A peer-reviewed version of this preprint was published in PeerJ on 8 September 2017.

View the peer-reviewed version (peerj.com/articles/3548), which is the preferred citable publication unless you specifically need to cite this preprint.

Campbell MA, Alfaro ME, Belasco M, López JA. 2017. Early-branching euteleost relationships: areas of congruence between concatenation and coalescent model inferences. PeerJ 5:e3548
https://doi.org/10.7717/peerj.3548

PeerJ Preprints NOT PEER-REVIEWED

Early-branching euteleost relationships: Areas of congruence between concatenation and coalescent model inferences

Matthew A Campbell Corresp., [1, 2] , **Michael E Alfaro** [3] , **Max Belasco** [3] , **J. Andres Lopez** [4, 5]

[1] Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California, United States of America
[2] Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, California, United States of America
[3] Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, California, United States of America
[4] School of Fisheries and Ocean Sciencs, University of Alaska Fairbanks, Fairbanks, Alaska, United States of America
[5] University of Alaska Museum, Fairbanks, Alaska, United States of America

Corresponding Author: Matthew A Campbell
Email address: drmaccampbell@gmail.com

Phylogenetic inference based on evidence from DNA sequences has led to significant strides in the development of a stable and robustly supported framework for the vertebrate tree of life. To date, the bulk of those advances have relied on sequence data from a small number of genome regions that have proven unable to produce satisfactory answers to consistently recalcitrant phylogenetic questions. Here, we re-examine phylogenetic relationships among early-branching euteleostean fish lineages classically grouped in the Protacanthopterygii using DNA sequence data surrounding ultraconserved elements. We report and examine a dataset of thirty-four OTUs with 17,957 aligned characters from fifty-three nuclear loci. Phylogenetic analysis is conducted both in concatenated and joint gene trees and species tree estimation frameworks. Both analytical frameworks yield supporting evidence for existing hypotheses of relationship for the placement of *Lepidogalaxias salamandroides*, monophyly of the Stomiatii and the presence of an esociform + salmonid clade. *Lepidogalxias salamandroides* and the Esociformes + Salmoniformes are successive sister lineages to all other euteleosts in the two analysis types receiving high support values for this arrangement. However, inter-relationships of Argentiniformes, Stomiatii and Neoteleostei remain uncertain as they varied by analysis type while receiving strong and contradictory indices of support. Topological differences between analysis types are apparent within the Ostarioclupeomorpha and the percomorph taxa in the data set. Our results identify concordant areas with strong support for relationships within and between early-branching euteleost lineages but they also reveal limitations in the ability of larger datasets to conclusively resolve other aspects of that phylogeny.

PeerJ Preprints NOT PEER-REVIEWED

Title: Early-branching euteleost relationships: Areas of congruence between concatenation and coalescent model inferences

Authors and Affiliations:

Matthew A. Campbell[a,1*], Michael E. Alfaro[b], Max Belasco[b], J. Andrés López[c,d]

[a] *Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK 99775 USA*

[b] *Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA 90095 USA*

[c] *School of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK 99775, USA*

[d] *University of Alaska Museum, Fairbanks, AK 99775, USA*

*Corresponding Author: DrMacCampbell@gmail.com

[1] Present Address: *Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95064, USA*

Abstract

Phylogenetic inference based on evidence from DNA sequences has led to significant strides in the development of a stable and robustly supported framework for the vertebrate tree of life. To date, the bulk of those advances have relied on sequence data from a small number of genome regions that have proven unable to produce satisfactory answers to consistently recalcitrant phylogenetic questions. Here, we re-examine phylogenetic relationships among early-branching euteleostean fish lineages classically grouped in the Protacanthopterygii using DNA sequence data surrounding ultraconserved elements. We report and examine a dataset of thirty-four OTUs with 17,957 aligned characters from fifty-three nuclear loci. Phylogenetic analysis is conducted both in concatenated and joint gene trees and species tree estimation frameworks. Both analytical frameworks yield supporting evidence for existing hypotheses of relationship for the placement of *Lepidogalaxias salamandroides*, monophyly of the Stomiatii and the presence of an esociform + salmonid clade. *Lepidogalxias salamandroides* and the Esociformes + Salmoniformes are successive sister lineages to all other euteleosts in the two analysis types receiving high support values for this arrangement. However, inter-relationships of Argentiniformes, Stomiatii and Neoteleostei remain uncertain as they varied by analysis type while receiving strong and contradictory indices of support. Topological differences between analysis types are apparent within the Ostarioclupeomorpha and the percomorph taxa in the data set. Our results identify concordant areas with strong support for relationships within and between early-branching euteleost lineages but they also reveal limitations in the ability of larger datasets to conclusively resolve other aspects of that phylogeny.

Keywords: Basal Euteleosts, Protacanthopterygii, Incomplete Lineage Sorting, Concatenation, Species Tree Estimation

1. Introduction

Phylogenomic datasets comprising hundreds to thousands of genome segments produced through high throughput sequencing technology have shown promise to resolve difficult phylogenetic problems (*e.g.* Faircloth et al., 2013, 2012; Gilbert et al., 2015; Harrington et al., 2016; Lemmon and Lemmon, 2013). At the same time, novel and refined inference tools including implementations of the multispecies coalescent model to address incomplete lineage sorting (ILS) through Gene Trees-to-Species Tree (GT-ST) methods (Knowles and Kubatko, 2011) continue to extend the power and complexity of phylogenetic research. Despite these advances in genomic-scale dataset production and phylogenetic inference, difficult areas of the tree of life remain unresolved (Delsuc et al., 2005; Pyron, 2015; Rokas and Carroll, 2006). Relationships among early-branching euteleost lineages remain nebulous (*e.g.* Betancur-R. et al., 2013; Campbell et al., 2013; Li et al., 2010; Near et al., 2012) and stand out as one of the most contentious regions of the fish tree of life. Although this question has been studied from morphological and molecular perspectives consensus has yet to emerge.

