads on UCE targets was 24% (±95 CI=8%; Supplementary Table S2, available on Dryad). We also identified and extracted 640 unique UCE loci from the genome sequence of *D. rerio*.

After alignment and trimming of each UCE locus, we created two data sets of 50% and 75% completeness. The 50% complete alignment retained a minimum of 17 taxa per alignment and 567 loci within the data set. Alignments within the 50% complete data set had an average, post-trimming length of 257.3 bp (±95 CI=6.57; range 104–563 bp). The average number of taxa in each of these alignments was 26.3 (±95 CI=0.4; range 17–35 taxa). The supermatrix we generated from these data contained 145,897 characters; 83,464 informative sites; and 69,765 site patterns. The 75% complete alignment retained a minimum of 26 taxa per alignment and 353 loci within the data set. Alignments within the 75% complete matrix had an average, post-trimming length of 269.9 (±95 CI=8.0; range 116–563 bp). The average number of taxa in each of these alignments was 29.3 (±95 CI=0.21; range 26–35 taxa). The supermatrix we generated from these data contained 95,274 characters; 26,555 informative sites; and 44,643 site patterns.

Phylogenetic Relationships of Ostariophysan Lineages

A summary of our phylogenetic results is depicted in Figure 3a–c, with a focus on the portion of the trees that were most incongruent between analyses: the relationships of Characiphysi (Gymnotiformes, Siluriformes, and Characiformes).

Morphology

Tree topologies inferred from analysis of morphological data from Fink and Fink (1981, 1996) are identical using Bayesian and ML methods (Supplementary Fig. S1, available on Dryad); all nodes in the phylogeny are supported with bootstrap support values of 100 and Bayesian posterior probabilities of >0.99. This phylogeny is also identical to ordinal relationships resolved based on parsimony analyses by Fink and Fink (1981, 1996; Fig. 1) where Gonorynchiformes is the sister clade to Otophysi (Cypriniformes + (Characiformes + (Siluriformes + Gymnotiformes))). The morphology-based phylogenetic results are notable for representing the only time our analyses recovered a monophyletic Characiformes as the sister lineage of a clade containing reciprocally monophyletic Gymnotiformes and Siluriformes (Fink and Fink 1981; Fig. 3b).

UCEs

RAxML bootstrapping under autoMRE settings for the 50% complete data set converged after 50 bootstrap replicates for both partitioned and unpartitioned analyses, and RAxML bootstrapping under autoMRE for the 75% complete data converged after 100 bootstrap replicates. Bayesian analyses of the partitioned, 50% complete data set reached convergence (ADSF 0.3%; average PSRF=1.0; ±95 CI PSRF=0.0007; average ESS=2338.6; ±95 CI ESS=114.3) after 5,000,000 iterations while Bayesian analysis of the partitioned 75% complete data set (ADSF 0.10%; average PSRF=1.0; ±95 CI PSRF=0.0003; average ESS=2250.5; ±95 CI ESS=134.02) reached convergence after 2.6 million iterations. The PSRF was $\leqslant 1.05$ for 95% of the variables, and an average SD of splits <0.01 were taken as thresholds for convergence.

We ran 50% complete UCE data sets using ML with no partitioning of loci (Supplementary Fig. S2, available on Dryad) as well as partitioned Bayesian (Supplementary Fig. S3, available on Dryad) and ML (Supplementary Fig. S4, available on Dryad) analyses. Partitioning versus nonpartitioning in ML analyses for the 50% complete UCE data set resulted in identical topologies and little change in bootstrap support for nodes having low support. We recovered Characiformes as paraphyletic in all 50% complete Bayesian and ML analyses: Citharinoidei is the sister group to a clade composed of Characoidei + Siluriformes. Gymnotiformes is the sister group to all other Characiphysi.

The concatenated ML (partitioned and not partitioned) and Bayesian UCE 75% complete phylogenies (Supplementary Figs. S5 and S6, available on Dryad) had identical topologies and recovered the outgroup Clupeiformes as the sister group

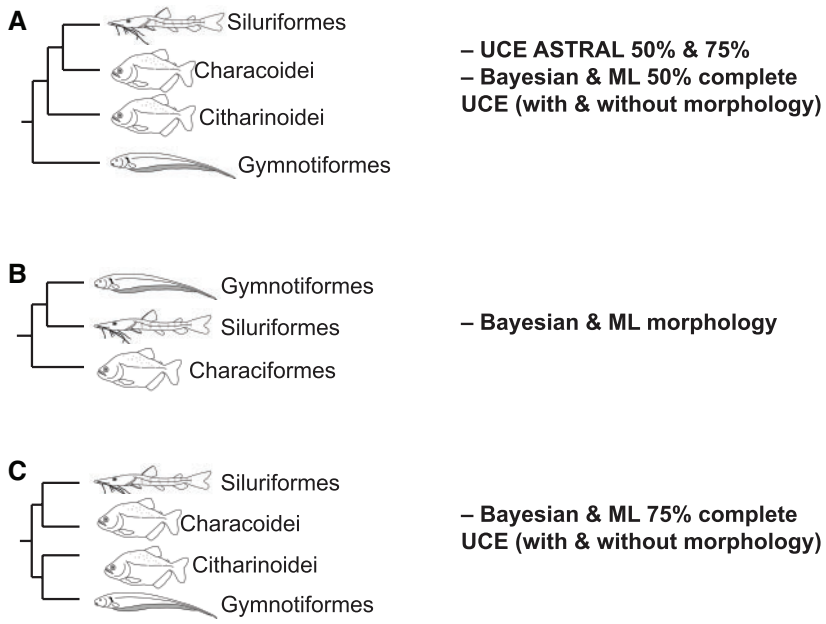

FIGURE 3. Summary of incongruous relationships for major Characiphysi lineages observed between analyses in this study. Labels "A", "B", and "C" refer to the three major categories of results that are referenced in the text.

of a monophyletic Ostariophysi. We recovered a monophyletic Gonorhynchiformes as the sister lineage of the remaining Ostariophysi and a monophyletic Cypriniformes as the sister group of the remaining Otophysi, matching the recovered relationships of Fink and Fink (1981) (Fig. 1). However, the relationships recovered within Characiphysi by the ML and Bayesian 75% complete UCE concatenated analyses are novel. We recovered each of the characiform suborders as monophyletic but not as each other's closest relatives, specifically, we recovered the Citharinoidei is resolved as the sister clade of the Gymnotiformes, and the Characoidei as the sister group of the Siluriformes.

