

In the format provided by the authors and unedited.

# Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life

6 Dahiana Arcila<sup>1,2</sup>, Guillermo Ortí<sup>1</sup>, Richard Vari<sup>2§</sup>, Jonathan W. Armbruster<sup>3</sup>, Melanie L.  
7 J. Stiassny<sup>4</sup>, Kyung D. Ko<sup>1</sup>, Mark H. Sabaj<sup>5</sup>, John Lundberg<sup>5</sup>, Liam J. Revell<sup>6</sup> & Ricardo  
8 Betancur-R.<sup>7,2\*</sup>

9

10

11 <sup>1</sup>Department of Biological Sciences, The George Washington University, 2023 G St.  
12 NW, Washington, DC, 20052, United States

13 <sup>2</sup>Department of Vertebrate Zoology, National Museum of Natural History Smithsonian  
14 Institution, PO Box 37012, MRC 159, Washington, DC, 20013, United States

15 <sup>3</sup>Department of Biological Sciences, Auburn University, Auburn, AL 36849, United  
16 States

17 <sup>4</sup>Department of Ichthyology, Division of Vertebrate Zoology, American Museum of  
18 Natural History, New York, New York, United States

19 <sup>5</sup>Department of Ichthyology, The Academy of Natural Sciences, 1900 Benjamin Franklin  
20 Parkway, Philadelphia, Pennsylvania 19103, United States

21 <sup>6</sup>Department of Biology, University of Massachusetts Boston, Boston, Massachusetts  
22 02125, United States

23 <sup>7</sup>Department of Biology, University of Puerto Rico – Río Piedras, PO Box 23360, San  
24 Juan, Puerto Rico

25

26

## 27 **This PDF file includes:**

28

29       Supplementary Methods

30

30       Supplementary Notes

31

31       Figs. S1 to S7

32

32       Tables S1 to S8

33

## 34 **Other Supplementary Materials for this manuscript:**

35

36       A compressed folder containing all data files is available from Zenodo.org  
37       (<http://dx.doi.org/10.5281/zenodo.51603>). These include:

38

39       Database S1. List of target loci and probes designed for pilot study

40

40       Database S2. Scripts for *quasi de novo* assembler

41

41       Database S3. DNA sequence alignments, gene and multi-locus trees for pilot  
42       study

43

43       Database S4. List of loci selected for Otophysi

44

44       Database S5. List of target loci and probe designed for Otophysi

45

45       Database S6. DNA and protein sequence alignments for Otophysi

- 46 Database S7. List of presence and absence of species and genes examined for  
 47 Otophysi  
 48 Database S8. Gene sequence annotation using Blast2Go. GO = Gene ontology  
 49 Database S9. Summary of gene properties and selection criteria for subset  
 50 assembly (selected loci for each subset given in bold)  
 51 Database S10. Gene and multi-locus trees for Otophysi  
 52 Database S11. GGI tutorial dataset  
 53 Database S12. GGI trees (Otophysi) and species trees analyses using STAR and  
 54 ASTRAL-2  
 55 Database S13. GGI trees metazoans, birds, mammals, and yeast

## 57 Supplementary Methods

58  
 59 **Pilot study (ray-finned fishes).** We ran an initial experiment to test the efficacy of target  
 60 capture techniques to obtain genome-level data for a large number of exons across the  
 61 diversity of fishes. Taxonomic sampling for this experiment consisted of 4 otophysans  
 62 (one representative per order) and 10 distantly-related ray-finned fish species (Table S5;  
 63 Database S1). Exon markers were selected using the EvolMarker pipeline<sup>1</sup> to screen  
 64 genomes of zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*) via pairwise BLAST  
 65 searches to identify single-copy, slowly-evolving exon regions according to a 50%  
 66 dissimilarity criterion. A set of 3,957 putatively orthologous exons longer than 200 bp  
 67 was defined by this method. Custom biotinylated RNA bait libraries for the pilot study  
 68 were designed based on zebrafish sequences following specifications of the MYBait  
 69 target enrichment system (<http://www.mycroarray.com>). A total of 15,376  
 70 oligonucleotide baits (120 bp long) tiling over the 3,957 exon sequences with 2x density  
 71 were designed using the py\_tiler.py script [https://github.com/faircloth-lab/uce-probe-](https://github.com/faircloth-lab/uce-probe-design)  
 72 [design](https://github.com/faircloth-lab/uce-probe-design); <sup>2</sup>].

73 DNA extractions were performed using DNeasy extraction kits (Qiagen, Inc.)  
 74 and resulting genomic DNAs were sheared to a target size of 500 bp using a Covaris®  
 75 sonicator. Library preparation and indexing followed standard Illumina protocols. Target  
 76 enrichment (TE) was conducted using a modification of the Mamanova et al. protocol<sup>3</sup>,  
 77 as optimized by Faircloth et al.<sup>2</sup> Indexed libraries enriched for the 3,957 exons were  
 78 pooled and sequenced in two lanes of Illumina HiScan.

79 Reads were assembled into contigs with a *quasi de novo* approach<sup>4</sup> using custom  
 80 scripts to perform the following steps (Dataset S2). First, reads from each of the 14  
 81 species were mapped against the reference zebrafish genome using GNUMAP v3.0.2<sup>5</sup>,  
 82 setting similarity (a), genome mer-size (m), and matching seed hashes (k) to 0.55, 7, and  
 83 5, respectively. Matching reads and their pair-mates (FASTQ files) were collected  
 84 separately for each species and locus. Second, each set of collected reads was assembled  
 85 separately using the *de novo* algorithm with pair-mate mode implemented in Trinity<sup>6</sup>. The  
 86 number of contigs per locus obtained for each species varied from zero to four. Third,  
 87 pairwise alignments were conducted against the reference for contigs from each locus  
 88 separately using BLAST searches. If multiple contigs were retrieved for a single locus  
 89 with a similarity value < 97%, they were flagged as potential paralogs and discarded for  
 90 downstream analyses; if two contigs assembled had a similarity value >97%, presumably  
 91 representing allelic variants, they were subsequently collapsed using IUPAC ambiguity

92 codes. Finally, the coding region for each orthologous exon was extracted from the  
 93 assembled contigs for all species and subsequently aligned using MAFFT v7.023<sup>7</sup> (Table  
 94 S5).

95 For 279 loci, assemblies produced more than one contig for at least one species  
 96 with <97% similarity; therefore, these were removed from further analyses. In addition,  
 97 2637 exons also were discarded because they were captured with low efficiency and had  
 98 either extremely low sequencing coverage or were successfully assembled in less than 6  
 99 species. The final dataset assembled consisted of 1041 single-copy genes (288,581 sites).  
 100 To test the efficacy of these markers in resolving ray-finned fish phylogeny, two different  
 101 analyses were conducted. First, a Maximum Likelihood (ML) analysis of the  
 102 concatenated dataset in RAxML v8.2.6<sup>8</sup> based on the GTRGAMMA model; second, a  
 103 summary multi-species coalescent analysis using ASTRAL-2<sup>9</sup>, using as input the 1041  
 104 gene trees obtained with RAxML (Database S3). Both methods converged on a single  
 105 well-supported topology that is highly congruent with previous phylogenetic hypotheses  
 106 for fishes<sup>10,11; fig. S3</sup>). This preferred set of 1041 single-copy exons was used to create a  
 107 new capture library customized for otophysan taxa.

108

### 109 **Supplementary Notes**

110

111 **Testing the assumption of subclade monophyly of individual gene genealogies.** The  
 112 main goal of the GGI approach is to address a systematic question when a number of  
 113 well-supported clades of undisputed monophyly have unresolved relationships (in our  
 114 case, the relationships among otophysan orders/suborders). The assumption of  
 115 monophyly for these “undisputed” clades is based on organismal phylogenies supported  
 116 by several lines of evidence and is more likely to be met for ancient divergences spanning  
 117 tenths or hundreds of millions of years of evolution (e.g., Otophysi, and the other groups  
 118 analyzed here) than in shallow species divergences (e.g., humans, gorillas and  
 119 chimpanzees)<sup>12</sup>. The question is whether in these cases forcing the monophyly of  
 120 organismal clades onto individual gene trees is acceptable or reasonable. Rosenberg<sup>13</sup>  
 121 calculated that under a neutral coalescent model the vast majority of genes in a genome  
 122 (99.99–99.999%) require 5.3–8.3 coalescence time units to achieve monophyly (for either  
 123 one species or two sister species). While it is theoretically possible to directly estimate  
 124 branch lengths in coalescent time units using multi-species coalescent approaches (e.g.,  
 125 ASTRAL-2), in reality these can be severely underestimated due to high levels of gene  
 126 tree error<sup>14</sup>, which is evidently the case with our dataset (Figs. S1, S2). Thus, to test  
 127 whether this assumption is met for all major otophysan groups (i.e., cypriniforms,  
 128 gymnotiforms, siluriforms characoids and citharinoids; Fig. 1), we first estimated the  
 129 length of the branches subtending these clades in millions of years using fossil-calibrated  
 130 phylogenies. A time-calibrated version of our tree and trees published by others<sup>15,16</sup> show  
 131 that the subtending stem lineages leading to these clades are 20–65 million years (MY)  
 132 (Table S7). With these branch length values in MY (T) it is thus possible to estimate  
 133 coalescent time units, given by  $T/(2N_e \times \text{generation time})$  for diploid species, where  $N_e$  is  
 134 effective population size<sup>12</sup>. Assuming a conservative estimate of  $N_e = 100000$  and a  
 135 generation time of 5 years, the minimum estimated length in coalescent units for all  
 136 otophysan stem lineages is 20–65 (see also range of estimates in Table S8). Note that  
 137 these parameters are conservatively estimated. For instance, freshwater fishes are

138 typically confined to small geographic regions with restricted breeding and their  
139 population sizes are likely smaller<sup>17</sup>. Furthermore, the mean age at maturation for several  
140 freshwater fishes is estimated at 2.5 years (North American species)<sup>18</sup>. These estimates of  
141 coalescent units nonetheless suggest that the assumption of subclade monophyly is met in  
142 this case for the vast majority of genes in the otophysan dataset (Table S8). It is possible  
143 that for some genes allelic polymorphisms are maintained by selection for much longer  
144 periods of time resulting in deeper coalescence events than those predicted under the  
145 neutral model. Nevertheless, as discussed in the main text, it seems unlikely that those  
146 isolated instances of deep coalescences would introduce systematic bias into our GGI  
147 procedure.