The name Euteleostei was first applied to a diverse group of fishes that includes all teleosts outside of the superorders Elopomorpha, Osteoglossomorpha and Clupeomorpha by phyletic analysis (Greenwood et al., 1967, 1966). Rosen (1985) excluded esocoids from the Euteleostei based on cladistic analyses of morphological characters, while Johnson and Patterson (1996) included esocoids but excluded ostariophysans. Subsequent phylogenetic studies of mitochondrial (*e.g.* López et al. 2004; Lavoué et al., 2008) and nuclear DNA (*e.g.* Betancur-R. et al., 2013; Near et al., 2012) supported a monophyletic Euteleostei including esocoids but excluding Ostariophysi and the Alepocephaliformes (previously classified in Argentiniformes nested in the Euteleostei).

64 Recent phylogenetic studies based on molecular evidence consistently support the
65 monophyly of five major euteleost lineages (Betancur-R. et al., 2013; Campbell et al., 2013; Li et
66 al., 2010; Near et al., 2012): 1) a clade formed by Esociformes and Salmoniformes; 2) the
67 Stomiatii *sensu* Betancur-R. et al. (2013) consisting of Osmeriformes (excluding Galaxiiformes)
68 and Stomiiformes; 3) the Argentiniformes (excluding the Alepocephaliformes); 4) the
69 Galaxiiformes (excluding *Lepidogalaxias*); and 5) the Neoteleostei. In addition, these studies
70 agree on placing the monotypic *Lepidogalaxias* as the sister group of all other euteleosts. Aside
71 from the placement of *Lepidogalaxias,* there is little congruence among different studies
72 regarding relationships among the five lineages (*e.g.* Betancur-R. et al., 2013; Campbell et al.,
73 2013; Li et al., 2010; Near et al., 2012). The early branching patterns of euteleosts are still in
74 need of further study and represent a difficult problem for traditional morphological and
75 molecular phylogenetics.
76 Here we apply the "new and general theory of molecular systematics" (Edwards, 2009) to
77 examine early-branching euteleost relationships using multi-locus datasets generated by targeted
78 enrichment of conserved nuclear DNA sequences. Concatenated and GT-ST phylogenetic
79 inference frameworks are used to assess the stability and strength of evidence for alternative
80 arrangements in this poorly resolved section of the fish tree of life.
81

82 **2. Material and Methods**

83 *2.1 Taxon and character sampling*

84 We targeted species representing five of the six major euteleost lineages as well as several non-
85 euteleost outgroups (Supplementary Table S1). We prepared genomic DNA libraries with 500-
86 600 bp inserts by shearing total genomic DNA extracts to size using a sonicator (Diagenode, Inc)

87 and ligating a set of custom-indexed Illumina Tru-Seq compatible adapters (Faircloth and Glenn,
88 2012) to the sheared DNA using reagents from a library preparation kit (KapaBiosystems, Inc.).
89 Adapter-ligated DNA was amplified with 16-18 cycles of PCR. To obtain sequences from
90 homologous loci across the taxonomic sample, we performed targeted enrichment of
91 ultraconserved element (UCEs) loci shared among acanthopterygians following protocols
92 outlined in Faircloth et al., (2013). We modified the capture protocol by pooling eight, indexed
93 sequencing libraries at equimolar ratios prior to enrichment and performing 12-16 cycles of
94 PCR-recovery after enrichment. Following the enrichment procedure, we quantified enriched,
95 amplified libraries using a commercial qPCR quantification kit (KapaBiosystems, Inc.), and we
96 prepared an equimolar pool of pooled libraries for sequencing on an Illumina HiSeq 2500
97 instrument using 100 base pair, paired-end sequencing chemistry in rapid run mode (UCLA
98 Neuroscience Genomics Core). To extend our taxon sampling, we included previously published
99 UCE data (Faircloth et al., 2013) in our analyses (Supplementary Table S1).
100
101 *2.2 Raw sequence data processing*
102 Demultiplexed reads were edited for length, overall quality and adapter contamination using
103 Trimmomatic v. 0.32 (Bolger et al., 2014). We assembled a subset of cleaned reads across
104 various kmers with Velvet v. 1.2.10 (Zerbino and Birney, 2008) to establish a range of suitable
105 kmers for assembly. We then assembled sequences for each species using two different
106 approaches. For non-salmonids, we assembled reads using VelvetOptimiser v. 2.2.5 across the
107 optimal range of kmers we identified (57 to 83). For salmonids, assemblies from Velvet were
108 produced for each value between 57 and 83. However, as the optimization performed by
109 VelvetOptimiser is designed for haploid or diploid organisms, an alternative selection criterion

110 of the maximum number of single copy UCE loci was chosen to accommodate the effect of
111 ancestral polyploidy in salmonid genomes (Allendorf and Thorgaard, 1984). A single dataset
112 assembly was retained downstream analyses from each alternative approach to data assembly.
113 We identified homologous UCE loci and prepared sequences for alignment with the PHYLUCE
114 pipeline (Faircloth, 2015). During orthology assessment, the PHYLUCE package screens for and
115 removes from analysis reciprocally duplicate enriched loci, which may represent paralogs.
116

117 *2.2 Alignment and phylogenetic analysis*

118 Following orthology assessment, the taxon set consisted of thirty-four Operational Taxonomic
119 Units (OTUs) representing outgroups and basal euteleost lineages. We ensured this taxon set
120 included loci sequenced in at least 31 of the 34 OTUs. We aligned data from all loci in with
121 MAFFT v. 7.130b (Katoh et al., 2002) through the PHYLUCE pipeline (Faircloth, 2015).
122 We analyzed the 34-OTU dataset under the Maximum-Likelihood (ML) framework as
123 implemented in RAxML v. 8.1.24 (Stamatakis, 2014). Each UCE locus was modeled as a
124 partition evolving under the general time reversible (GTR) model of sequence evolution with
125 gamma distributed rate variation (Γ). We set ML pseudoreplicate searches to automatically stop
126 when stable bootstrap indices were detected (autoMRE). A joint gene trees and species tree
127 estimation was conducted in a Bayesian framework with *BEAST (Heled and Drummond, 2010)
128 as implemented in BEAST v. 2.1.3 (Drummond et al., 2012). We analyzed data using a constant
129 coalescent model under a Hasegawa-Kishino-Yano (HKY) model of sequence evolution with a
130 four-category gamma distributed rate variation (Γ) and empirical base frequencies to each locus.
131 Convergence and sufficient effective sample sizes (ESSs, > 200) of all parameters were reached
132 by combining three chains of 800 Million generations with 40% burn-in. Two additional