UCE Multispecies Coalescent Approach

Species tree analysis of the 50% and 75% complete UCE data sets produced identical topologies with the exception of *Sternopygus macrurus* being recovered as the sister group to *Apteronotus albifrons* in the 50% complete species tree (Supplementary Fig. S7, available on Dryad) versus the sister group to the remaining Gymnotiformes in the 75% species tree (Fig. 4). Bootstrap values were largely the same across the two trees. Both species tree analyses recovered a non-monophyletic Characiformes due to the placement of Characoidei as the sister lineage of Siluriformes, and Citharinodei being the sister group of that clade (Characoidei + Siluriformes).

We found four internodes in three regions of the species trees that potentially are in the anomaly zone. In these internodes, the descending branches were shorter than the limit of the anomaly zone as determined by $a(x)$. The internodes with lengths likely to produce anomalous gene trees were the same for both the 50% and the 75% complete species tree (shown with dashed lines in Fig. 4 and Supplementary Fig. S7 available on Dryad): the internode between the crown node of Gymnotiformes and the common ancestor with the rest of Characiphysi, the internode leading to the Characoidei, and the two deepest internodes within the Siluriformes.

UCEs + Morphology

RAxML bootstrapping under autoMRE settings for the 50% complete data set plus morphological data from Fink and Fink (1981, 1996) converged after 50 bootstrap replicates for both partitioned and unpartitioned analyses, and RAxML boostrapping under autoMRE for the 75% complete data converged after 150 bootstrap replicates for the partitioned data and 100 boostrap replicates for the unpartitioned data. Bayesian analyses of the partitioned, 50% complete data set reached convergence (ADSF 0.04%; average PSRF = 1.0; ±95 CI PSRF = 0.0004; average ESS = 1570.4; ±95 CI ESS = 74.7) after 5,000,000 iterations and Bayesian analysis of the partitioned 75% complete data set (ADSF 0.03%; average PSRF = 1.0; ±95 CI PSRF = 0.0001; average ESS = 4430.7; ±95 CI ESS = 249.5) reached convergence after 5,000,000 iterations.

The 50% complete UCE + morphology data recovered the same topology with both Bayesian (Supplementary Fig. S8, available on Dryad) and ML analyses (Supplementary Fig. S9, available on Dryad) and the addition of the morphological characters did not change the topology from the results inferred from the 50% complete UCE data alone (Fig. 3a). Support values across approaches were similar.

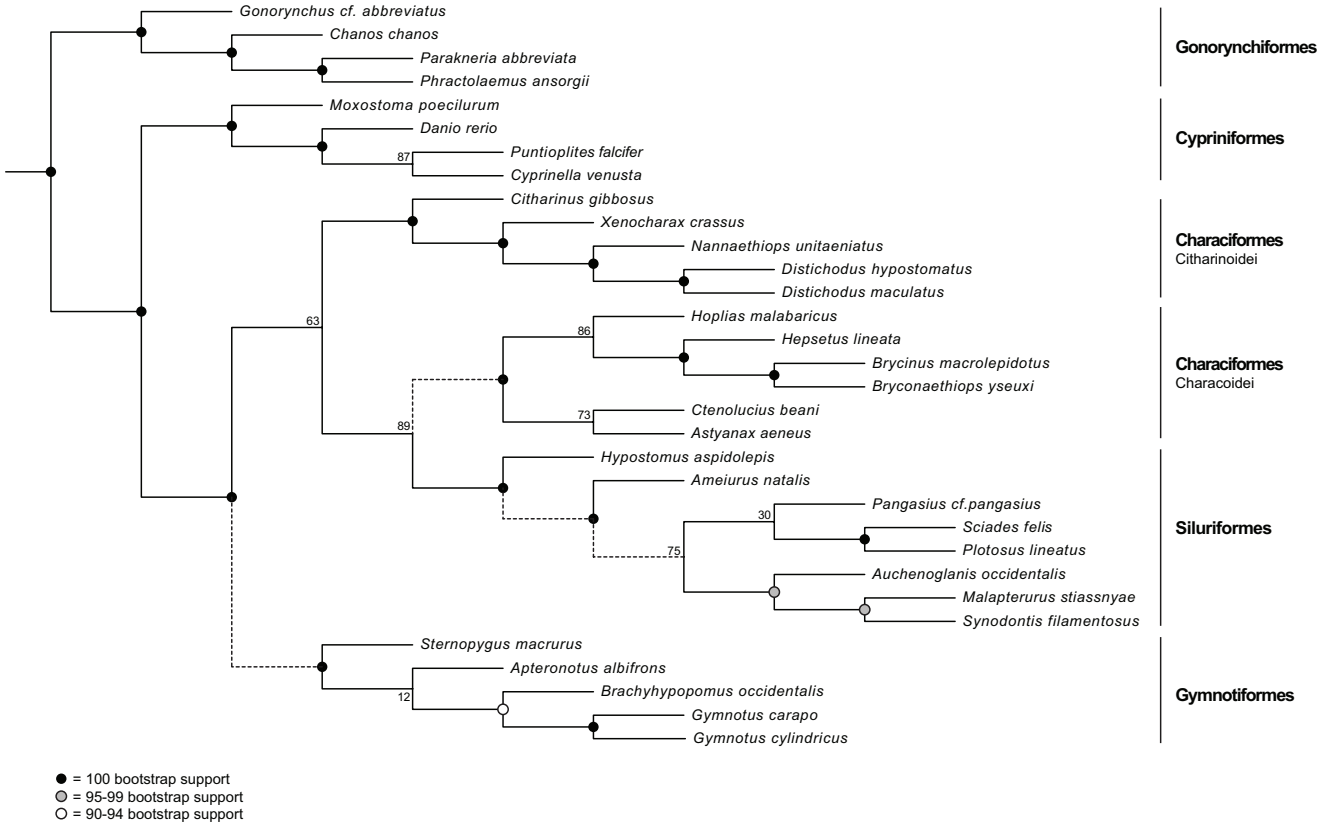

FIGURE 4. ASTRAL phylogeny of ostariophysans using ultraconserved elements with 75% data coverage, using a coalescent species tree approach. Internodes with lengths that were shorter than the limit of $a(x)$ and therefore potentially in the anomaly zone are shown with dashed lines (see text).