148 Fig S1 (right) shows that individual (unconstrained) gene trees support  
149 monophyly of these undisputed clades with very low frequency. But Fig. S2 also shows  
150 that gene tree error is rampant (hence our motivation to devise GGI initially). Because the  
151 signal supporting the monophyly of each subclade is strong (morphological,  
152 mitochondrial, multi-locus, genomic) and the coalescent estimations above indicate  
153 ample time for gene trees to achieve subclade monophyly, the reason that gene trees do  
154 not resolve these subclades in most cases is likely due to a combination of poor signal  
155 and systematic error, as we discuss in the main text, rather than ILS.

156 In addition to Otophysi, Table S8 presents estimates of branch lengths in  
157 coalescent time units for subclades in the birds and mammal datasets (no fossil-calibrated  
158 trees were found for metazoans and yeasts). Note that although some of these estimates  
159 fall below the minimum thresholds for genome-wide monophyly (i.e., 5.3–8.3 units<sup>13</sup>),  
160 modified GGI-based coalescent analyses that use a combination of unconstrained and  
161 constrained gene trees obtained results identical to those using the constrained-only  
162 version (see below and Figs. 5, S6, S7).

#### 164 **GGI analyses based on a mixture of constrained and unconstrained tree searches.**

165 During early stages of our study, an anonymous reviewer proposed a different approach  
166 for GGI to relax the constraints of subclade monophyly applied to each gene tree. The  
167 proposal involved conducting AU tests to compare the constrained gene tree topologies  
168 (shown in Fig. 1) against the unconstrained ML tree for each gene. If the unconstrained  
169 ML tree is not significantly better than the second-ranked constrained tree according to  
170 the AU test, then it is reasonable and justified to take the constrained tree as a better  
171 estimate of the gene tree. But in cases where the unconstrained ML tree is significantly  
172 better than all constrained trees then the former should be selected and used for species  
173 tree analyses. For the Otophysi dataset that implies using 16 rather than 15 gene tree  
174 searches for each AU test (15 constrained and one unconstrained trees).

175 We performed such tests using both DNA and protein gene alignments and find  
176 that for 60–70% of the genes in the Otophysi dataset the unconstrained tree is  
177 significantly better than the second-ranked constrained tree (Fig. S6). This result is,  
178 however, neither unexpected (see Figs. S1, S2) nor entirely relevant for the GGI approach  
179 since we are interested in knowing which of the 15 plausible hypotheses is a better fit for  
180 the data in each gene alignment. For example, assume that for a given gene history deep  
181 coalescences result in gymnotiform paraphyly (i.e., some gymnotiforms are closer to  
182 siluriforms than to other gymnotiforms), but the monophyly of all other clades in the  $H_0$   
183 tree is still supported (i.e., citharinoids, characoids, characiforms, siluriforms,

184 gymnotiforms + siluriforms). In this case while the unconstrained tree is expected to be  
 185 significantly better than any of the constrained trees due to gymnotiform paraphyly, the  
 186 predefined genealogy to be favored would still be  $H_0$  (i.e., as in a model-fitting  
 187 framework). Several factors other than ILS associated with gene tree estimation error  
 188 may lead to significant incongruence between an unconstrained ML gene tree and each of  
 189 the reasonable hypotheses represented by the 15 constraints (e.g., model misspecification,  
 190 systematic biases, horizontal gene transfer, hidden paralogy<sup>19</sup>), but understanding the  
 191 forces behind this incongruence is not the main goal of this procedure. Instead, we are  
 192 attempting to identify the models (plausible evolutionary histories) that are the best fit for  
 193 each gene partition by ranking them according to the AU test. In addition, we emphasize  
 194 that the coalescent time unit estimations shown above suggest that all otophysan  
 195 subclades should be monophyletic for the vast majority of gene trees.

196 Noteworthy, the overall results obtained with GGI are robust to the usage of  
 197 unconstrained ML gene trees (Fig. S6). We find that although most unconstrained ML  
 198 gene trees are statistically better than rank 2 constrained trees (714 for DNA and 594 for  
 199 proteins;  $P=0.05$ ), analyses including these unconstrained ML trees still find the same  
 200 result obtained with constrained searches alone (Figs. 2A-B):  $H_0$  is supported much more  
 201 often than  $H_{a10}$ , the second most frequent topology (260 vs. 25 for DNA, and 275 vs. 32  
 202 for proteins; Fig. S6). Furthermore, the GGI-based species tree analyses using a  
 203 combination of constrained and significantly better unconstrained ML gene trees as input  
 204 for ASTRAL-2 support the  $H_0$  tree for both DNA or protein alignments (Fig. S6).

205 In addition to otophysans, the AU tests comparing unconstrained and constrained  
 206 ML gene trees find similar results for metazoans, with the GGI-based ASTRAL-2 tree  
 207 supporting the Ctenophora-sister hypothesis (Fig. S7A). Similar results were also  
 208 obtained for mammals and yeasts, except that in these two cases the number of  
 209 unconstrained trees that are significant is much smaller (4–8%; Figs. S7B-C). For both  
 210 datasets, the ASTRAL-2 analyses find the same topologies as those favored by GGI  
 211 (Figs. 2E-F). While we also analysed the Neoaves dataset using the unconstrained and  
 212 constrained ML gene trees (Fig. S7D) we did not use these as input for ASTRAL-2  
 213 because, as discussed in the paper, this requires testing 105 competing topologies in  
 214 addition to the unconstrained trees – an analysis beyond the scope of this study. The  
 215 pattern of support among constrained gene trees for birds (Fig. S7D) is, however, similar  
 216 to that reported in Fig. 2D.

217 Taken together, these analyses indicate that our initial results are robust to  
 218 possible violations of the assumption of subclade monophyly across the vast majority of  
 219 genes. Even if independent estimates of branch lengths in coalescent time units suggest  
 220 high probability of monophyly (Table S8), we advise potential users of GGI to test the  
 221 robustness of their results using a combination of constrained and unconstrained trees  
 222 searches as input for summary coalescent methods.

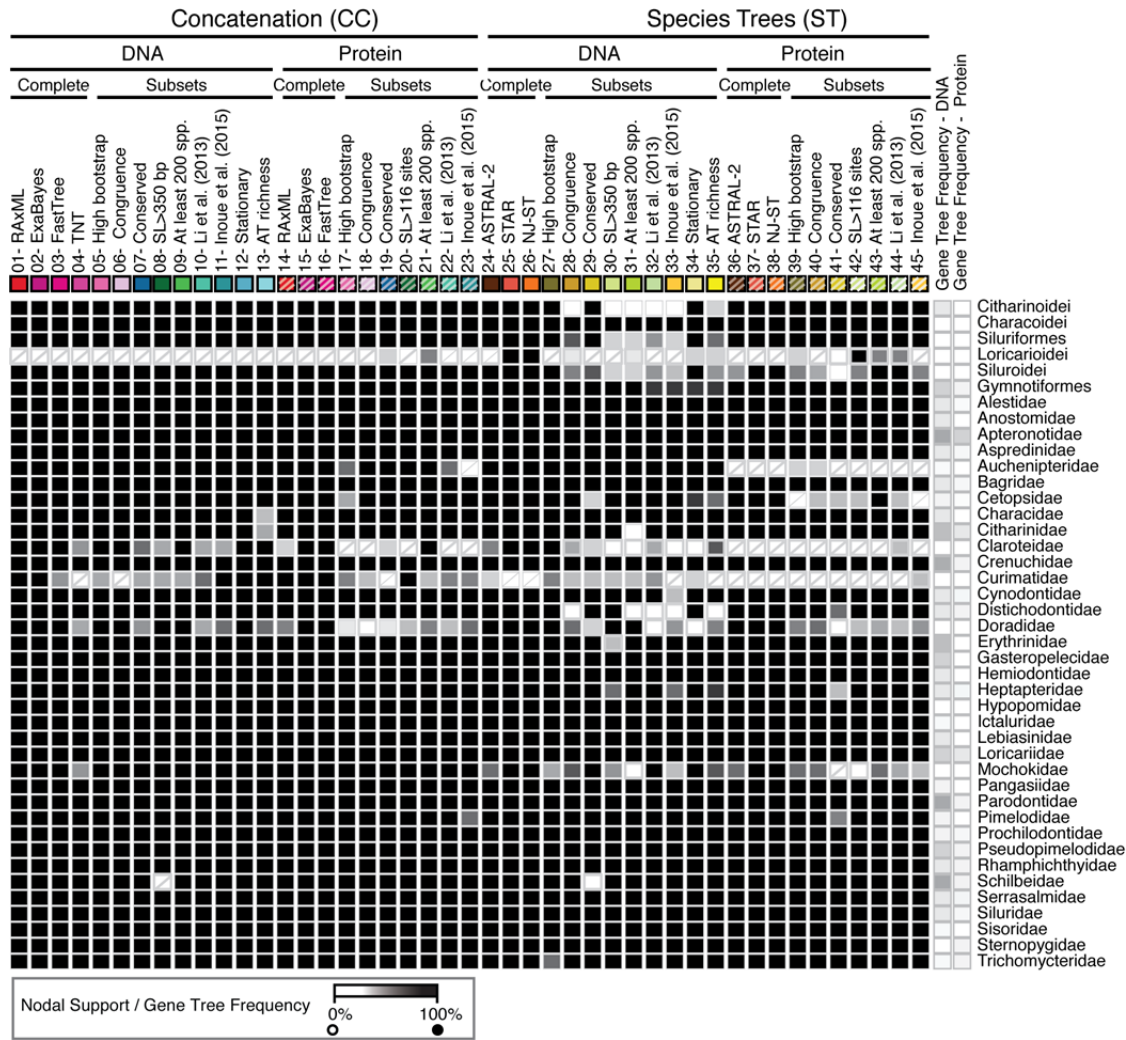
223

- 224 1 Li, C., Riethoven, J. J. & Naylor, G. J. P. EvolMarkers: a database for mining  
 225 exon and intron markers for evolution, ecology and conservation studies. *Mol*  
 226 *Ecol Resour* **12**, 967-971 (2012).  
 227 2 Faircloth, B. C. *et al.* Ultraconserved Elements Anchor Thousands of Genetic  
 228 Markers Spanning Multiple Evolutionary Timescales. *Systematic Biology* **61**,  
 229 717–726, doi:10.1093/sysbio/sys004 (2012).