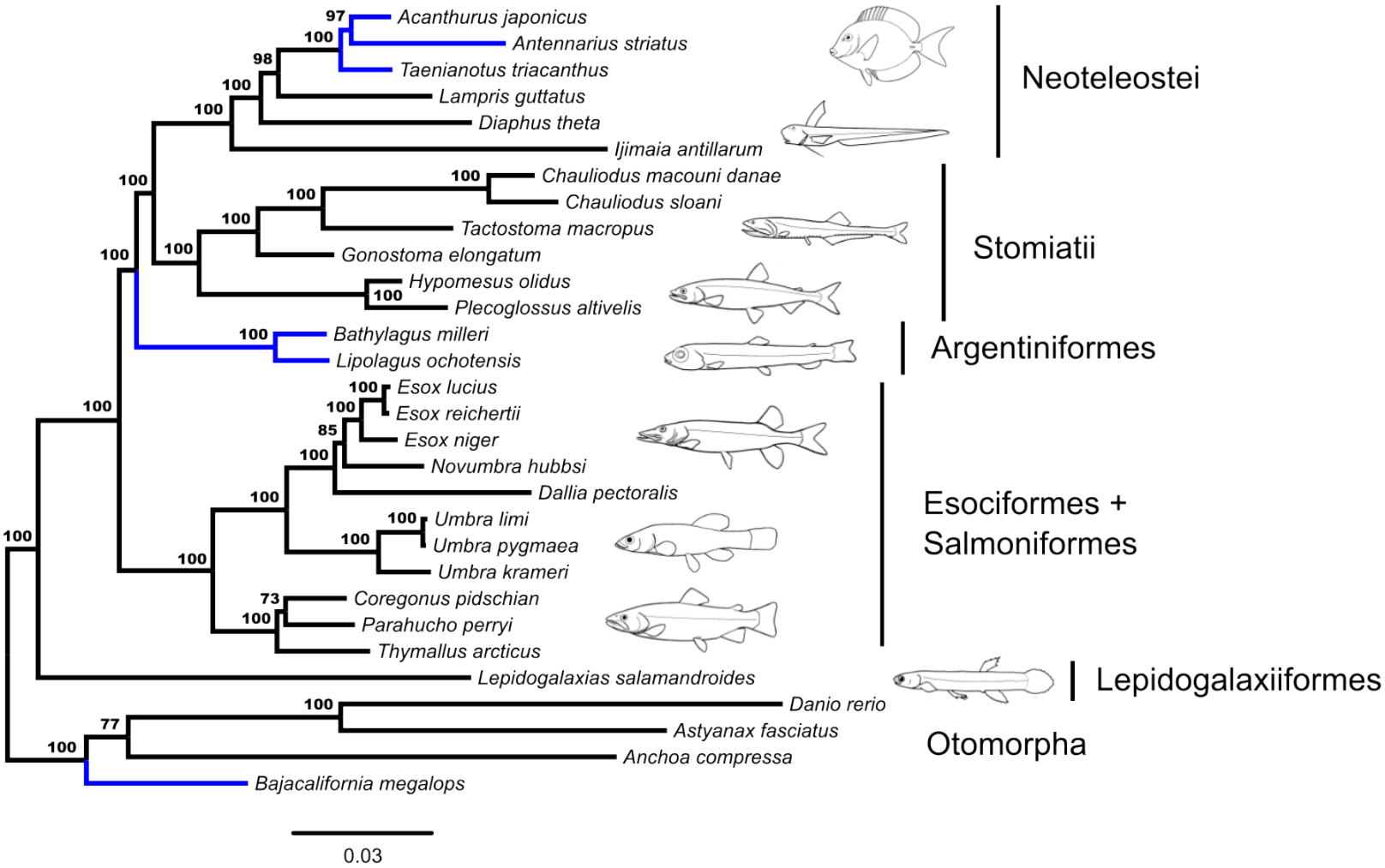
133 analyses were conducted. To verify that partitioning in the ML analysis by gene does not
134 influence early-branching euteleost relationships and support values, objective partitioning was
135 investigated. To verify that the use of the coalescent model in *BEAST resulted in an alternative
136 arrangement of early-branching euteleost lineages, not the choice of nucleotide evolution model,
137 a concatenated Bayesian analysis with BEAST 2 with the same nucleotide evolution model for
138 each UCE locus as *BEAST (HKY+Γ with empirical base frequencies) was undertaken
139 (Supplemental Document S1).
140
141 *2.3 Topology tests and occurrence of particular arrangements in the Bayesian tree posterior*
142 *sample*
143 To determine the significance of UCE evidence corroborating or refuting alternative
144 phylogenetic arrangements, we tested the following topologies resulting from concatenated and
145 GT-ST analysis against each other: (1) the best-scoring ML topology; (2) the consensus species-
146 tree topology from *BEAST; and (3) a Protacanthopterygii *sensu* Betancur-R. et al. (2013) as the
147 sister lineage to the Stomiatii. A best scoring ML tree (1 from above) and constrained trees (2
148 and 3 from above) were generated with RAxML v. 8.2.3 partitioned by UCE using a GTR + Γ
149 model of nucleotide evolution. We tested the trees against each other by generating per site
150 likelihoods with RAxML and analyzing the output with CONSEL v. 0.20 (Shimodaira and
151 Hasegawa, 2001). CONSEL implements several hypothesis tests allowing a more rigorous
152 comparison between alternative hypotheses than solely comparing likelihood values.
153 As the *BEAST posterior tree presented as the consensus species-tree topology
154 represents the combination of many different species trees, we searched the combined post burn-
155 in posterior tree sample from the separate *BEAST chains (180,003 trees) for alternative