The 75% complete UCE + morphology data recovered the same topology for both Bayesian (Supplementary Fig. S10, available on Dryad) and ML analyses (Supplementary Fig. S11, available on Dryad) and again the addition of the morphological characters did not change the topology. Support values across approaches were similar. The ML and Bayesian 75% complete UCE topology (with or without morphology) continued to support a unique result compared with all other analyses of other data sets examined for Ostariophysi (Fig. 3c).

All SH tests for the 50% complete UCE data set (with and without morphology) showed that a topology with a monophyletic Characifomes was significantly worse ($P < 0.01$) than the recovered phylogenies. For the 75% complete UCE data set with morphology, the SH test was not significant; however, it was significant at $P < 0.05$ without morphology.

DISCUSSION

Our examination of ostariophysan relationships using UCEs and morphology recovered the species-poor clade Gonorhynchiformes as the sister group to all remaining ostariophysans (collectively Otophysi), and Cypriniformes as the sister lineage to all other Otophysi (collectively Characiphysi). As with past studies, our analyses of different data sets show that a consensus for the relationships within Characiphysi (which include Characiformes, Siluriformes, and Gymnotiformes; Fig. 1) remains elusive. However, our phylogenetic analyses using UCE loci (with or without morphology) consistently fail to resolve Characiformes as monophyletic. Using a coalescent approach, we identified several internodes as potentially being in the anomaly zone (highlighted in Fig. 4 and Supplementary Fig. S7 available on Dryad), and likely the root cause of discordance between the results from our coalescent and concatenated methods.

Resolution of Ostariophysan Relationships: The Conflict Continues

Depending on analytical method or data set used, our results either mirrored or suggested new potential relationships to the conflicting variety of proposed relationships among major groups of ostariophysans (Figs. 2 and 3). For example, analyses based on the concatenated 50% complete UCE data set (with and without morphology) or the multispecies coalescent approach (with both 50% and 75% complete UCE data sets) recover a paraphyletic Characiformes, with the Citharinoidei as the sister group to Characoidei +

Siluriformes (Fig. 3a). This topology is consistent with most recent studies that have included reasonable taxon sampling (Nakatani et al. 2011; Chen et al. 2013; but see Arcila et al. 2017). In contrast, the phylogeny inferred from the concatenated 75% complete UCE data set (with or without morphology) suggests a novel pattern of higher level relationships that supports Siluriformes as the sister group of Characoidei, and Citharinoidei as the sister group of Gymnotiformes (Fig. 3c). This topology has not been previously recovered.

A potential explanation for this novel topology is the symmetrical arrangement of the major clades of Characiphysi separated by a short internal branch. This arrangement is a hallmark of an anomaly zone (Rosenberg 2013; Linkem et al. 2016) as the probability of producing anomalous gene trees increases as parental branches x get shorter, and the limit of the anomaly zone $a(x)$ approaches infinity (Degnan and Rosenberg 2006; Linkem et al. 2016). For a group as large as Ostariophysi, taxon sampling represents a particularly important consideration for resolution of nodes in anomaly zones. Previous studies have generally sampled relatively few taxa (<1% of the 10,000+ species); with members of Citharinoidei particularly poorly represented in most data matrices (something we tried to remedy with our sampling efforts). Although our sampling of species is similarly limited, our selection does include representatives of most major clades of both Citharinoidei and Characoidei. This distinction is important because sampling taxa that diverge near the node of interest, in this case the most recent common ancestor of Citharinoidei and Characoidei, is expected to produce the largest potential change in tree topology (Townsend and Lopez-Giraldez 2010). Simply sampling additional taxa, all of which diverge far from the MRCA of Citharinoidei or Characoidei, is therefore unlikely to impact topological reconstruction concerning the monophyly of Characiformes (Townsend and Lopez-Giraldez 2010; Fig. 4). The pace of species discovery (both fossil and extant) in ostariophysan fishes remains high (Almirón et al. 2015; Armbruster et al. 2015; Carvalho and Albert 2015), raising the possibility that additional deeply divergent lineages are yet to be discovered (Albert et al. 2011). Although this possibility remains speculative, such organisms would be incredibly valuable for resolving the evolutionary relationships of this clade with confidence.

Short, deep internodes are among the hardest phylogenetic problems (Townsend et al. 2012), and can present an exceptional challenge for genomic-scale phylogenetic analyses (Jeffroy et al. 2006; Philippe et al. 2011; Townsend 2012; Anisimova et al. 2013; Posada 2013; Romiguier et al. 2013; Eytan et al. 2015) as is reflected in our analyses of the different UCE data sets. For example, the difference between the 50% and 75% complete UCE data may be driven partially by heterogeneity in substitution rates, as the 75% complete data should have fewer rapidly evolving markers than the 50% complete data (which are less conserved). The 50% complete data set agreed with most of our reconstructions of the relationships of Characiphysi (Fig. 3a), the fact that the concatenated 75% complete ML and Bayesian analyses produce a novel result is notable. Unfortunately, there have been few studies investigating the behavior of UCE data (Romiguier et al. 2013) and a thorough investigation of this data set is outside the scope and aim of this study. Regardless of which factors underlie the conflict between data sets, our results are in agreement with the increasingly common conclusion that more data (particularly next-generation sequence data) does not always lead to more stable or robust conclusions (Rodríguez-Ezpeleta et al. 2007; Philippe et al. 2011; Eytan et al. 2015), suggesting we are still far from achieving definitive resolution of ostariophysans.