- 230 3 Mamanova, L. *et al.* Target-enrichment strategies for next-generation sequencing.  
231 *Nat Meth* **7**, 111-118,  
232 doi:[http://www.nature.com/nmeth/journal/v7/n2/supinfo/nmeth.1419\\_S1.html](http://www.nature.com/nmeth/journal/v7/n2/supinfo/nmeth.1419_S1.html)  
233 (2010).
- 234 4 Lemmon, A. R., Emme, S. A. & Lemmon, E. M. Anchored Hybrid Enrichment  
235 for Massively High-Throughput Phylogenomics. *Systematic Biology* **61**, 727-744,  
236 doi:10.1093/sysbio/sys049 (2012).
- 237 5 Clement, N. L. *et al.* The GNUMAP algorithm: unbiased probabilistic mapping of  
238 oligonucleotides from next-generation sequencing. *Bioinformatics (Oxford,*  
239 *England)* **26**, 38-45, doi:10.1093/bioinformatics/btp614 (2010).
- 240 6 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using  
241 the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-  
242 1512, doi:10.1038/nprot.2013.084 (2013).
- 243 7 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software  
244 Version 7: Improvements in Performance and Usability. *Molecular Biology and*  
245 *Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
- 246 8 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-  
247 analysis of large phylogenies. *Bioinformatics (Oxford, England)* **30**, 1312-1313,  
248 doi:10.1093/bioinformatics/btu033 (2014).
- 249 9 Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree  
250 estimation. *Bioinformatics (Oxford, England)* **30**, i541-548,  
251 doi:10.1093/bioinformatics/btu462 (2014).
- 252 10 Near, T. J. *et al.* Resolution of ray-finned fish phylogeny and timing of  
253 diversification. *Proceedings of the National Academy of Sciences* **109**, 13698-  
254 13703, doi:10.1073/pnas.1206625109 (2012).
- 255 11 Betancur-R., R. *et al.* The tree of life and a new classification of bony fishes.  
256 *PLoS Currents Tree of Life* **2013 Apr 18**,  
257 doi:10.1371/currents.tol.53ba26640df0cacee75bb165c8c26288 (2013).
- 258 12 Degnan, J. H. & Rosenberg, N. A. Gene tree discordance, phylogenetic inference  
259 and the multispecies coalescent. *Trends in Ecology & Evolution* **24**, 332-340,  
260 doi:10.1016/j.tree.2009.01.009 (2009).
- 261 13 Rosenberg, N. A. The shapes of neutral gene genealogies in two species:  
262 probabilities of monophyly, paraphyly, and polyphyly in a coalescent model.  
263 *Evolution* **57**, 1465-1477 (2003).
- 264 14 Sayyari, E. & Mirarab, S. Fast Coalescent-Based Computation of Local Branch  
265 Support from Quartet Frequencies. *Mol Biol Evol* **33**, 1654-1668,  
266 doi:10.1093/molbev/msw079 (2016).
- 267 15 Chen, W. J., Lavoue, S. & Mayden, R. L. Evolutionary origin and early  
268 biogeography of otophysan fishes (Ostariophysi: Teleostei). *Evolution* **67**, 2218-  
269 2239, doi:10.1111/evo.12104 (2013).
- 270 16 Nakatani, M., Miya, M., Mabuchi, K., Saitoh, K. & Nishida, M. Evolutionary  
271 history of Otophysi (Teleostei), a major clade of the modern freshwater fishes:  
272 Pangaeian origin and Mesozoic radiation. *BMC Evolutionary Biology* **11**, 177,  
273 doi:10.1186/1471-2148-11-177 (2011).

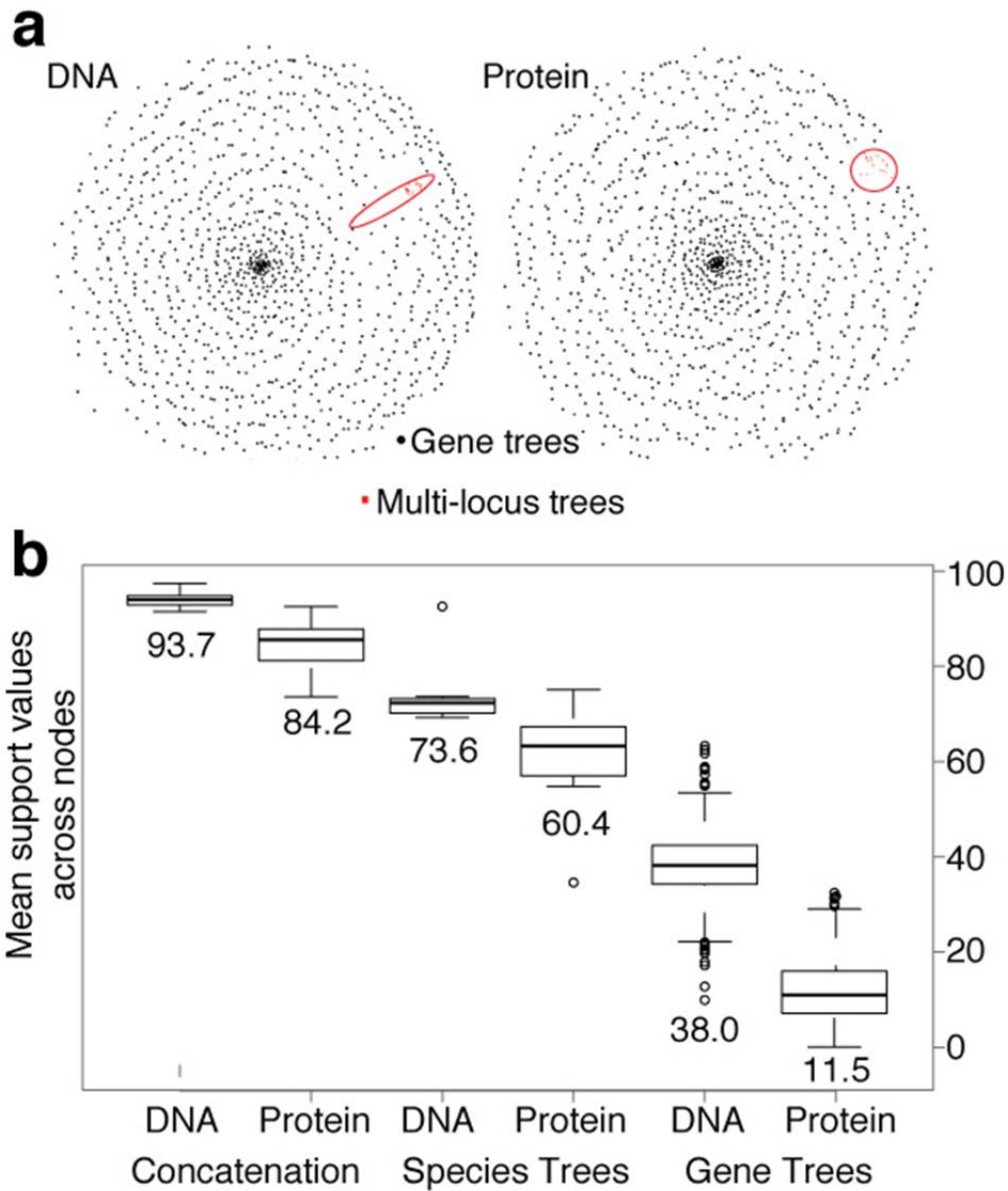
- 274 17 Yi, S. & Streebman, J. T. Genome size is negatively correlated with effective  
275 population size in ray-finned fish. *Trends Genet* **21**, 643-646,  
276 doi:10.1016/j.tig.2005.09.003 (2005).
- 277 18 Mims, M. C., Olden, J. D., Shattuck, Z. R. & Poff, N. L. Life history trait  
278 diversity of native freshwater fishes in North America. *Ecology of Freshwater*  
279 *Fish* **19**, 390-400, doi:10.1111/j.1600-0633.2010.00422.x (2010).
- 280 19 Maddison, W. P. Gene Trees in Species Trees. *Systematic Biology* **46**, 523-536  
281 (1997).
- 282 20 Fink, S. V. & Fink, W. L. Interrelationships of the Ostariophysan Fishes  
283 (Teleostei). *Zoological Journal of the Linnean Society* **72**, 297-353 (1981).
- 284 21 Dimmick, W. W. & Larson, A. A molecular and morphological perspective on the  
285 phylogenetic relationships of the otophysan fishes. *Molecular Phylogenetics and*  
286 *Evolution* **6**, 120-133 (1996).
- 287 22 Alves-Gomes, J. A. in *Gonorynchiformes and Ostariophysan relationships* (eds  
288 T. Grande, F. J. Potayo-Ariza, & R. Diogo) (Science Publishers, 2010).
- 289 23 Saitoh, K., Miya, M., Inoue, J. G., Ishiguro, N. B. & Nishida, M. Mitochondrial  
290 genomics of ostariophysan fishes: Perspectives on phylogeny and biogeography.  
291 *Journal of Molecular Evolution* **56**, 464-472, doi:10.1007/s00239-002-2417-y  
292 (2003).
- 293 24 Lavoue, S. *et al.* Molecular systematics of the gonorynchiform fishes (Teleostei)  
294 based on whole mitogenome sequences: Implications for higher-level  
295 relationships within the Otocephala. *Molecular Phylogenetics and Evolution* **37**,  
296 165-177, doi:10.1016/J.Ympev.2005.03.024 (2005).
- 297 25 Peng, Z., Diogo, R. & He, S. Teleost fishes (Teleostei). *The timetree of life*, 335 -  
298 337 (2009).
- 299 26 Chakrabarty, P., McMahan, C., Fink, W., Stiassny, M. L. & Alfaro, M. in *ASIH –*  
300 *American Society of Ichthyologists and Herpetologists*. (eds M. L. Crump & M.  
301 A. Donnelly).
- 302 27 Betancur, R. R., Orti, G. & Pyron, R. A. Fossil-based comparative analyses reveal  
303 ancient marine ancestry erased by extinction in ray-finned fishes. *Ecol Lett* **18**,  
304 441-450, doi:10.1111/ele.12423 (2015).
- 305 28 Meredith, R. W. *et al.* Impacts of the Cretaceous Terrestrial Revolution and KPg  
306 Extinction on Mammal Diversification. *Science* **334**, 521-524,  
307 doi:10.1126/science.1211028 (2011).
- 308 29 Hedges, S. B., Marin, J., Suleski, M., M., P. & Kumar, S. Tree of Life Reveals  
309 Clock-Like Speciation and Diversification. *Mol Biol Evol* **2015**, 835-845 (2015).
- 310 30 Martin, A. P. & Palumbi, S. R. Body size, metabolic rate, generation time, and the  
311 molecular clock. *Proc. Natl. Acad. Sci. USA*, 4087-4091 (1993).
- 312 31 Charlesworth, B. Fundamental concepts in genetics: effective population size and  
313 patterns of molecular evolution and variation. *Nat Rev Genet* **10**, 195-205 (2009).
- 314 32 Prum, R. O. *et al.* A comprehensive phylogeny of birds (Aves) using targeted  
315 next-generation DNA sequencing. *Nature* **526**, 569-573, doi:10.1038/nature15697  
316 (2015).
- 317 33 Saether, B. E. *et al.* Generation time and temporal scaling of bird population  
318 dynamics. *Nature* **436**, 99-102 (2005).