156 phylogenetic hypotheses to determine if the *BEAST algorithm considered these alternatives.
157 The *BEAST posterior tree sample was searched for the best scoring ML topology and a
158 monophyletic Protacanthopterygii *sensu* Betancur-R. et al. (2013) with Python scripts (Moravec,
159 2015).
160
161 **3. Results**
162 *3.1 Characteristics of UCE dataset*
163 Following orthology assessment and filtering for loci not present in 31 of 34 OTUs, the dataset is
164 composed of a total of 53 UCE loci, 17,957 characters, 9,576 distinct alignment patterns and
165 22.11% gaps or missing data. We present details of the number of UCE loci recovered for each
166 taxon, the average length of UCE matching contigs, average coverage of contigs matching UCEs
167 and number of duplicate loci removed in Supplementary Table S1. The assemblies and alignment
168 are available within the Data Supplement.
169
170 *3.2 Early-branching euteleost relationships*
171 Concatenated ML analysis supports a monophyletic Euteleostei, excluding Ostariophysi and
172 Alepocephaliformes (Bootstrap Support [bs] = 100%). Figure 1 shows the inferred branching
173 pattern among main euteleost groups from the 34-OTU dataset. Relationships among main
174 euteleost lineages in the concatenated ML topology are (*Lepidogalaxias*, ((Esociformes,
175 Salmoniformes), (Argentiniformes, (Stomiatii, Neoteleostei)))) with all nodes among those
176 lineages receiving strong support (bs = 100%).
177 GT-ST analysis of the dataset in *BEAST indicates a monophyletic Euteleostei with high
178 support, posterior probability (pp) = 1.00 (Figure 2). A topology of (*Lepidogalaxias,*

((Esociformes, Salmoniformes), ((Argentiniformes, Stomiatii), Neoteleostei))))) is generated in this analysis. Support values for the placement of main euteleost lineages are high throughout the consensus tree. The placement of *Lepidogalaxias* and the Esociformes + Salmoniformes receive very high support (pp = 1.00). Argentiniformes + Stomiatii as the sister lineage of the neoteolosts received strong support (pp = 0.99). A sister relationship between the Argentiniformes and Stomiatii was also well supported (pp = 0.96). The GT-ST and ML inferred phylogenies differ on the relationships among argentiniforms, stomiatians and neoteleosts.

Through the additional concatenated analyses presented in the Supplemental Document S1, conflicts between ML and GT-ST results presented in Figures 1 & 2 are shown to the product of the distinct analytical frameworks not from how data are modeled. The additional concatenated analyses in Supplemental Document S1 show identical branching patterns for main early-branching euteleost lineages to the concatenated ML analysis presented in Figure 1 with high support values. Retaining the same model but changing the partitioning strategy with RAxML demonstrates that the inferred phylogeny from the ML analysis presented in Figure 1 is not sensitive to partitioning (Supplemental Figure S1). Not implementing a *BEAST model, while retaining the same nucleotide evolution and partitioning scheme for a concatenated analysis with BEAST 2 also produces a phylogeny (Supplemental Figure S2) with the branching of main early-branching euteleost lineages matching that of the concatenated ML analysis presented in Figure 1, not the *BEAST GT-ST analysis presented in Figure 2. Consequently, the topological differences between phylogenies shown in Supplemental Figure S2 and Figure 2 may be attributed to whether a concatenated or coalescent approach is implemented.