The recovery of Characiformes as non-monphyletic, both in the various forms presented here and in other independent studies, provides support for the legitimacy of this topological hypothesis; however, a recent study by Arcila et al. (2017) provides an alternative view. Using over 1500 exons, Arcila et al. (2017) recovered the traditional relationships presented in the morphological work of Fink and Fink (1981) using a new approach called gene genealogy interrogation (GGI). This approach which chooses the gene tree topologies that most frequently support one partially constrained tree (out of a set of alternatives). The argument for advancing the GGI method is that gene tree estimation error rather than incomplete lineage sorting is the dominant hurdle in studies of deep-time phylogenomic studies. While theory predicts gene tree estimation errors to be particularly problematic for small, deep internodes, the phylogenetic information content of slow-evolving neutral markers such as UCE's have not been investigated. An expanded discussion about when it is appropriate to select among different phylogenetic signals when there is extensive "gene tree error" is a needed area of future research.

Issues of Orthology/Paralogy

Paralogy and orthology have long been central issues in molecular phylogenetics, and molecular data obtained from massively parallel sequencing has amplified this problem because many software programs intended to examine paralogy do not perform well with large data matrices (e.g. Bousseau et al. 2013). Our data selection criteria removed multiple copies of duplicate UCE loci to avoid using non-homologous loci in phylogenetic analyses. However, orthology/paralogy issues may affect next-generation sequencing studies, particularly when the taxonomic group under study has undergone whole genome duplication events, such as in teleost fishes (Christoffels et al. 2004). In teleost fishes, such as ostariophysans, it is possible that we may be examining paralogous loci instead of homologous loci if only a single copy of a locus actually present in multiple copies is recovered during data processing. Assuming that this problem is widespread, and paralagous copies of a given locus are aligned in a matrix, this error could result in inaccurate phylogenetic reconstruction

(Eytan et al. 2015). No current methods exist to determine if such problematic alignments of non-orthologous loci have been accidently included in an analysis of teleosts. For example, using a BLAST search of target loci against a known genome to indicate areas multiple loci may be prevalent (a technique used mainly in exon capture analyses—see Ilves and López-Fernández 2014) is potentially highly misleading given the few genomes available for Ostariophysans. Currently, only four reference genomes are available for the group (*D. rerio*, *Astyanax mexicanus*, *E. electricus*, and *Ictalurus punctatus*). However, considerable genomic diversity including repeated losses of entire gene families or duplications of individual loci are a common feature of teleost genome evolution (Star et al. 2011; Malmstrøm et al. 2016) that can mislead a detailed assessment of gene synteny. Our approach of deleting loci observed to have multiple variable copies is commonly used in phylogenomic pipelines (see also McCormack et al. 2013) and further investigation into the orthology/paralogy issue in organisms with duplicated genomes is an important avenue of research for future phylogenomic studies.

Molecules, Morphology, and Characiform Relationships

Assembling morphological evidence for phylogenetic relationships of species-rich and ancient clades such as those examined here is among the most arduous tasks in systematics (Wiley et al. 2011; Giribet 2015). Characiphysi, in particular, is a difficult group to study using phenotypic traits alone. This species-rich group includes all catfishes, knifefishes, characins, and citharinoids, encompassing exceptional phenotypic disparity and species diversity and dominating continental freshwater ecosystems across most of the globe. While Characiphysi appears to have diverged rapidly prior to the final break-up of Gondwana during the Upper Cretaceous (Lundberg 1993; Buckup 1998; Malabarba and Malabarba 2010), known fossils provide few insights into the phylogenetic relationships of major characiphysan lineages (López–Fernández and Albert 2011). This is largely due to the fact that early members of Characiphysi may have resembled the oldest known otophysan fossil *Santanichthys diasii* from the early Cretaceous,which closely resembles a generalized characiform in general body form (Filleul and Maisey 2004). Despite this difficulty, the data set of Fink and Fink (1981) has been argued to provide strong support for ostariophysan relationships and remains highly regarded as one of the most comprehensive morphological data sets of its kind (Chen et al. 2013).

While strong signal from a morphological character matrix can change topological relationships in combined phylogenetic analyses (Nylander et al. 2004; Wiens et al. 2010), characiform monophyly is only recovered when analyzing the Fink and Fink (1981) morphological data set in the absence of molecular data (Fig. 3b). Notably, the only instance in which molecular data have ever recovered a monophyletic Characiformes is when this morphological data set (from Fink and Fink 1981) was also included (Dimmick and Larson 1996), when few (two) members of the characiform suborder Citharinoidei were included (Betancur-R et al. 2013), or when a data parsing approach was applied (Arcila et al. 2017). In contrast, all other molecular phylogenetic studies support the non-monophyly of characiforms (summarized in Figs. 2 and 3).

In our analyses, the addition of morphological characters to UCE data did not change the tree relative to topologies produced by UCE data analyzed alone (Supplementary Figs. S9–S12, available on Dryad). Because the impact of morphological characters in combined-data phylogenetic inferences is proportional to the relative information content of each data set, the sheer size of the UCE data sets (each including many hundreds of loci) are possibly numerically overwhelming the information content of the morphological data (Nylander et al. 2004). A separate parsimony analysis of the UCE + morphology (not shown) found that the tree constrained to recover a monophyletic Characiformes was 47 steps longer than the unconstrained tree (which recovered the same topology as that from the UCE data alone). This high number of additional steps needed to support characiform monophyly might explain why support values were relatively unchanged with the addition of morphological traits. However, it should be noted that the SH test results for the 75% complete UCE data set with morphological characters included did not show a significant difference between the likelihood score of the recovered ML phylogeny and the ML phylogeny which recovered a monophyletic Characiformes (all other SH test comparisons were significant), demonstrating the influence of morphological characters even in a phylogenomic analysis. Further, we argue that morphological data will become anything but obsolete with the widespread use of data-rich next-generation sequencing approaches (i.e., phylogenomics). For example, studies using fossil tip dating (Pyron 2011; Ronquist et al. 2012; Chen et al. 2015) to integrate extinct and extant species into a time-calibrated Tree of Life strongly illustrate the utility of morphological data in addition to large molecular data sets (Wood et al. 2013; Dornburg et al. 2015a, 2015b; Giribet 2015; Hsiang et al. 2015) suggesting an exciting frontier for ostariophysan phylogenetics.