320  
321  
322  
323  
324  
325  
326  
327  
328  
329

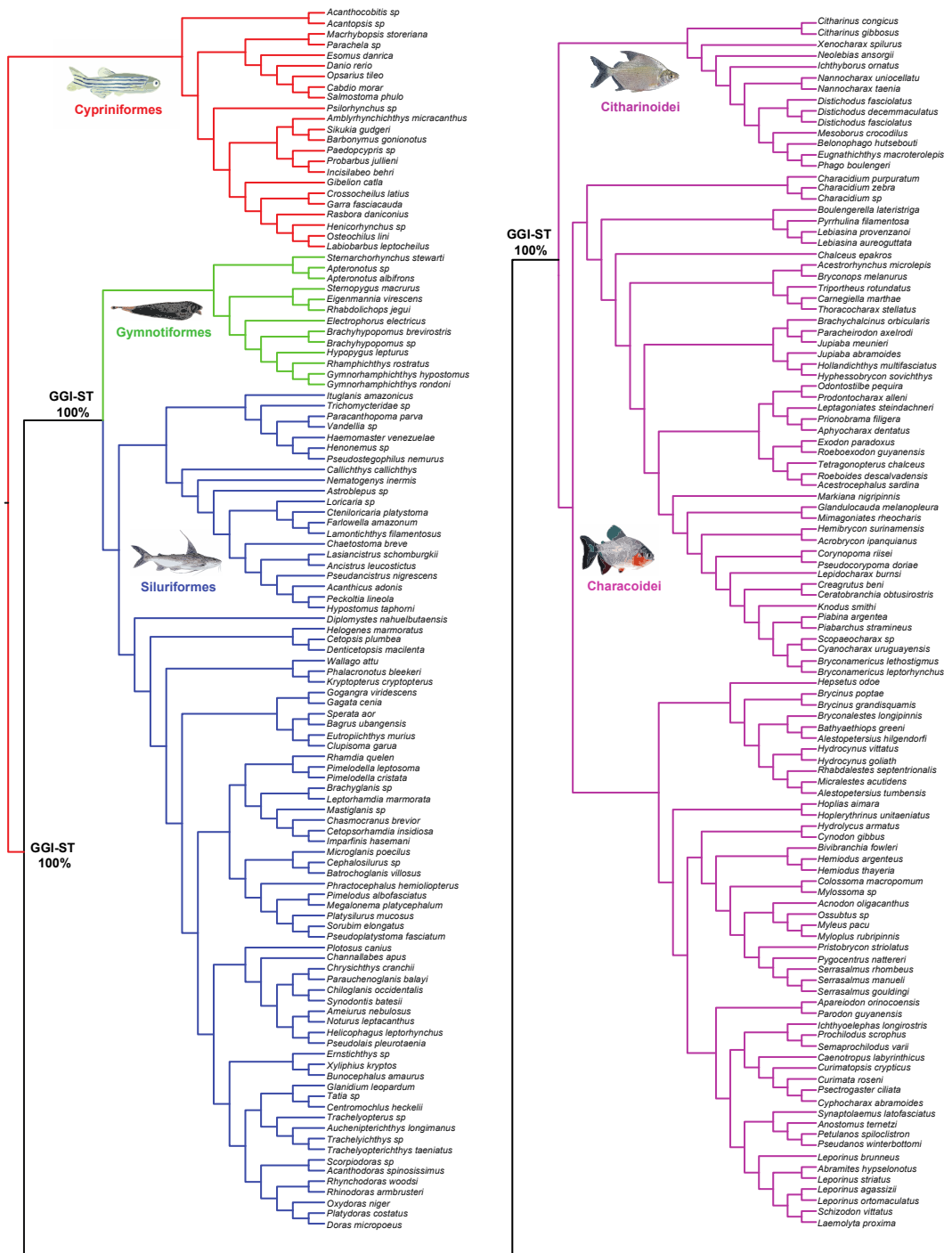
**Figure S1. Resolution and support for undisputed taxonomic groups (orders, suborders and families) that are independent of the main hypotheses tested (Fig. 1) |** Results from 45 multi-locus analyses show high levels of concordance across methods in terms of presence/absence of clades (absences denoted by oblique lines) as well as nodal support (grey scale). By contrast, a large proportion of individual gene trees fail to resolve these groups. These comparisons highlight differences in phylogenetic accuracy and congruence between multi-locus analyses and individual gene trees. Twelve families represented by single individuals in our dataset were not tested.





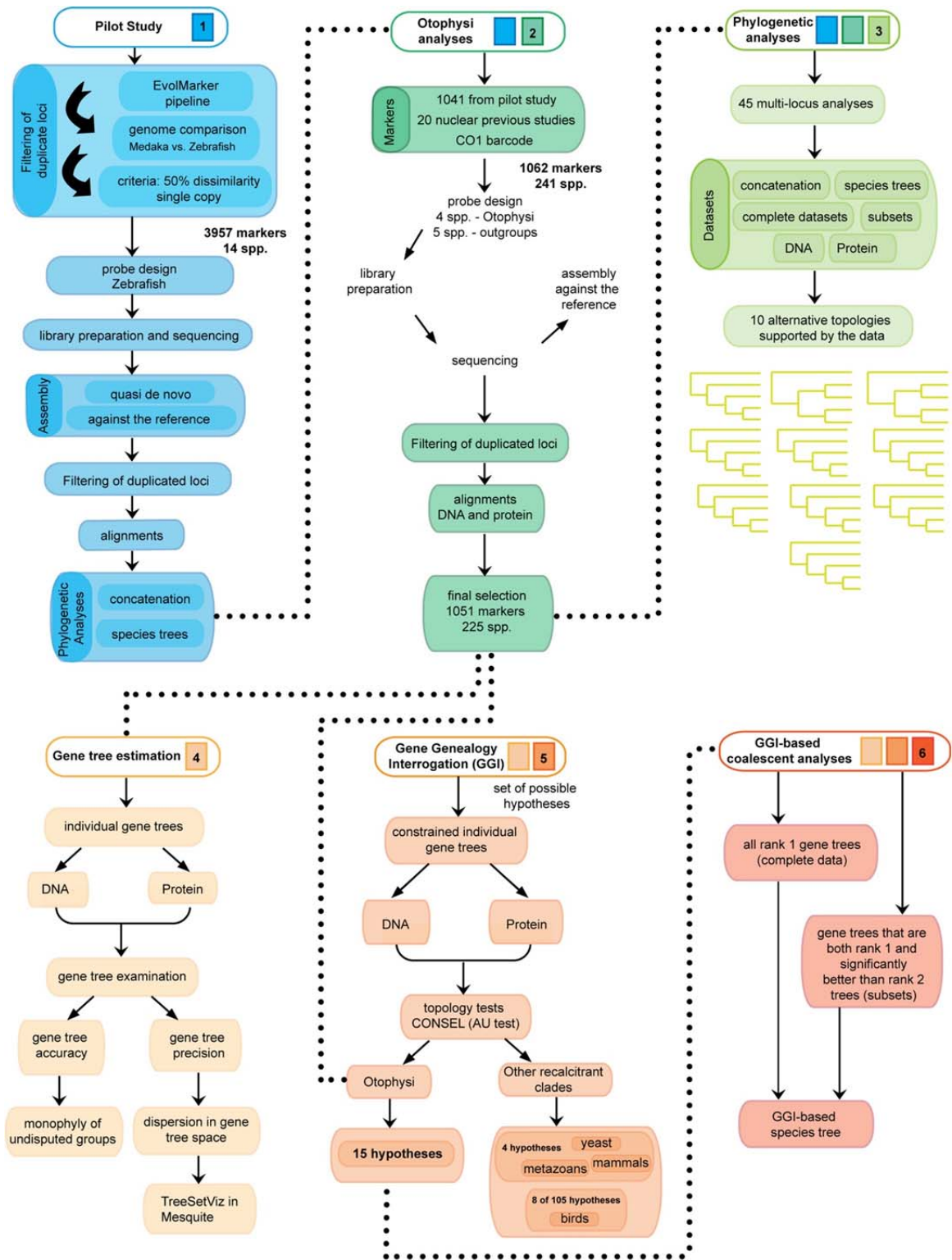
339  
340  
341  
342  
343  
344  
345

**Figure S2. Assessments of phylogenetic precision for multi-locus trees and individual gene trees** | **a**, dispersion in multidimensional tree space based on un-weighted Robinson-Foulds distances; **b**, plots of mean support values across all nodes for each tree obtained with different methods (ExaBayes: posterior probabilities; all other analyses: bootstrap values).