PeerJ Preprints NOT PEER-REVIEWED

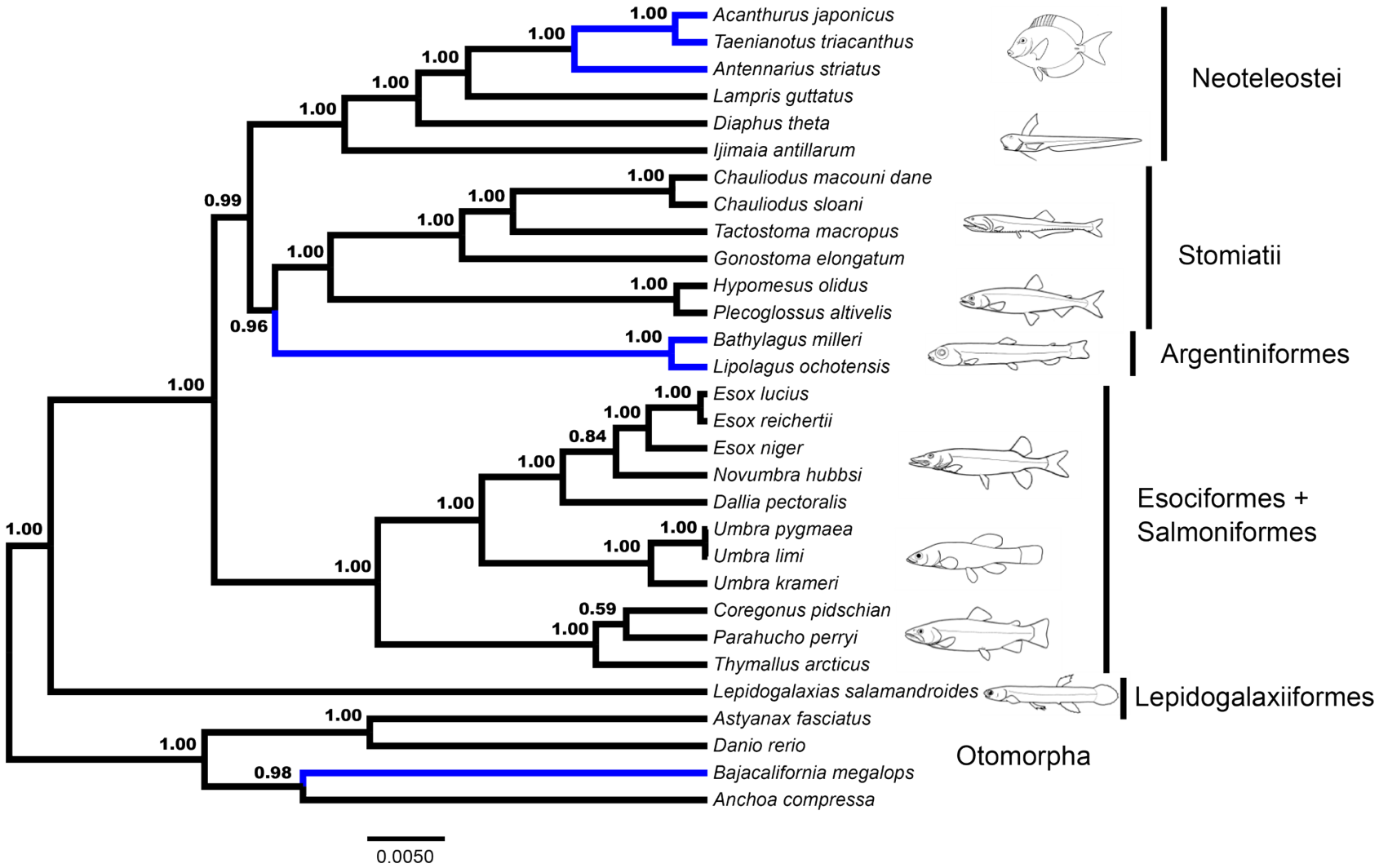


PeerJ Preprints NOT PEER-REVIEWED

202 **Figure 1:** Phylogenetic tree from fifty-three ultraconserved element (UCE) loci generated in a concatenated framework with RAxML.
203 Each locus is designated as a partition and modeled under a GTR + Γ model of nucleotide evolution. Values from automatic stopping
204 of bootstrap replicates are indicated at each node. The tree is rooted by *Polypterus senegalus*, this taxon, *Amia calva, Osteoglossum*
205 *bicirrhosum,* and *Pantodon buchholzi* are omitted from figure. Early-branching euteleost taxa are labeled and indicated by drawings of
206 a representative taxon (Nelson, 2006). From the Neoteleostei, Ateleopodiformes and Acanthuriformes drawings are included.
207 Placements of taxa that are different from the GT-ST topology (Figure 2) are indicated in blue.
208

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2903v1 | CC BY 4.0 Open Access | rec: 1 Apr 2017, publ: 1 Apr 2017

PeerJ Preprints NOT PEER-REVIEWED



PeerJ Preprints NOT PEER-REVIEWED

211 **Figure 2:** Species tree from *BEAST. Fifty-three ultraconserved element (UCE) loci are modeled under an HKY model of nucleotide
212 sequence evolution with a four category gamma distribution characterizing rate variation among sites (Γ). Each model of sequence
213 evolution has independent model parameters. This tree represents the combination of three independent *BEAST runs with the
214 posterior probability of each node indicated. Early-branching euteleost lineages are labeled and indicated with representations of a
215 representative taxon (Nelson, 2006). Images of neoteleost lineages from Acanthuriformes and Ateleopodiformes are also included.
216 The tree is rooted by *Polypterus senegalus*, this taxon, *Amia calva, Osteoglossum bicirrhosum,* and *Pantodon buchholzi* are omitted
217 from figure. Placements of taxa that are different from the concatenated topology (Figure 1) are indicated in blue.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2903v1 | CC BY 4.0 Open Access | rec: 1 Apr 2017, publ: 1 Apr 2017

3.3 Topology tests and occurrence of particular arrangements in the Bayesian tree posterior sample

Testing with CONSEL indicates the best-scoring ML tree, with a topology of (*Lepidogalaxias,* ((Esociformes, Salmoniformes), (Argentiniformes, (Stomiatii, Neoteleostei)))), is significantly better than the topology generated by GT-ST analysis with both the approximately unbiased test ($p = 1 \times 10^{-5}$) and the weighted Shimodaira-Hasegawa test ($p = 1 \times 10^{-3}$). A monophyletic assemblage of protacanthopterygian taxa *sensu* Betancur-R. et al. (2013) sister to the Stomiatii is significantly worse than the best-scoring ML tree with both the approximately unbiased test ($p = 8 \times 10^{-6}$) and the weighted Shimodaira-Hasegawa test ($p = 1 \times 10^{-4}$). The posterior set of 180,003 trees generated by *BEAST did not include a single occurrence of either the ML best tree topology or a monophyletic Protacanthopterygii *sensu* Betancur-R. et al. (2013).

4. Discussion

4.1 Hypotheses of early-branching euteleost relationships

Our phylogenomic analysis provides strong support for relationships of early diverging euteleosts that consist of *Lepidogalaxias* and esociforms + salmoniforms as successive sister lineages to a clade containing argentiniforms, stomiatiids and neoteleosts. Despite the most intensive character sampling of this group to date, our analyses do not resolve two conflicting hypotheses for relationships among the Argentiniformes, Stomiatii and Neoteleostei. The concatenated ML derived topology resolves argentiniforms and stomatiids as successive sister lineages to the neoteleosts, while the GT-ST analysis recovers an argentiniform + stomiatiids clade as the sister group to neoteleosts.

240 Combined, our analyses yield strong support for the Esociformes + Salmoniformes clade,
241 which has found robust and consistent support in molecular phylogenetic studies (López et al.,
242 2004), reviewed by (Campbell et al., 2013), despite weak or conflicting evidence from
243 morphology (Johnson and Patterson, 1996; Wilson and Williams, 2010). We also recover the
244 Stomiatii (Osmeriformes + Stomiiformes) with high support values in both analyses in this
245 study. On the other hand, we do not find a close relationship between the clade of Esociformes +
246 Salmoniformes and any other major group of early-branching euteleosts such as Argentiniformes
247 (Near et al., 2012). Instead, as shown in mitogenomic phylogenies (Campbell et al., 2013; Inoue
248 et al., 2003) or analyses of combined mitochondrial and nuclear data (Burridge et al., 2012), we
249 find Esociformes and Salmoniformes as sister to all other euteleosts in the study, with the
250 exclusion of *Lepidogalaxias.*
251
252 *4.2 Support for hypotheses of early-branching euteleost lineages*
253 Unlike other molecular (and morphological) studies of the euteleost phylogeny (*e.g.* Betancur-R.
254 et al., 2013; Li et al., 2010; Near et al., 2012), our conflicting topologies are strongly supported
255 by both bootstrap values and Bayesian posterior probabilities.
256 Earlier studies typically yield low or moderate support for relationships along this section of the
257 teleost phylogeny backbone. For example, the placement of the Argentiniformes and
258 Salmoniformes + Esociformes sister to the remaining three major euteleost lineages (Stomiati,
259 Galaxiiformes, and neoteleosts) receives a bootstrap support value between 70-89% in Near et al.
260 (2012). Other nodes supporting the branching order of the five major euteleost lineages are
261 supported by 90-99% bootstrap support values. The bootstrap support for a sister
262 Protacanthopterygii *sensu* Betancur-R. et al. (2013) and Stomiatii is 36%, and the monophyly of