The Evolutionary History of Electroreception

The phylogenetic results obtained here are consistent with several interpretations regarding the evolution of passive electroreception. One interpretation is that the many derived osteological, neurological, and physiological traits shared by extant catfishes and gymnotiform knifefishes arose as convergent adaptations to nocturnal and benthic habitats. Indeed, a very similar form of passive electroreception evolved in a distantly related group of teleost fishes in the Old World, Notopteroidei (members of Osteoglossiformes), a clade that includes

Asian and African knifefishes (Notopteridae), and the exclusively African Mormyroidea.

The passive electroreceptive system of Notopteroidei resembles that of Siluriformes and Gymnotiformes in many aspects, including similar hair-cell electroreceptor cells, ampullary-shaped electroreceptor organs, laterosensory pathways in the peripheral and central nervous system, and the embryological tissues giving rise to these electrosensory structures (see reviews in Albert and Crampton 2005; Freitas et al. 2006). However, all these attributes are shared by other vertebrate laterosensory systems, including the mechanosensory (vibration sensitive) lateral line systems present in all teleost fishes. In other words, these traits are perhaps best interpreted as plesiomorphic, rather than convergent (Finger et al. 1986).

The passive electrosensory systems of Siluriformes and Gymnotiformes also share many additional and uniquely derived traits that distinguish them from both the notopteroid and the more general vertebrate passive electrosensory systems. These traits include subtle but functionally important details of neural architecture, cytology, and physiology (Fink and Fink 1996; Albert et al. 1998). Therefore, although the passive electrosensory systems of Siluriformes and Gymnotiformes may in fact be convergent, such a hypothesis would require substantially more evolutionary work than that of a single evolutionary transformation. Other interpretations are possible. The similar morphologies of Characoidei and Citharinoidei might represent instances of convergent evolution, or the retention of pleisiomorphic traits of Characiphysi. Resolution of these questions will require additional morphological study of both extant and fossil forms. Another interpretation may be that the morphology is providing the true signal of relationships of these fishes, and the molecular data are not. We cannot falsify this later statement, and without corroborative data about their evolutionary relationships we may never know the "true tree" of Ostariophysi. Although the failure of our phylogenomic data to unequivocally resolve this long-standing controversy in ichthyological systematics is frustrating, we are hopeful that this study and others will help to usher in a new age of discovery for the phenotypic traits needed to further our understanding of problematic regions on the tree of life.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.n9g60.

Funding

This work was supported by National Science Foundation (NSF) in computational portions (data processing) [DEB-1242260 to B.C.F.; DEB-0614334, DEB-0741450, and DEB-1354511 to J.S.A.; DEB-0916695 and DEB-1354149 to P.C.]. Portions of this research (phylogenetic analyses) were conducted with high-performance computing resources provided by Louisiana State University (http://www.hpc.lsu.edu). New samples collected by Louisiana State University personnel followed IACUC Protocol #09-002.

Acknowledgments

We thank our many field collaborators in Central and South America, Asia, and Africa. Drs. Bill and Sara Fink helped with previous drafts of the manuscript.

Data Availability

UCE raw read and assembled contig data are available from the NCBI SRA and GenBank (PRJNA363064). PHYLUCE source code is available under an open source license from https://github.com/faircloth-lab/phyluce/.

References

Aberer A.J., Kobert K., Stamatakis A. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31:2553–2556.

Albert J.S. 2001. Species diversity and phylogenetic systematics of American knifefishes (Gymnotiformes, Teleostei). Misc. Pub. UMMZ 190:1–127.

Albert J.S., Crampton W.G.R. 2005. Electroreception and electrogenesis. In: Evans D., editor. The physiology of fishes. New York (NY): CRC. p. 429–470.

Albert J.S., Lannoo M.J., Yuri T. 1998. Testing hypotheses of neural evolution in gymnotiform electric fishes using phylogenetic character data. Evolution 52:1760–1780.

Albert J.S., Petry P., Reis R.E. 2011. Major biogeographic and phylogenetic patterns. In: Albert J.S., Petry P., Reis R.E., editors. Historical biogeography of neotropical freshwater fishes. Berkeley: University of California Press. p. 21–58.

Almirón A., Casciotta J., Piálek L., Doubnerová K., Říčan O. 2015. *Oligosarcus amome* (Ostariophysi: Characidae), a new species from the río Uruguay basin, Misiones, Argentina. Zootaxa 3915:581–590.

Anisimova M., Liberles D.A., Philippe H., Provan J., Pupko T., von Haeseler A. 2013. State-of the art methodologies dictate new standards for phylogenetic analysis. BMC Evol. Biol. 13:161.

Arcila D., Ortí G., Vari R., Armbruster J.W., Stiassny M.L.J., Ko K.D., Sabaj M.H., Lundberg J., Revell L.J., Betancur R., R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat. Ecol. Evol. 1:0020. doi:10.1038/s41559-016-0020.

Armbruster J.W., Werneke D.C., Tan M. 2015. Three new species of saddled loricariid catfishes, and a review of *Hemiancistrus*, *Peckoltia*, and allied genera (*Siluriformes*: Loricariidae). ZooKeys 480:97–123.

Arroyave J.A., Denton J.S.S., Stiassny M.L.J. 2013. Are characiform fishes Gondwanan in origin? Insights from a time-scaled molecular phylogeny of the Citharinoidei (Ostariophysi: Characiformes). PLoS One 8:e77269.

Berg L.S. 1912. Faune de la Russie et des pays limitrophes. Poissons (Marsipobranchii et Pisces). Vol. III. Ostariophysi. Part 1. St. Petersbourg, 336 pp., Pls. 1–2 [in Russian].

Betancur-R R., Broughton R.E., Wiley E.O., Carpenter K., Lopez J.A., Li C., Arratia G., Holcroft N.I., Arcila D., Sanciangco M., Cureton J., Zhang F., Buser T., Campbell M., Rowley T., Ballesteros J.A., Lu G., Grande T., Ortí G. 2013. The tree of life and a new classification of bony fishes. PLoS Curr. Tree of Life. doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

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

Braun C.B., Grande T. 2008. Evolution of peripheral mechanisms for the enhancement of sound reception. In: Popper A.N., Fay R.R., Webb J.L., editors. Springer Handbook of Auditory Research: Fish Bioacoustics. New York: Springer-Verlag. p. 99–144.