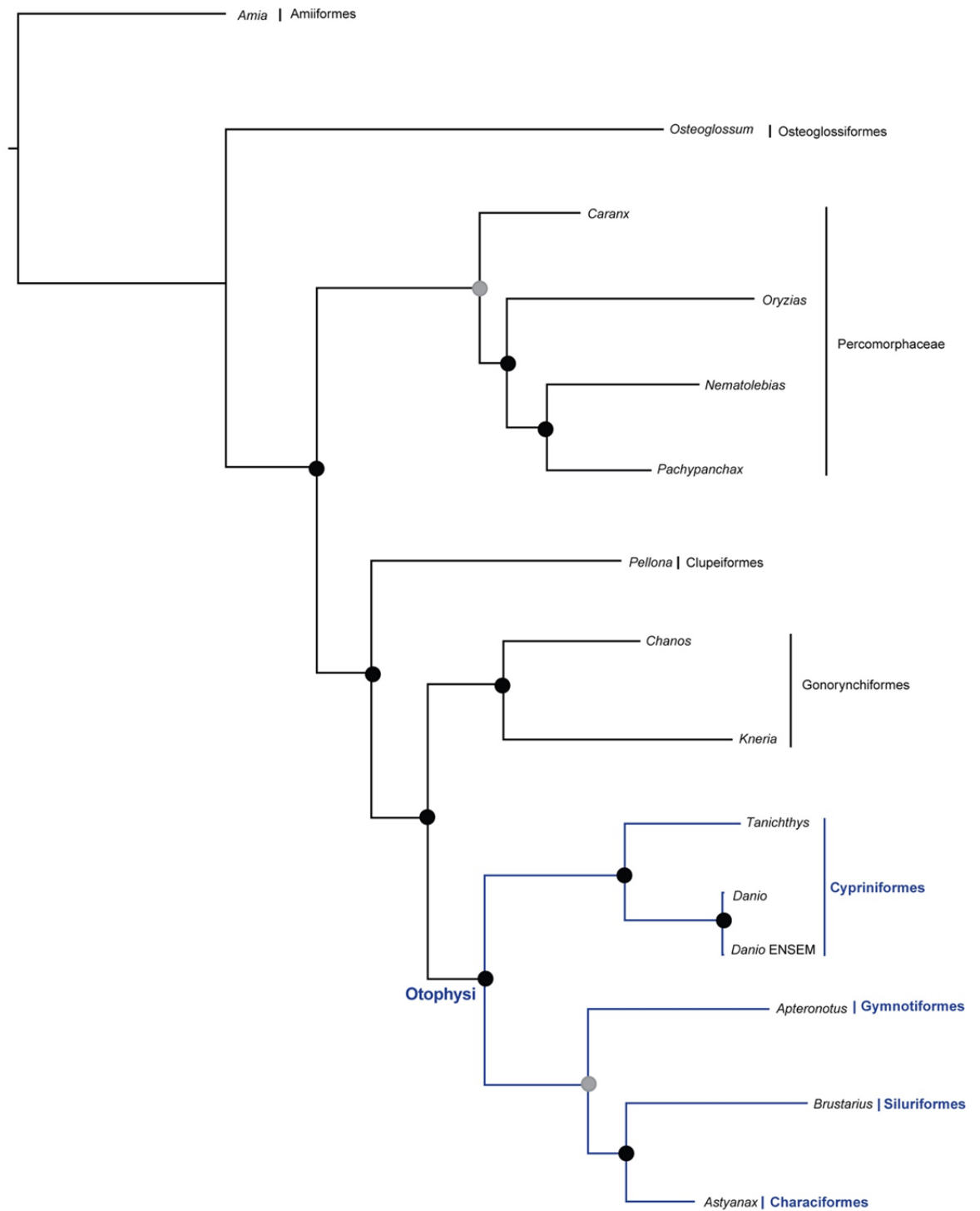
353  
354  
355

**Figure S3. GGI-based species trees for Otophysi using ASTRAL-2 (complete dataset).**



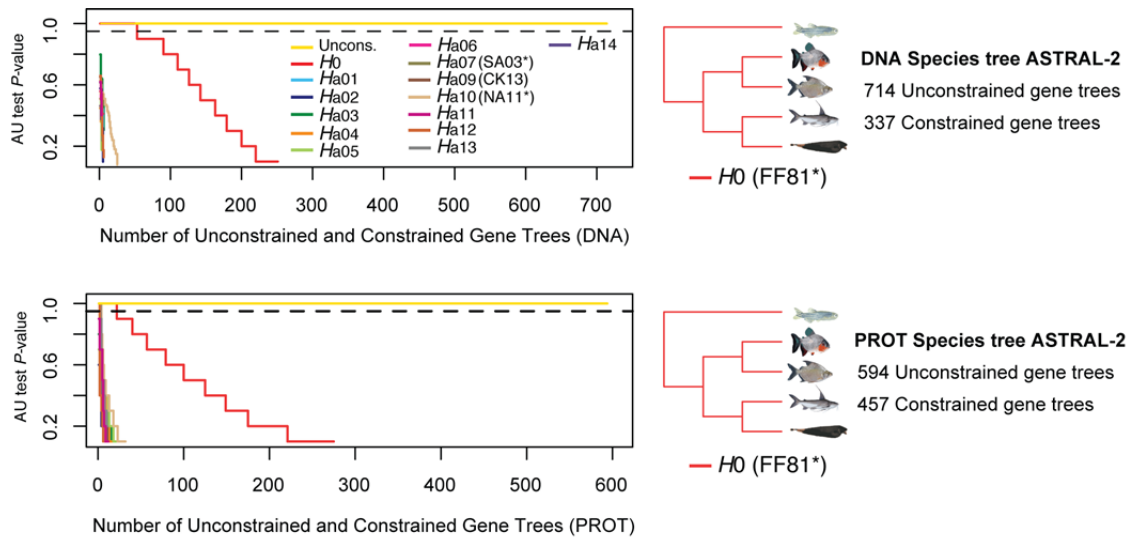
357  
358

**Figure S4. Flowchart of experimental design and methodological approaches used.**



359  
360

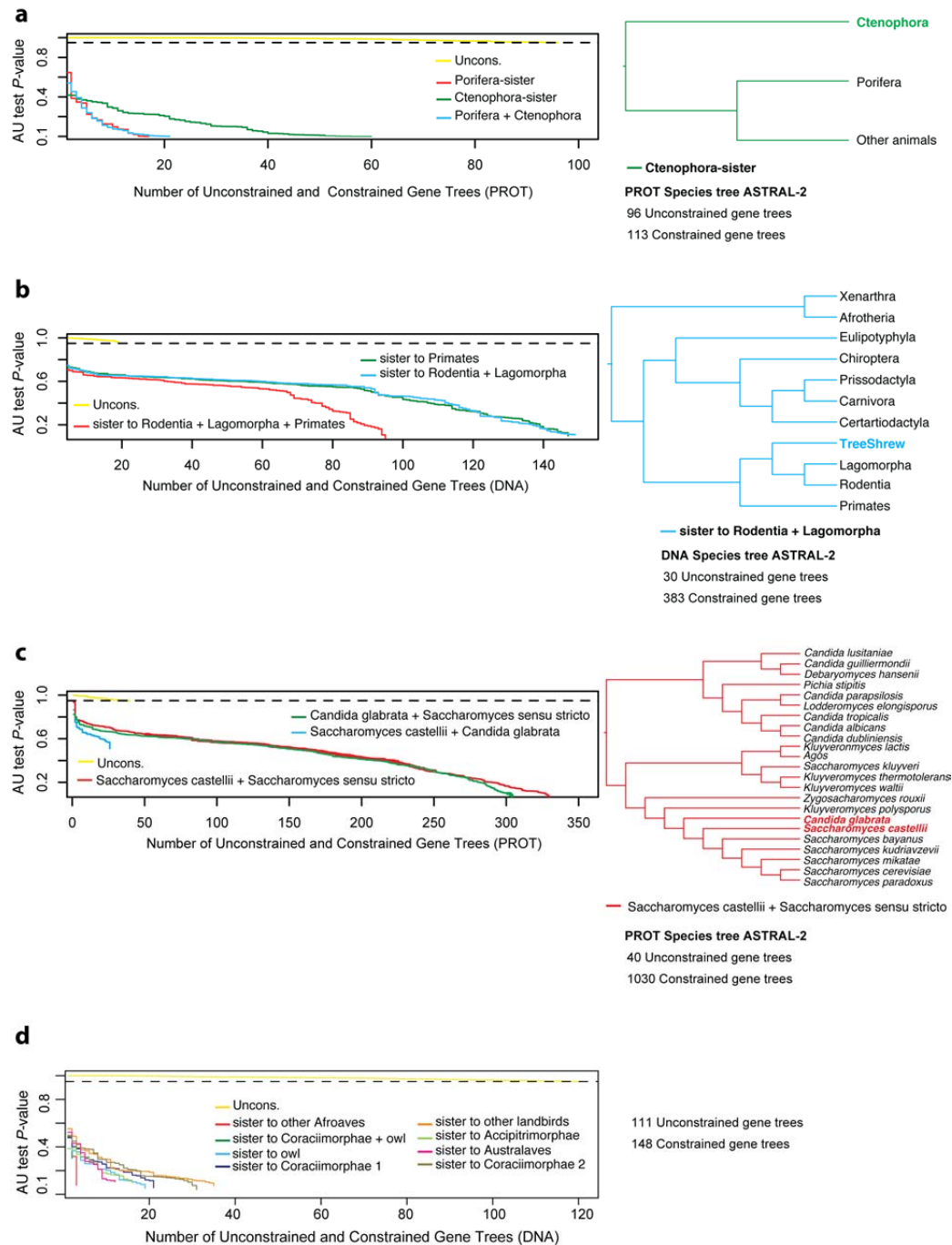
361 **Figure S5. Phylogenomic tree of selected ray-finned fish from pilot study** | Both  
362 phylogenetic methods conducted (concatenation in RAxML; species tree in ASTRAL-2)  
363 resulted in the same topology, with high bootstrap support (black circles 100% BS; grey  
364 circles 75–85% BS).



366

367 **Figure S6. Testing the assumption of subclade monophyly under GGI, using a**  
 368 **combination of significantly better unconstrained and constrained gene trees for**  
 369 **Otophysi** | Lines represent the cumulative number of genes with phylogenetic signal (X-  
 370 axis) rejecting or accepting any of the 15 alternative topologies with highest probability  
 371 in favour of the unconstrained or a constrained ML tree and their associated *P*-values (Y-  
 372 axis) according to the approximately unbiased (AU) topology test. Coalescent species  
 373 trees analyses used both constrained and significantly-better unconstrained gene  
 374 trees as input for ASTRAL-2 (\*Table S1).