Stomiatii receives a bootstrap support of 73% in the hypothesis presented by Betancur-R. et al. (2013). In a mitochondrial genome based study, a sister relationship of Argentiniformes to the Salmoniformes + Esociformes receives a bootstrap support of 74% (Li et al., 2010). In the same study, the Argentiniformes, Salmoniformes and Esociformes are the sister lineage of the Stomiatii, supported by an 81% bootstrap support value (Li et al., 2010).

While we find uncharacteristically high support for branching relationships among all of the four major euteleost lineages represented in this study in a concatenated ML framework, gauging the significance of high bootstrap values in analyses of large data matrices is problematic. Bootstrap values may be high even with conflict or systematic error (Felsenstein, 1978; Hillis and Bull, 1993; Huelsenbeck, 1997). Concatenated ML phylogenomic analysis has previously been demonstrated with 1,070 genes in yeasts to produce 100% bootstrap support for all internodes, despite incorrect branching likely present (Salichos and Rokas, 2013). The GT-ST analysis also produces high support values; however, posterior probability values themselves are both conditioned on the model of evolution and are not guaranteed to have good frequentist statistical behavior (Alfaro et al., 2003; Alfaro and Holder, 2006) and may be misleading under certain conditions (Suzuki et al., 2002; Salichos and Rokas, 2013).

4.2 Hypothesis testing and alternative topologies in the Bayesian posterior tree sample

In a hypothesis testing framework, the optimal topology from the GT-ST framework is a significantly worse fit compared to the concatenated ML best tree. Conversely the concatenated ML best tree topology is absent from the 180,003 posterior trees produced in the GT-ST analysis. Combined, these demonstrate that strong conflicting signal underlies these topological differences. Recent studies have alternatively suggested that concatenation may perform better

286 than GT-ST when individual loci are not long enough to resolve phylogenies (Gatesy and
287 Springer, 2014), that concatenation and GT-ST methods should behave similarly under a range
288 of conditions (Tonini et al., 2015), and that phylogenomic scale data sets may exacerbate
289 problems of model misspecification (Liu et al., 2015). For additional discussion around these
290 issues see also Edwards et al., (2016) and Springer and Gatesy (2016). At present, the
291 relationships of the argentiniforms and stomatiids to neoteleosts remain unclear and may depend
292 strongly on the inclusion of the Galaxiidae. The placement of galaxiids has been unstable
293 (Betancur-R. et al., 2013; Burridge et al., 2012; Campbell et al., 2013; Ishiguro et al., 2003; Li et
294 al., 2010; López et al., 2004; Near et al., 2012), although independent studies *e.g.* (Campbell et
295 al., 2013; Near et al., 2012) suggest that galaxiids may be the sister lineage of the Neoteleostei.
296

297 *4.3 Lack of evidence for the monophyly of protacanthopterygians*

298 The Protacanthopterygii is a historically important taxon of early-branching euteleosts with its
299 definition and content repeatedly modified (*e.g.* Greenwood et al., 1966; Johnson and Patterson,
300 1996; Lauder and Liem, 1983; Rosen, 1973; Rosen and Greenwood, 1970; Rosen and Patterson,
301 1969). Protacanthopterygian monophyly as defined by morphology (*e.g.* Johnson and Patterson,
302 1996) was questioned by molecular phylogenetics (Ishiguro et al., 2003). More recently, the
303 Protacanthopterygii was redefined by Betancur-R. et al. (2013) with molecular phylogenetics (bs
304 of 37%) containing the Argentiniformes, Galaxiiformes, Esociformes and Salmoniformes.
305 Although we were unable to obtain representatives of Galaxiiformes, our analyses demonstrate
306 that the Argentiniformes are not most closely related to the Esociformes + Salmoniformes. A
307 topology test using available taxa in this dataset further rejected the Protacanthopterygii *sensu*
308 Betancur-R. et al. (2013).

309

310 **5. Conclusions**

311 Two the approaches (concatenation and GT-ST) implemented in this study indicated large areas
312 of congruence in topology resolving several relationships within the euteleosts. However, the
313 disagreements highlight some of the potential caveats in resolving all relationships of the early-
314 branching euteleosts. We report the first study using a joint GT-ST method to examine the
315 question of early-branching euteleost relationships. A joint estimation of species tree and gene
316 trees was chosen over other "shortcut" methods (Gatesy and Springer, 2014) and produced a
317 slightly different hypothesis of relationships when compared to concatenated analyses. A test of
318 topology rejects the species-tree topology over the best scoring concatenated ML topology.
319 Likewise, posterior support for the Bayesian species tree hypothesis is high for early-branching
320 euteleost nodes, indicating very few occurrences of alternative topologies in the tree search. For
321 major euteleost lineages, relationships among Argentiniformes, Neoteleostei and Stomiatii
322 differed in the results of concatenated ML and Bayesian joint GT-ST analyses. This is in line
323 with previous research on early-branching euteleost relationships. The lack of agreement
324 between studies of early-branching euteleost relationships may be caused by short internode
325 distances deep in the evolutionary past, leading to the formation and preservation of few
326 informative characters linking these old lineages. A related but less likely possibility is that short
327 internodes associated with very rapid diversification created conditions conducive to pervasive
328 ILS at the base of the euteleost radiation resulting in conflicting histories across euteleost
329 genomes and incongruent results between studies of early-branching euteleost relationships.