Broughton R.E. 2010. Phylogeny of teleosts based on mitochondrial DNA sequences. In: Nelson J.S., Schultze H.P., Wilson M.V.H., editors. Origin and phylogenetic interrelationships of teleosts. München: Verlag Dr. Friedrich Pfeil. p. 61–76.

Broughton R.E., Betancur R., R., Li, Chenhong, Arratia G., Ortí G. 2013. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. Tree of Life. doi:10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e.

Boussau B., Szollosi G.J., Duret L., Gouy M., Tannier E., Daubin V. 2013. Genome-scale coestimation of species and gene trees. Genome Res. 2:323–330. doi: 10.1101/gr.141978.112.

Buckup P.A. 1998. Relationships of the Characidiinae and phylogeny of characiform fishes (Teleostei: Ostariophysi) In: Malabarba L., Reis R., Vari R., Lucena Z.M.S., Lucena C., editors. Phylogeny and classification of neotropical fishes. Porto Alegre: EDIPUCRS. p. 123–143.

Carvalho T.P., Albert J.S. 2015. A new species of *Rhamphichthys* (Gymnotiformes: Rhamphichthyidae) from the Amazon Basin. Copeia 103:34–41.

Chakrabarty P., Warren M., Page L., Baldwin C. 2013. GenSeq: an updated nomenclature and ranking for genetic sequences from type and non-type sources. ZooKeys 346:29–41.

Chen, M-Y., Liang D., Zhang P. 2015. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. Syst. Biol. 64:1104–1120.

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

Chivers D.P., Wisenden B.D., Hindman C.J., Michalak T.A., Kusch R.C., Kaminskyj S.G.W., Jack K.L., Ferrari M.C.O., Pollock R.J., Halbgewachs C.F., Pollock M.S., Alemadi S., James C.T., Savaloja R.K., Goater C.P., Corwin A., Mirza R.S., Kiesecker J.M., Brown G.E., Adrian J.C. Jr, Krone P.H., Blaustein A.R., Mathis A. 2007. Epidermal 'alarm substance' cells of fishes maintained by non-alarm functions: possible defense against pathogens, parasites and UVB radiation. Proc. Roy. Soc. Lond. B 274:2611–2619.

Christoffels A., Koh E.G., Chia, J.-M., Brenner S., Aparicio S., Venkatesh B. 2004. *Fugu* genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol. Biol. Evol. 21:1146–1151.

Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8: 783–786.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68.

DePristo M.A., Banks E., Poplin R., Garimella K.V., Maguire J.R., Hartl C., Philippakis A.A., del Angel G., Rivas M.A., Hanna M., McKenna A., Fennell T.J., Kernytsky A.M., Sivachenko A.Y., Cibulskis K., Gabriel S.B., Altshuler D., Daly M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Dimmick W.W., Larson A. 1996. A molecular and morphological perspective on the phylogenetic relationships of the Otophysan fishes. Mol. Phylogenet. Evol. 6:120–133.

Dornburg A., Moore J., Beaulieu J.M., Eytan R.J., Near T.J. 2015a. The impact of shifts in marine biodiversity hotspots on patterns of range evolution: evidence from the Holocentridae (squirrelfishes and soldierfishes). Evolutio 69:146–161.

Dornburg A., Friedman M., Near T.J. 2015b. Phylogenetic analysis of molecular and morphological data highlights uncertainty in the relationships of fossil and living species of Elopomorpha (Actinopterygii: Teleostei). Mol. Phylogenet. Evol. 89:205–218.

Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Eschmeyer W.N., Fong J.D. 2015. Species by Family/Subfamily. Available from: http://researchcalacademyorg/research/ichthyology/catalog/SpeciesByFamilyasp (accessed 25 March 2015).

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 Evol. Biol. 15:113.

Faircloth B.C., Glenn T.C. 2012. Not all sequence tags are created equal: Designing and validating sequence identification tags robust to indels. PLoS One 7:e42543.

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. 61:717–726.

Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8:e65923. doi:10.1371/journal.pone.0065923.

Faircloth B.C., Branstetter M.G., White N.D., Brady S. 2014. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. 15:489–501.

Filleul A., Maisey J.G. 2004. Redescription of *Santanichthys diasii* (Otophysi, Characiformes) from the Albian of the Santana Formation and comments on its implications for Otophysan relationships. Am. Mus. Novit. 3455:1–21.

Fink S.V., Fink W.L. 1981. Interrelationships of the ostariophysan fishes (Teleostei). Zool. J. Linn. Soc. 72: 297–353.

Fink S.V., Fink W.L. 1996. Interrelationships of ostariophysan fishes (Teleostei). In: Stiassny M.L.J., Parenti L.R., Johnson G.D. editors. Interrelationships of fishes. San Diego (CA): Academic Press. p. 209–249.

Finger T.E., Bell C.C., Carr C.E. 1986. Comparisons among electroreceptive teleosts: why are electrosensory systems so similar. In: Electroreception, Bull T.H., Heiligenberg W. editors. New York: Wiley. p. 465–481.

Fisher S., Barry A., Abreu J., Minie B., Nolan J., Delorey T.M., Young G., Fennell T.J., Allen A., Ambrogio L., Berlin A.M., Blumenstiel B., Cibulskis K., Friedrich D., Johnson R., Juhn F., Reilly B., Shammas R., Stalker J., Sykes S.M., Thompson J., Walsh J., Zimmer A., Zwirko Z., Gabriel S., Nicol R., Nusbaum C. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol. 12:R1.

Friedman M., Sallan L.C. 2012. Five hundred million years of extinction and recovery: a Phanerozoic survey of large-scale diversity patterns in fishes. Palaeontology 55:707–742.

Freitas R., Zhang G., Albert J.S., Evans D.H., Cohn M.J. 2006. Developmental origin of shark electrosensory organs. Evol. Dev. 8:74–80.