375

376 **Figure S7. Testing the assumption of subclade monophyly under GGI, using a**  
 377 **combination of significantly better unconstrained and constrained gene trees for**  
 378 **other groups** | Lines represent the cumulative number of genes with phylogenetic signal  
 379 (X-axis) rejecting or accepting any of the 15 alternative topologies with highest  
 380 probability in favour of the unconstrained or a constrained ML tree and their associated  
 381  $P$ -values (Y-axis) according to the approximately unbiased (AU) topology test.  
 382 Coalescent species trees analyses used both constrained and significantly-better  
 383 unconstrained gene trees as input for ASTRAL-2. **a**, metazoans; **b**, eutherian  
 384 mammals (treeshrew); **c**, yeast phylogeny; **d**, Neoaves (mousebird).

385 **Table S1.** Data used by previous studies addressing interrelationships among major  
 386 lineages in *Otophysi*.

Study	Code	Markers	Length bp/chrs	Number of Species
Fink and Fink <sup>20</sup>	FF81	Morphology	127	
Dimmick and Larson <sup>21</sup>	DL96	1 nuclear gene and mt rDNA	2,477	9
Alves-Gomes <sup>22</sup>	AG10	mt rDNA	701	60
Saitoh et al. <sup>23</sup>	SA03	mt genome	8,096	13
Near et al. <sup>10</sup>	NE12	10 nuclear genes	7,587	16
Lavoue et al. <sup>24</sup>	LA05	mt genome	10,395	13
Peng et al. <sup>25</sup>	PO09	mt genome	6,198	17
Betancur-R et al. <sup>11</sup>	BE13	20 nuclear genes and mt rDNA	20,853	98
Chakrabarty et al. <sup>26</sup>	CK13	351 UCEs	133,424	35
Nakatani et al. <sup>16</sup>	NA11	10 nuclear genes and mt genome	17,918	66
Chen et al. <sup>15</sup>	CH13	5 nuclear genes	4,518	95

387 mt= mitochondrial

412 **Table S2.** Description of the 45 multi-locus datasets analysed, including information on  
 413 number of loci, total length of the alignments (sites), and number of parsimony  
 414 informative sites.

Analyses	Method	DNA			Protein		
		Loci	Sites	Parsim- IS	Loci	Sites	Parsim- IS
RAxML	Concatenation	1051	279,012	123,714	1051	93,457	27,055
ExaBayes		1051	279,012	123,714	1051	93,457	27,055
FastTree		1051	279,012	123,714	1051	93,457	27,055
TNT		1051	279,012	123,714	1051	93,457	27,055
High bootstrap		200	65,709	31,189	200	20,954	8,987
Congruence		210	61,914	28,065	210	21,154	8,394
Conserved		200	48,348	17,660	200	15,791	1,487
SL > 350*/100** sites		205	60,225	27,446	205	28,907	10,026
At least 200 spp.		231	68,682	30,657	231	22,441	5,965
Li et al. (2013)		243	60,147	26,595	243	20,049	5,106
Inoue et al. (2015)		175	44,559	19,651	175	14,853	3,878
Stationarity (Low DI)		200	46,677	20,118	n/a	n/a	n/a
AT richness		200	52,809	23,763	n/a	n/a	n/a
ASTRAL-2		Species Trees	1051	279,012	123,714	1051	93,457
STAR	1050		279,012	123,714	1051	93,457	27,055
NJst	1050		279,012	123,714	1051	93,457	27,055
High bootstrap	200		65,709	31,189	200	20,954	8,987
Congruence	210		61,914	28,065	210	21,154	8,394
Conserved	200		48,348	17,660	200	15,791	1,487
SL > 350*/100** sites	205		60,225	27,446	205	28,907	10,026
At least 200 spp.	231		68,682	30,657	231	22,441	5,965
Li et al. (2013)	243		60,147	26,595	243	20,049	5,106
Inoue et al. (2015)	175		44,559	19,651	175	14,853	3,878
Stationarity (Low DI)	200		46,677	20,118	n/a	n/a	n/a
AT richness	200	52,809	23,763	n/a	n/a	n/a	

415 SL> = sequence length greater than 350 sites (\*DNA alignments) or 96 sites (\*\*protein  
 416 alignments); Pars-IS = parsimony informative sites; DI = disparity index; n/a = not  
 417 applicable

424 **Table S3.** Summary of the results using gene genealogy interrogation (GGI) on the five  
425 datasets analysed

Hypothesis tested	DNA		PROT	
	Number of rank 1 gene trees	Number of rank 1 gene trees that are also significant	Number of rank 1 gene trees	Number of rank 1 gene trees that are also significant
<b>Otophysi</b>				
<i>HO</i> (FF81*)	459	327	314	197
<i>Ha01</i>	31	0	57	1
<i>Ha02</i>	44	0	51	4
<i>Ha03</i>	40	0	47	9
<i>Ha04</i>	40	0	56	0
<i>Ha05</i>	51	0	86	5
<i>Ha06</i>	44	1	72	5
<i>Ha07</i> (SA03*)	22	0	56	4
<i>Ha08</i> (LA05*)	0	0	0	0
<i>Ha09</i> (CK13)	33	0	55	6
<i>Ha10</i> (NA11*)	174	69	146	39
<i>Ha11</i>	39	0	45	2
<i>Ha12</i>	25	0	22	1
<i>Ha13</i>	24	0	23	2
<i>Ha14</i>	25	0	21	0
<b>Metazoans</b>				
Porifera-sister			38	0
Ctenophora-sister			134	7
Porifera + Ctenophora			37	1
<b>Yeasts</b>				
<i>Saccharomyces castellii</i> + <i>Saccharomyces s.s</i>			426	38
<i>Candida glabrata</i> + <i>Saccharomyces s.s</i>			332	24
<i>Saccharomyces castellii</i> + <i>Candida glabrata</i>			312	17
<b>Eutherian mammals</b>				
Sister to Rodentia + Lagomorpha + Primates	97	1		
Sister to Primates	155	3		
Sister to Rodentia + Lagomorpha	162	7		
<b>Neoavian birds</b>				
Sister to other Afroaves	12	9		
Sister to Coraciimorphae + owl	4	0		
Sister to owl	35	2		
Sister to Coraciimorphae 1	31	1		
Sister to other landbirds	66	2		
Sister to Accipitrimorphae	31	1		
Sister to Australaves	25	0		
Sister to Coraciimorphae 2	55	1		

426 \* Hypotheses resolved by more than one previous study; *s.s* = sensu strict

427 **Table S4.** General properties of all datasets analysed to test the generality of the Gene  
 428 Genealogy Interrogation approach.  
 429

<b>Dataset description</b>	<b>Number of Genes</b>	<b>Number of Taxa</b>	<b>Number of Hypotheses</b>	<b>Number of Sites</b>
Otophysi	1,051	225	15	279,012
Metazoa	209	76	3	60,768
Birds (mousebird)	259	90	8*	390,000
Mammals (tree shrew)	413	18	3	139,600
Yeast	1,070	23	3	20,289

430 \* Only 8 (out of 105) possible unrooted trees were



466  
467

**Table S5.** List of specimens examined and catalog numbers.

Order	Family	Genus	Species	Catalog Number
Characiformes	Acestrorhynchidae	<i>Acestrorhynchus</i>	<i>microlepis</i>	AUM 3537
Characiformes	Alestidae	<i>Alestopetersius</i>	<i>hilgendorfi</i>	AMNH 034-3321
Characiformes	Alestidae	<i>Bathyaethiops</i>	<i>greeni</i>	AMNH 024-2395
Characiformes	Alestidae	<i>Brycinus</i>	<i>grandisquamis</i>	AMNH 115-11429
Characiformes	Alestidae	<i>Brycinus</i>	<i>poptae</i>	AMNH 025-2455
Characiformes	Alestidae	<i>Bryconalestes</i>	<i>longipinnis</i>	AUM 5548
Characiformes	Alestidae	<i>Alestopetersius</i>	<i>tumbensis</i>	AMNH 052-5189
Characiformes	Alestidae	<i>Hydrocynus</i>	<i>goliath</i>	AMNH 022-2138
Characiformes	Alestidae	<i>Hydrocynus</i>	<i>vittatus</i>	AMNH 075-7445
Characiformes	Alestidae	<i>Micralestes</i>	<i>acutidens</i>	AUM 5674
Characiformes	Alestidae	<i>Rhabdalestes</i>	<i>septentrionalis</i>	AMNH 113-11282
Characiformes	Anostomidae	<i>Abramites</i>	<i>hypselonotus</i>	GO 77
Characiformes	Anostomidae	<i>Anostomus</i>	<i>ternetzi</i>	AUM 3075
Characiformes	Anostomidae	<i>Laemolyta</i>	<i>proxima</i>	AUM 3059
Characiformes	Anostomidae	<i>Leporinus</i>	<i>agassizii</i>	AUM 3060
Characiformes	Anostomidae	<i>Leporinus</i>	<i>brunneus</i>	AUM 3071
Characiformes	Anostomidae	<i>Leporinus</i>	<i>ortomaculatus</i>	AUM 3065
Characiformes	Anostomidae	<i>Leporinus</i>	<i>striatus</i>	AUM 3064
Characiformes	Anostomidae	<i>Petulanos</i>	<i>spiloclistron</i>	AUM 3057
Characiformes	Anostomidae	<i>Pseudanos</i>	<i>winterbottomi</i>	AUM 3084
Characiformes	Anostomidae	<i>Schizodon</i>	<i>vittatus</i>	AMNH 075-7445
Characiformes	Anostomidae	<i>Synaptolaemus</i>	<i>latofasciatus</i>	AUM 3064
Characiformes	Chalceidae	<i>Chalceus</i>	<i>epakros</i>	AUM 3601
Characiformes	Characidae	<i>Acestrocephalus</i>	<i>sardina</i>	AUM 3532
Characiformes	Characidae	<i>Acrobrycon</i>	<i>ipanquianus</i>	ANSP 180776
Characiformes	Characidae	<i>Aphyocharax</i>	<i>dentatus</i>	UFRGS13585 TEC1707A
Characiformes	Characidae	<i>Jupiaba</i>	<i>abramoides</i>	AUM 4709
Characiformes	Characidae	<i>Brachyhalcinus</i>	<i>orbicularis</i>	AUM 3056
Characiformes	Characidae	<i>Piabarchus</i>	<i>stramineus</i>	UFRGS12898 TEC1049
Characiformes	Characidae	<i>Ceratobranchia</i>	<i>obtusirostris</i>	AUM 4046
Characiformes	Characidae	<i>Paracheirodon</i>	<i>axelrodi</i>	GO 213
Characiformes	Characidae	<i>Corynopoma</i>	<i>riisei</i>	ROM 8900
Characiformes	Characidae	<i>Creagrutus</i>	<i>beni</i>	UFRGS 12811 TEC1521
Characiformes	Characidae	<i>Cyanocharax</i>	<i>uruguayensis</i>	UFRGS 10692 TEC393B
Characiformes	Characidae	<i>Exodon</i>	<i>paradoxus</i>	AUM 3536
Characiformes	Characidae	<i>Glandulocauda</i>	<i>melanopleura</i>	UFRGS 12885 TEC831
Characiformes	Characidae	<i>Hemibrycon</i>	<i>surinamensis</i>	AUM 4711
Characiformes	Characidae	<i>Hollandichthys</i>	<i>multifasciatus</i>	UFRGS 11792 TEC 841A
Characiformes	Characidae	<i>Odontostilbe</i>	<i>pequira</i>	AUM 4047
Characiformes	Characidae	<i>Hyphessobrycon</i>	<i>sovichthys</i>	AUM 4799
Characiformes	Characidae	<i>Bryconamericus</i>	<i>leptorhynchus</i>	UFRGS1812B
Characiformes	Characidae	<i>Jupiaba</i>	<i>meunieri</i>	AUM 4688
Characiformes	Characidae	<i>Knodus</i>	<i>smithi</i>	MUSM 35458
Characiformes	Characidae	<i>Lepidocharax</i>	<i>burnsi</i>	UFRGS 12886 TEC1018
Characiformes	Characidae	<i>Leptagoniates</i>	<i>steindachneri</i>	GO 123
Characiformes	Characidae	<i>Markiana</i>	<i>nigripinnis</i>	UFRGS 1160A
Characiformes	Characidae	<i>Mimagoniates</i>	<i>rheocharis</i>	UFRGS 12896 911
Characiformes	Characidae	<i>Bryconamericus</i>	<i>lethostigmus</i>	UFRGS12537 TEC1239
Characiformes	Characidae	<i>Piabina</i>	<i>argentea</i>	UFRGS 11373 TEC1211
Characiformes	Characidae	<i>Prionobrama</i>	<i>filigera</i>	ANSP 180765
Characiformes	Characidae	<i>Prodontocharax</i>	<i>alleni</i>	ANSP 182951