330 We evaluated identical datasets under concatenated and GT-ST frameworks and find
331 three areas of incongruence: 1) argentiniform sister lineage, 2) the placement of the

alepocephaliform lineage *Bajacalifornia*, and 3) the arrangement of the three neoteleost lineages *Antennarius*, *Acanthurus* and *Taenianotus*. The percomorph taxa (*Antennarius*, *Acanthurus* and *Taenianotus*) belong in a set of fish lineages whose relationship have been particularly difficult to elucidate (Nelson, 1989). The incongruent inferences we observed between the two approaches may be differential effects of ILS on coalescent versus non-coalescent phylogenetic approaches.

A final question to consider in this manuscript is: Which analysis to prefer? There is not clear evidence to prefer a particular analysis framework to another. In terms of main early-branching euteleost lineages, only the placement of Argentiniformes between concatenated and GT-ST hypotheses varied. The placement of the argentiniform fishes is unresolved by this study and that branching between the Neotelostei, Stomiati and Argentiniformes may be considered a soft polytomy. We find that phylogenomics and the application of the coalescent model in phylogenetics strengthen support for the earliest splits in the euteleostean radiation. However, key aspects of early euteleost phylogeny remain unresolved and leave open the question of whether extant genomes from these lineages retain historical signal that can be retrieved above the noise accumulated over hundreds of millions of years of independent evolution.

Acknowledgements

We would like to acknowledge the following institutions and individuals that contributed to sampling in this project: Burke Museum of Natural History and Culture, Kansas University Biodiversity Institute and Natural History Museum, Academy of Natural Sciences of Philadelphia, Robert Marcotte (Univeristy of Alaska Fairbanks), Molly Hallock (Washington Department of Fish and Wildlife, retired), Motohiro Kikuchi (Chitose Salmon Park) and Peter

Unmack (University of Canberra). Sébastien Lavoué (National Taiwan University), Thaddaeus Buser (Oregon State University) and Kerry Reid (University of California Santa Cruz) provided very helpful comments on drafts of the manuscript. We would like to thank Kevin McCracken (University of Miami) for his support in promoting Next-Generation Sequencing and Bioinformatics at the University of Alaska Fairbanks.

360 **References**

361 Alfaro, M.E., Holder, M.T., 2006. The posterior and the prior in Bayesian phylogenetics. Annu.
362 Rev. Ecol. Evol. Syst. 37, 19-42. doi:10.1146/annurev.ecolsys.37.091305.110021
363 Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing
364 the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in
365 assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.
366 doi:10.1093/molbev/msg028
367 Allendorf, F., Thorgaard, G., 1984. Tetraploidy and the evolution of salmonid Fishes, in: Turner,
368 B. (Ed.), Evolutionary Genetics of Fishes, Monographs in Evolutionary Biology.
369 Springer US, pp. 1–53.
370 Betancur-R., R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I.,
371 Arcila, D., Sanciangco, M., Cureton II, J.C., Zhang, F., Buser, T., Campbell, M.A.,
372 Ballesteros, J.A., Roa-Varon, A., Willis, S., Borden, W.C., Rowley, T., Reneau, P.C.,
373 Hough, D.J., Lu, G., Grande, T., Arratia, G., Ortí, G., 2013. The tree of life and a new
374 classification of bony fishes. PLoS Curr, 5, 1-33.
375 doi:10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288
376 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina
377 sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170
378 Burridge, C.P., McDowall, R.M., Craw, D., Wilson, M.V.H., Waters, J.M., 2012. Marine
379 dispersal as a pre-requisite for Gondwanan vicariance among elements of the galaxiid
380 fish fauna. J. Biogeogr. 39, 306–321. doi:10.1111/j.1365-2699.2011.02600.x
381 Campbell, M.A., López, J.A., Sado, T., Miya, M., 2013. Pike and salmon as sister taxa: Detailed
382 intraclade resolution and divergence time estimation of Esociformes + Salmoniformes

383 based on whole mitochondrial genome sequences. Gene 530, 57–65.
384 doi:10.1016/j.gene.2013.07.068
385 Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree
386 of life. Nat. Rev. Genet. 6, 361–375. doi:10.1038/nrg1603
387 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with
388 BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–73. doi:10.1093/molbev/mss075
389 Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution
390 63, 1–19. doi:10.1111/j.1558-5646.2008.00549.x
391 Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu,
392 S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016.
393 Implementing and testing the multispecies coalescent model: A valuable paradigm for
394 phylogenomics. Mol. Phylogenet. Evol. 94, Part A, 447–462.
395 doi:10.1016/j.ympev.2015.10.027
396 Faircloth, B.C., 2015. PHYLUCE is a software package for the analysis of conserved genomic
397 loci. Bioinformatics, btv646 doi:10.1093/bioinformatics/btv646
398 Faircloth, B.C., Glenn, T.C., 2012. Not all sequence tags are created equal: designing and
399 validating sequence identification tags robust to indels. PloS One 7, e42543.
400 Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C.,
401 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple
402 evolutionary timescales. Syst. Biol, sys004. doi:10.1093/sysbio/sys004
403 Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the
404 radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements
405 (UCEs). PLoS ONE 8, e65923. doi:10.1371/journal.pone.0065923

406 Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively
407 misleading. Syst. Biol. 27, 401–410. doi:10.1093/sysbio/27.4.401
408 Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees,
409 bypassed hidden support, and the coalescence/concatalescence conundrum. Mol.
410 Phylogenet. Evol. 80, 231–266. doi:10.1016/j.ympev.2014.08.013
411 Gilbert, P.S., Chang, J., Pan, C., Sobel, E.M., Sinsheimer, J.S., Faircloth, B.C., Alfaro, M.E.,
412 2015. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness
413 than traditional gene markers in percomorph fishes. Mol. Phylogenet. Evol. 92, 140–146.
414 doi:10.1016/j.ympev.2015.05.027
415 Greenwood, P.H., Myers, G.S., Rosen, D.E., Weitzman, S.I.I., 1967. Named main divisions of
416 teleostean fishes. Proc. Biol. Soc. Wash. 80, 227–228.
417 Greenwood, P.H., Rosen, D.E., Weitzman, S.H., Myers, G.S., 1966. Phyletic studies of
418 teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat.
419 Hist. 131, 339–456.
420 Harrington R.C., Faircloth B.C., Eytan R.I., Smith W.L., Near T.J., Alfaro M.E., Friedman M.,
421 2016. Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in
422 a blink of the evolutionary eye. BMC Evol. Biol. 16, 224. doi: 10.1186/s12862-016-
423 0786-x
424 Heled, J., Drummond, A.J., 2010. Bayesian Inference of Species Trees from Multilocus Data.
425 Mol. Biol. Evol. 27, 570–580. doi:10.1093/molbev/msp274
426 Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing
427 confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
428 doi:10.1093/sysbio/42.2.182