Frisch K.V. 1942. Über einen Schreckstoff der Fischhaut und seine biologische Bedeutung. Zeitschrift für vergleichende. Physiologie 29:46–145.

Gilbert P.S., Chang J., Pan C., Sobel E.M., Sinsheimer J.S., Faircloth B.C., Alfaro M.E. 2015. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. Mol. Phylogenet. Evol. 92:140–146.

Giribet G. 2015. Morphology should not be forgotten in the era of genomics–a phylogenetic perspective. Zool. Anz. 256:96–103.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raktima R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–654.

Greenwood P.H., Rosen D.E., Weitzman S.H., Myers G.S. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. 131:339–456.

Harrington R.C., Faircloth B.C., Eytan R.I., Smith W.L., Near T.J., Alfaro M.E., Friedman M. 2016. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the evolutionary eye. BMC Evol. Biol. 16:224.

Helfman G.S., Collette B.B., Facey D.E., Bowen B.W. 2009. The diversity of fishes: biology, evolution, and ecology. 2nd ed. London, England: Wiley-Blackwell.

Hsiang A.Y., Field D.J., Webster T.H., Behlke A.D.B., Davis M.B., Racicot R.A., Gauthier J.A. 2015. The origin of snakes: revealing the ecology,

behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. BMC Evol. Biol. 15:87.

Ilves K.L., López-Fernández H. 2014. A targeted next-generation sequencing toolkit for exon-capture based cichlid phylogenomics. Mol. Ecol. Res. 14: 802–811.

Jeffroy O., Brinkmann H., Delsuc F., Philippe H. 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22:225–231.

Kappas I., Vittas S., Pantzartzi C.N., Drosopoulou E., Scouras Z.G. 2016. A time-calibrated mitogenome phylogeny of catfish (Teleostei:Siluriformes) PLoS One 11:e0166988 doi: 10.1371/journal.pone.0166988

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

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

Kubatko K., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17-24.

Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses.Mol. Biol. Evol. 29:1695–1701.

Lanfear R., Calcott B., Kainer D., Mayer C., Stamatakis A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14:82.

Lavoué S., Miya M., Inoue J.G., Saitoh K., Ishiguro N.B., Nishida M. 2005. Molecular systematics of the gonorynchiform fishes (Teleostei) based on whole mitogenome sequences: Implications for higher-level relationships within the Otocephala. Mol. Phylogenet. Evol. 37:165–177.

Li C., Lu G., Ortí G. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Syst. Biol. 57:519–539.

Li H., Handsaker B., Wysoker A., Fennell T. Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN].

Liu L., Edwards S.V. 2009. Phylogenetic analysis in the anomaly zone. Syst. Biol. 58:452–460.

Linkem C.W., Minin V.N., Leaché A.D. 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). Syst. Biol. 65:465–477.

Liu D.-W., Lu Y., Yan H.Y., Zakon H.H. 2016. South American weakly electric fish (Gymnotiformes) are long-wavelength-sensitive cone monochromats. Brain Behav. Evol. 88:204–212.

Liu K., Raghavan S., Nelesen S., Linder C.R., Warnow T. 2009. Rapid and accurate large scale coestimation of sequence alignments and phylogenetic trees. Science. 324:1561–1564.

Liu K., Warnow T.J., Holder M.T., Nelesen S., Yu J., Stamatakis A. Linder C.R. 2012. SATé-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. Syst. Biol. 61:90–106.

López–Fernández H., Albert J.S. 2011. Paleogene radiations. In: Albert J.S., Petry P., Reis R.E., editors. Historical biogeography of neotropical freshwater fishes. Berkeley (CA): ~~Fishes~~ University of California Press. p. 105–118.

Lundberg J.G. 1993. African-South American freshwater fish clades and continental drift: problems with a paradigm. In: Goldblatt P., editor. Biological relationships between Africa and South America. New Haven (CT): Yale University Press. p. 156–199.

Malabarba L.R., Malabarba M.C. 2010. Biogeography of Characiformes: an evaluation of the available information of fossil and extant taxa. In: Nelson J.S., Schultze H.-P., Wilson M.V.H., editors. Origin and phylogenetic interrelationships of teleosts. ~~München~~: Verlag Dr. Friedrich Pfeil. p. 317–336.

Maniatis T., Sambrook J., Fritisch E.F. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbour (NY): Cold Spring Harbour Laboratory.

Marçais G. Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27: 764–770.

Malmstrøm M., Matschiner M., Tørresen O.K., Star B., Snipen L.G., Hansen T.F., Baalsrud H.T., Nederbragt A.J., Hanel R., Salzburger W., Stenseth N.C., Jakobsen K.S., Jentoft S. 2016. Evolution of the immune system influences speciation rates in teleost fishes. ~~Nature~~ Genetics 48: 1204–1210.

McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22:746–754.

McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., Brumfield R.T. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66:526–538.

McGee M.D., Faircloth B.C., Borstein S.R., Zheng J., Darrin Hulsey C., Wainwright P.C., Alfaro M.E. 2016. Replicated divergence in cichlid radiations mirrors a major vertebrate innovation. Proc. Biol. Sci. 283. http://rspb.royalsocietypublishing.org/content/royprsb/283/1822/20151413.full.pdf.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–303.

Mirarab S., ~~Tandy W.~~ 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Nakatani M., Miya M., Mabuchi K., Saitoh K., Nishida M. 2011. Evolutionary history of Otophysi (Teleostei), a major clade of the modern freshwater fishes: Pangaean origin and Mesozoic radiation. BMC Evol. Biol. 11:177.

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. 109:13698–13703.

Near T.J., Dornberg A., Eytan R.I., Keck B., 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. Natl. Acad. Sci. USA. 110: 12738–12743.

Nelson J.S. 2006. Fishes of the world. 4th ed. Hoboken: John Wiley.

Novacek M.J., Marshall L.G. 1976. Early biogeographic history of ostariophysan fishes. Copeia 1976:1–12.

Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53: 47–67.