Characiformes	Characidae	<i>Pseudocorypoma</i>	<i>doriae</i>	UFRGS 12389 TEC693A
Characiformes	Characidae	<i>Roeboexodon</i>	<i>guyanensis</i>	AUM 3605
Characiformes	Characidae	<i>Roebooides</i>	<i>descalvadensis</i>	UFRGS13496 TEC1618B
Characiformes	Characidae	<i>Scopaeocharax</i>	<i>sp</i>	UFRGS 10692 TEC393B
Characiformes	Characidae	<i>Tetragonopterus</i>	<i>chalceus</i>	AUM 3599
Characiformes	Chilodontidae	<i>Caenotropus</i>	<i>labyrinthicus</i>	AUM 3096
Characiformes	Citharinidae	<i>Citharinus</i>	<i>congicus</i>	AMNH 081-8014
Characiformes	Citharinidae	<i>Citharinus</i>	<i>gibbosus</i>	AMNH 102-10177
Characiformes	Crenuchidae	<i>Characidium</i>	<i>sp</i>	AUM 4777
Characiformes	Crenuchidae	<i>Characidium</i>	<i>purpuratum</i>	AUM 4061
Characiformes	Crenuchidae	<i>Characidium</i>	<i>zebra</i>	AUM 4856
Characiformes	Ctenoluciidae	<i>Boulengerella</i>	<i>lateristriga</i>	AUM 3257
Characiformes	Curimatidae	<i>Curimata</i>	<i>roseni</i>	ANSP 189094
Characiformes	Curimatidae	<i>Curimatopsis</i>	<i>crypticus</i>	ANSP 189091
Characiformes	Curimatidae	<i>Cyphocharax</i>	<i>abramoides</i>	AUM 4709
Characiformes	Curimatidae	<i>Psectrogaster</i>	<i>ciliata</i>	ANSP 189093
Characiformes	Cynodontidae	<i>Cynodon</i>	<i>gibbus</i>	AUM 3244
Characiformes	Cynodontidae	<i>Hydrolycus</i>	<i>armatus</i>	AUM 3529
Characiformes	Distichodontidae	<i>Belonophago</i>	<i>hutsebouti</i>	AMNH 025-2468
Characiformes	Distichodontidae	<i>Distichodus</i>	<i>decemmaculatus</i>	AMNH 026-2531
Characiformes	Distichodontidae	<i>Distichodus</i>	<i>fasciolatus</i>	AUM 5670
Characiformes	Distichodontidae	<i>Distichodus</i>	<i>fasciolatus</i>	AMNH 069-6958
Characiformes	Distichodontidae	<i>Eugnathichthys</i>	<i>macroterolepis</i>	AMNH 097-9609
Characiformes	Distichodontidae	<i>Nannocharax</i>	<i>uniocellatus</i>	AMNH 081-8052
Characiformes	Distichodontidae	<i>Ichthyborus</i>	<i>ornatus</i>	AMNH 034-3337
Characiformes	Distichodontidae	<i>Mesoborus</i>	<i>crocodilus</i>	AMNH 097-9623
Characiformes	Distichodontidae	<i>Nannocharax</i>	<i>taenia</i>	AMNH 108-10751
Characiformes	Distichodontidae	<i>Neolebias</i>	<i>ansorgii</i>	GO 48
Characiformes	Distichodontidae	<i>Phago</i>	<i>boulengeri</i>	AMNH 102-10181
Characiformes	Distichodontidae	<i>Xenocharax</i>	<i>spilurus</i>	AMNH 088-8739
Characiformes	Erythrinidae	<i>Hoplerythrinus</i>	<i>unitaeniatus</i>	AUM 4875
Characiformes	Erythrinidae	<i>Hoplias</i>	<i>aimara</i>	AUM 4938
Characiformes	Gasteropelecidae	<i>Carnegiella</i>	<i>marthae</i>	AUM3958
Characiformes	Gasteropelecidae	<i>Thoracocharax</i>	<i>stellatus</i>	AUM 4002
Characiformes	Hemiodontidae	<i>Bivibranchia</i>	<i>fowleri</i>	AUM 3245
Characiformes	Hemiodontidae	<i>Hemiodus</i>	<i>argenteus</i>	AUM 3250
Characiformes	Hemiodontidae	<i>Hemiodus</i>	<i>thayeria</i>	AUM 3258
Characiformes	Hepsetidae	<i>Hepsetus</i>	<i>odoe</i>	AUM 5790
Characiformes	Iguanodectidae	<i>Bryconops</i>	<i>melanurus</i>	AUM 4794
Characiformes	Lebiasinidae	<i>Lebiasina</i>	<i>aureoguttata</i>	AUM 3719
Characiformes	Lebiasinidae	<i>Lebiasina</i>	<i>provenzano</i>	AUM 4823
Characiformes	Lebiasinidae	<i>Pyrrhulina</i>	<i>filamentosa</i>	AUM 3938
Characiformes	Parodontidae	<i>Apareiodon</i>	<i>orinocoensis</i>	AUM 3957
Characiformes	Parodontidae	<i>Parodon</i>	<i>guyanensis</i>	AUM 3605
Characiformes	Prochilodontidae	<i>Ichthyoelephas</i>	<i>longirostris</i>	ANSP 1305
Characiformes	Prochilodontidae	<i>Prochilodus</i>	<i>scrophus</i>	AUM 3735
Characiformes	Prochilodontidae	<i>Semaprochilodus</i>	<i>varii</i>	ANSP 187435
Characiformes	Serrasalmididae	<i>Acnodon</i>	<i>oligacanthus</i>	GO 401
Characiformes	Serrasalmididae	<i>Colossoma</i>	<i>macropomum</i>	GO 332
Characiformes	Serrasalmididae	<i>Myleus</i>	<i>pacu</i>	AUM 3042
Characiformes	Serrasalmididae	<i>Myloplus</i>	<i>rubripinnis</i>	AUM 3051
Characiformes	Serrasalmididae	<i>Mylossoma</i>	<i>sp</i>	AUM 3733
Characiformes	Serrasalmididae	<i>Ossubtus</i>	<i>sp</i>	GO 253
Characiformes	Serrasalmididae	<i>Pristobrycon</i>	<i>striolatus</i>	GO 400