429 Huelsenbeck, J.P., 1997. Is the Felsenstein Zone a fly trap? Syst. Biol. 46, 69–74.
430 doi:10.1093/sysbio/46.1.69
431 Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M., 2003. Basal actinopterygian relationships: a
432 mitogenomic perspective on the phylogeny of the "ancient fish." Mol. Phylogenet. Evol.
433 26, 110–120. doi: 10.1016/S1055-7903(02)00331-7
434 Ishiguro, N.B., Miya, M., Nishida, M., 2003. Basal euteleostean relationships: a mitogenomic
435 perspective on the phylogenetic reality of the "Protacanthopterygii." Mol. Phylogenet.
436 Evol. 27, 476–488. doi: 10.1016/S1055-7903(02)00418-9
437 Johnson, G.D., Patterson, C., 1996. Relationships of lower euteloestean fishes, in: Stiassny,
438 M.L.J. (Ed.), Interrelationships of Fishes. Academic Press, pp. 251–328.
439 Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple
440 sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059 –3066.
441 doi:10.1093/nar/gkf436
442 Knowles, L.L., Kubatko, L.S., 2011. Estimating species trees: practical and theoretical aspects.
443 John Wiley and Sons.
444 Lauder, G., Liem, K., 1983. The evolution and interrelationships of the actinopterygian fishes.
445 Bull. Mus. Comp. Zool. 150, 95–197.
446 Lavoué, S., Miya, M., Poulsen, J.Y., Moller, P.R., Nishida, M., 2008. Monophyly, phylogenetic
447 position and inter-familial relationships of the Alepocephaliformes (Teleostei) based on
448 whole mitogenome sequences. Mol. Phylogenet. Evol. 47, 1111–1121. doi:
449 10.1016/j.ympev.2007.12.002

Lemmon, E.M., Lemmon, A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99–121. doi: 10.1146/annurev-ecolsys-110512-135822

Li, J., Xia, R., McDowall, R.M., López, J.A., Lei, G., Fu, C., 2010. Phylogenetic position of the enigmatic *Lepidogalaxias salamandroides* with comment on the orders of lower euteleostean fishes. Mol. Phylogenet. Evol. 57, 932–936. doi:10.1016/j.ympev.2010.07.016

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data. Ann. N. Y. Acad. Sci. 1360, 36–53. doi:10.1111/nyas.12747

López, J.A., Chen, W.J., Ortí, G., 2004. Esociform phylogeny. Copeia 2004, 449–464. doi: 10.1643/CG-03-087R1

Moravec, J., 2015. BayesMonophyly. Available at https://github.com/J-Moravec/BayesMonophyly

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. 109, 13698–13703. doi:10.1073/pnas.1206625109

Nelson, G., 1989. Phylogeny of major fish groups, in: Fernholm, B., Bremer, K., Brundin, L., Jornvall, H., Rutberg, L., Wanntorp, H.-E. (Eds.), The Hierarchy of Life. Elsevier, Amsterdam, pp. 325–336.

Nelson, J.S., 2006. Fishes of the world, 4th ed. John Wiley and Sons, Inc., New York.

Pyron, R.A., 2015. Post-molecular systematics and the future of phylogenetics. Trends Ecol. Evol. 30, 384–389. doi:10.1016/j.tree.2015.04.016

Rokas, A., Carroll, S.B., 2006. Bushes in the tree of life. PLoS Biol. 4, e352.

473 Rosen, D., 1973. Interrelationships of higher teleostean fishes, in: Greenwood, P., Miles, R.,
474 Patterson, C. (Eds.), Interrelationships of Fishes. Academic Press, London, pp. 397–513.
475 Rosen, D.E., 1985. An essay on euteleostean classification. Am. Mus. Novit. 2827, 1–57.
476 Rosen, D., Greenwood, P., 1970. Origin of the Weberian apparatus and the relationships of the
477 ostariophysan and gonorynchiform fishes. Am. Mus. Novit. 2428, 1–25.
478 Rosen, D., Patterson, C., 1969. The structure and relationships of the paracanthopterygian fishes.
479 Bull. Am. Mus. Nat. Hist. 141, 357–474.
480 Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong
481 phylogenetic signals. Nature 497, 327-331.
482 Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic
483 tree selection. Bioinformatics 17, 1246–1247. doi:10.1093/bioinformatics/17.12.1246
484 Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Mol. Phylogenet. Evol. 94, Part A, 1–
485 33. doi:10.1016/j.ympev.2015.07.018
486 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
487 large phylogenies. Bioinformatics 30, 1312–1313. doi:10.1093/bioinformatics/btu033
488 Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by
489 Bayesian phylogenetics. Proc. Natl. Acad. Sci. 99, 16138–16143.
490 doi:10.1073/pnas.212646199
491 Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and species
492 tree methods exhibit statistically indistinguishable accuracy under a range of simulated
493 conditions. PLoS Curr. 7, ecurrents.tol.34260cc27551a527b124ec5f6334b6be.
494 doi:10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be

495 Wilson, M.V.H., Williams, R.R.G., 2010. Salmoniform fishes: key fossils, supertree, and
496 possible morphological synapomorphies., in: Nelson, J.S., Schultze, H.P., Wilson,
497 M.V.H. (Eds.), Origin and Phylogenetic Interrelationships of Teleosts. Verlag Dr.
498 Friedrich Pfeil, München, pp. 379–409.
499 Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de
500 Bruijn graphs. Genome Res. 18, 821–829. doi:10.1101/gr.074492.107