Oliveira C., Avelino G.S., Abe K.T., Mariguela T.C., Benine R.C., Ortí G., Vari R.P., Corrêae Castro R. 2011. Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling. BMC Evol. Biol. 11:275.

Ortí G., Meyer A. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. Mol. Biol. Evol. 13:556–573.

Ortí G., Meyer A. 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst. Biol. 46:75–100.

Philippe H., Brinkmann H., Lavrov D.V., Littlewood T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.

Posada D. 2013. Phylogenetic models of molecular evolution: next-generation data, fit, and performance. J. Mol. Evol. 76:351–352.

Poulsen J.Y., Moller P.R., Lavoué S., Knudsen M., Nishida M. 2009. Higher and lower-level relationships of the deep-sea fish order Alepocephaliformes (Teleostei: Otocephala) inferred from whole mitogenome sequences. Biol. J. Linn. Soc. 98:923–936.

Pyron R.A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466–481.

Rabosky D.L., Santini F., Eastman J., Smith S.A., Sidlauskas B., Chang J., Alfaro M.E. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat. Commun. 4:1958.

Rambaut A., Suchard M.A., Xie D., Drummond A.J. 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer. Last accessed: 2/1/2017.

Rodríguez-Ezpeleta N.,Brinkmann H., Roure B., Lartillot N. Lang B.F., Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst. Biol. 56:389–399.

Rohland N., Reich D. 2012. Cost-effective, high throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22:939–946.

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. Evol. 30:213–2144

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Hohna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rosen D.E., Greenwood P.H. 1970. Origin of the Weberian apparatus and the relationships of the ostariophysan and gonorynchiform fishes. Am. Mus. Novit. 2428:1–25.

Rosen D.E. 1973. Interrelationships of higher euteleostean fishes. Zool. J. Linn. Soc. 53(Suppl 1): 397–513.

Rosenberg N.A. 2013. Discordance of species trees with their most likely gene trees: a unifying principle. Mol. Biol. Evol. 30:2709–2713.

Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log–likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D., Pérez-Emán J., Burney C.W., Xie X., Harvey M.G, Faircloth B.C., Glenn T.C., Derryberry E.P., Prejean J., Fields S., Brumfield R.T. 2014. The drivers of tropical speciation. Nature 515:406–409.

Stamatakis A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.

Star B., Nederbragt A.J., Jentoft S., Grimholt U., Malmstrøm M., Gregers T.F., Rounge T.B., Paulsen J., Solbakken M.H., Sharma A., Wetten O.F., Lanzén A., Winer R., Knight J., Vogel J.H., Aken B., Andersen O., Lagesen K., Tooming-Klunderud A., Edvardsen R.B., Tina K.G., Espelund M., Nepal C., Previti C., Karlsen B.O., Moum T., Skage M., Berg P.R., Gjøen T., Kuhl H., Thorson J. Malde K., Reinhardt R., Du L., Johansen S.D., Searle S., Lien S., Nilsen F., Jonassen I., Omholt S.W., Stenseth N.C., Jakobsen K.S. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477: 207–210.

Stout C.C., Tan M., Lemmon A.R., Lemmon E.M., Armbruster J.W. 2016. Resolving Cypriniformes relationships using an anchored enrichment approach. BMC Evol. Biol. 16:244.

Sukumaran J., Holder M.T. 2010. DendroPy: A Python library for phylogenetic computing. Bioinformatics 26: 1569–1571.

Thacker C.E., Satoh T.P., Katayama E., Harrington R.C., Eytan R.I., Near T.J. 2015. Molecular phylogeny of Percomorpha resolves *Trichonotu*s as the sister lineage of Goioidei (Teleostei: Gobiiformes) and confirms the polyphyly of Trachinoidei. Mol. Phylogenet. Evol. 93:172–179.

Townsend J.P., Lopez-Giraldez F. 2010.Optimal selection of gene and ingroup taxon sampling for resolving phylogenetic relationships. Syst. Biol. 59:44–457.

Townsend J.P., Su Z. Tekle Y.I. 2012. Phylogenetic signal and noise: Predicting the power of a dataset to resolve phylogeny. Syst. Biol. 61:835–849.

Van der Laan R., Eschmeyer W.N., Fricke R. 2014. Family-group names of recent fishes. Zootaxa 3882:1–230.

Van der Auwera G.A., Carneiro M., Hartl C., Poplin R., del Angel G., Levy-Moonshine A., Jordan T., Shakir K., Roazen D., Thibault J., Banks E., Garimella K., Altshuler D., Gabriel S., DePristo M. 2013. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43:11.10.1–11.10.33.

Wagner C.E., Keller I., Wittwer S. Selz O.M., Mwaiko S., Greuter L., Sivasundar A. Seehausen O. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22:787–798.

Wainwright P.C., Smith W.L., Price S.A., Tang K.L., Sparks J.S., Ferry L.A., Kuhn K.L., Eytan R.I., Near T.J. 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst. Biol. 61:1001–1027.

Wainwright P.C., McGee M.D., Longo S.J., Hernandez L.P. 2015. Origins, innovations, and diversification of suction feeding in vertebrates. IntegrComp Biol. 55:134–145.

Wheeler T.J. Kececioglu J.D. 2007. Multiple alignment by aligning alignments. Bioinformatics 23:i559–i568.

Wiens J.J., Kuczynski C.A., Townsend T., Reeder T. W., Mulcahy D.G., Sites J.W. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. Syst. Biol. 59:674–688.

Wiley E.O., Chakrabarty P., Craig M.T., Davis M.P., Holcroft N.I., Mayden R.L., Smith W.L. 2011. Will the real phylogeneticists please stand up? Zootaxa 2946:7–16.

Wood H. M., Matzke N.J., Gillespie R.G., Griswold C.E. 2013. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. Syst. Biol. 62: 264–284.

Yan H.Y., Fine M.L., Horn N.S., Colón W.E. 2000. Variability in the role of the gasbladder in fish audition. J. Comp. Physiol. A 186:595–602.

Yu J. Holder M.T., Sukumaran J., Mirarab S., Oaks J. 2013. "SATé version 2.2.7 Available from http://phylo.bio.ku.edu/software/sate/sate.html" (accessed 15 February 2013).