Characiformes	Serrasalminidae	<i>Pygocentrus</i>	<i>nattereri</i>	AUM 3043
Characiformes	Serrasalminidae	<i>Serrasalmus</i>	<i>gouldingi</i>	AUM 3045
Characiformes	Serrasalminidae	<i>Serrasalmus</i>	<i>manueli</i>	AUM 3054
Characiformes	Serrasalminidae	<i>Serrasalmus</i>	<i>rhombeus</i>	AUM 3041
Characiformes	Triporthidae	<i>Triporthus</i>	<i>rotundatus</i>	AUM 3556
Cypriniformes	Cobitidae	<i>Acanthocobitis</i>	<i>sp</i>	AUM 3795
Cypriniformes	Cobitidae	<i>Acantopsis</i>	<i>sp</i>	AUM 5030
Cypriniformes	Cyprinidae	<i>Amblyrhynchichthys</i>	<i>micracanthus</i>	AUM 5035
Cypriniformes	Cyprinidae	<i>Incisilabeo</i>	<i>behri</i>	AUM 5025
Cypriniformes	Cyprinidae	<i>Barbonymus</i>	<i>gonionotus</i>	AUM 5095
Cypriniformes	Cyprinidae	<i>Opsarius</i>	<i>tileo</i>	AUM 3793
Cypriniformes	Cyprinidae	<i>Cabdio</i>	<i>morar</i>	AUM 3800
Cypriniformes	Cyprinidae	<i>Gibelion</i>	<i>catla</i>	AUM 5042
Cypriniformes	Cyprinidae	<i>Crossocheilus</i>	<i>latius</i>	AUM 3794
Cypriniformes	Cyprinidae	<i>Esomus</i>	<i>danrica</i>	AUM 3760
Cypriniformes	Cyprinidae	<i>Garra</i>	<i>fasciacauda</i>	AUM 5009
Cypriniformes	Cyprinidae	<i>Henicorhynchus</i>	<i>sp</i>	AUM 5059
Cypriniformes	Cyprinidae	<i>Labiobarbus</i>	<i>leptocheilus</i>	AUM 5099
Cypriniformes	Cyprinidae	<i>Macrhybopsis</i>	<i>storeriana</i>	AUM 7
Cypriniformes	Cyprinidae	<i>Osteochilus</i>	<i>lini</i>	AUM 5040
Cypriniformes	Cyprinidae	<i>Parachela</i>	<i>sp</i>	AUM 5066
Cypriniformes	Cyprinidae	<i>Probarbus</i>	<i>jullieni</i>	AUM 5023
Cypriniformes	Cyprinidae	<i>Rasbora</i>	<i>daniconius</i>	AUM 5111
Cypriniformes	Cyprinidae	<i>Salmostoma</i>	<i>phulo</i>	AUM 3779
Cypriniformes	Cyprinidae	<i>Sikukia</i>	<i>gudgeri</i>	AUM 5083
Cypriniformes	Paedocypridae	<i>Paedocypris</i>	<i>sp</i>	MT-001
Cypriniformes	Psilorhynchidae	<i>Psilorhynchus</i>	<i>sp</i>	AUM 3784
Cypriniformes	Cyprinidae	<i>Danio</i>	<i>rerio</i>	GO 23
Gymnotiformes	Apteronotidae	<i>Apteronotus</i>	<i>sp</i>	AUM 3732
Gymnotiformes	Apteronotidae	<i>Apteronotus</i>	<i>albifrons</i>	AUM 3921
Gymnotiformes	Apteronotidae	<i>Sternarchorhynchus</i>	<i>stewarti</i>	AUM 3745
Gymnotiformes	Gymnotidae	<i>Electrophorus</i>	<i>electricus</i>	AUM 3843
Gymnotiformes	Hypopomidae	<i>Brachyhypopomus</i>	<i>brevirostris</i>	AUM 3222
Gymnotiformes	Hypopomidae	<i>Brachyhypopomus</i>	<i>sp</i>	AUM 3920
Gymnotiformes	Hypopomidae	<i>Hypopygus</i>	<i>lepturus</i>	AUM 3224
Gymnotiformes	Rhamphichthyidae	<i>Gymnorhamphichthys</i>	<i>hypostomus</i>	AUM 3202
Gymnotiformes	Rhamphichthyidae	<i>Gymnorhamphichthys</i>	<i>rondoni</i>	AUM 3201
Gymnotiformes	Rhamphichthyidae	<i>Rhamphichthys</i>	<i>rostratus</i>	AUM 3220
Gymnotiformes	Sternopygidae	<i>Eigenmannia</i>	<i>virescens</i>	AUM 3218
Gymnotiformes	Sternopygidae	<i>Rhabdolichops</i>	<i>jegui</i>	ANSP 189021
Gymnotiformes	Sternopygidae	<i>Sternopygus</i>	<i>macrurus</i>	AUM 3198
Siluriformes	Aspredinidae	<i>Bunocephalus</i>	<i>amaurus</i>	AUM 3909
Siluriformes	Aspredinidae	<i>Ernstichthys</i>	<i>sp</i>	AUM 3988
Siluriformes	Aspredinidae	<i>Xyliphius</i>	<i>kryptos</i>	AUM 4054
Siluriformes	Astroblepidae	<i>Astroblepus</i>	<i>sp</i>	AUM 5348
Siluriformes	Auchenipteridae	<i>Auchenipterichthys</i>	<i>longimanus</i>	AUM 3321
Siluriformes	Auchenipteridae	<i>Centromochlus</i>	<i>heckelii</i>	AUM 3328
Siluriformes	Auchenipteridae	<i>Glanidium</i>	<i>leopardum</i>	AUM 4687
Siluriformes	Auchenipteridae	<i>Tatia</i>	<i>sp</i>	AUM 3320
Siluriformes	Auchenipteridae	<i>Trachelyichthys</i>	<i>sp</i>	AUM 3325
Siluriformes	Auchenipteridae	<i>Trachelyopterichthys</i>	<i>taeniatus</i>	AUM 3316
Siluriformes	Auchenipteridae	<i>Trachelyopterus</i>	<i>sp</i>	AUM 3330
Siluriformes	Bagridae	<i>Bagrus</i>	<i>ubangensis</i>	AUM 5668
Siluriformes	Bagridae	<i>Sperata</i>	<i>aor</i>	AUM 3818

Siluriformes	Callichthyidae	<i>Callichthys</i>	<i>callichthys</i>	ANSP 179110
Siluriformes	Cetopsidae	<i>Cetopsis</i>	<i>plumbea</i>	AUM 4035
Siluriformes	Cetopsidae	<i>Denticetopsis</i>	<i>macilenta</i>	AUM 3310
Siluriformes	Cetopsidae	<i>Helogenes</i>	<i>marmoratus</i>	AUM 3952
Siluriformes	Clariidae	<i>Channallabes</i>	<i>apus</i>	AUM 5737
Siluriformes	Claroteidae	<i>Chrysichthys</i>	<i>cranchii</i>	AUM 5676
Siluriformes	Claroteidae	<i>Parauchenoglanis</i>	<i>balayi</i>	AUM 5825
Siluriformes	Diplomystidae	<i>Diplomystes</i>	<i>nahuelbutaensis</i>	ANSP 180476
Siluriformes	Doradidae	<i>Acanthodoras</i>	<i>spinosissimus</i>	AUM 3267
Siluriformes	Doradidae	<i>Doras</i>	<i>micropoeus</i>	AUM 3262
Siluriformes	Doradidae	<i>Oxydoras</i>	<i>niger</i>	AUM 3301
Siluriformes	Doradidae	<i>Platydoras</i>	<i>costatus</i>	AUM 3268
Siluriformes	Doradidae	<i>Rhinodoras</i>	<i>armbrusteri</i>	AUM 3298
Siluriformes	Doradidae	<i>Rhynchodoras</i>	<i>woodsii</i>	AUM 4038
Siluriformes	Doradidae	<i>Scorpiodoras</i>	<i>sp</i>	AUM 3269
Siluriformes	Heptapteridae	<i>Brachyglanis</i>	<i>sp</i>	AUM 3179
Siluriformes	Heptapteridae	<i>Cetopsorhamdia</i>	<i>insidiosa</i>	AUM 3104
Siluriformes	Heptapteridae	<i>Chasmocranus</i>	<i>brevior</i>	AUM 4730
Siluriformes	Heptapteridae	<i>Imparfinis</i>	<i>hasemani</i>	AUM 3157
Siluriformes	Heptapteridae	<i>Leptorhamdia</i>	<i>marmorata</i>	AUM 4821
Siluriformes	Heptapteridae	<i>Mastiglanis</i>	<i>sp</i>	AUM 3122
Siluriformes	Heptapteridae	<i>Pimelodella</i>	<i>cristata</i>	AUM 3147
Siluriformes	Heptapteridae	<i>Pimelodella</i>	<i>leptosoma</i>	AUM 3106
Siluriformes	Heptapteridae	<i>Rhamdia</i>	<i>quelen</i>	AUM 3105
Siluriformes	Ictaluridae	<i>Ameiurus</i>	<i>nebulosus</i>	ANSP 185099
Siluriformes	Ictaluridae	<i>Noturus</i>	<i>leptacanthus</i>	AUM 37
Siluriformes	Loricariidae	<i>Acanthicus</i>	<i>adonis</i>	AUM 5787
Siluriformes	Loricariidae	<i>Ancistrus</i>	<i>leucostictus</i>	AUM 3936
Siluriformes	Loricariidae	<i>Chaetostoma</i>	<i>breve</i>	AUM 4063
Siluriformes	Loricariidae	<i>Farlowella</i>	<i>amazonum</i>	ANSP 199910
Siluriformes	Loricariidae	<i>Cteniloricaria</i>	<i>platystoma</i>	AUM 3894
Siluriformes	Loricariidae	<i>Hypostomus</i>	<i>taphorni</i>	AUM 4262
Siluriformes	Loricariidae	<i>Lamontichthys</i>	<i>filamentosus</i>	AUM 4024
Siluriformes	Loricariidae	<i>Lasiancistrus</i>	<i>schomburgkii</i>	AUM 4056
Siluriformes	Loricariidae	<i>Loricaria</i>	<i>sp</i>	AUM 3656
Siluriformes	Loricariidae	<i>Peckoltia</i>	<i>lineola</i>	AUM 3881
Siluriformes	Loricariidae	<i>Pseudancistrus</i>	<i>nigrescens</i>	AUM 3609
Siluriformes	Mochokidae	<i>Chiloglanis</i>	<i>occidentalis</i>	AUM 5622
Siluriformes	Mochokidae	<i>Synodontis</i>	<i>batesii</i>	AUM 5839
Siluriformes	Nematogenyidae	<i>Nematogenys</i>	<i>inermis</i>	ANSP 180477
Siluriformes	Pangasiidae	<i>Helicophagus</i>	<i>leptorhynchus</i>	AUM 5067
Siluriformes	Pangasiidae	<i>Pseudolais</i>	<i>pleurotaenia</i>	AUM 5052
Siluriformes	Pimelodidae	<i>Megalonema</i>	<i>platycephalum</i>	AUM 3103
Siluriformes	Pimelodidae	<i>Phractocephalus</i>	<i>hemioiopterus</i>	AUM 3118
Siluriformes	Pimelodidae	<i>Pimelodus</i>	<i>albofasciatus</i>	AUM 3100
Siluriformes	Pimelodidae	<i>Platysilurus</i>	<i>mucosus</i>	AUM 4019
Siluriformes	Pimelodidae	<i>Pseudoplatystoma</i>	<i>fasciatum</i>	AUM 3099
Siluriformes	Pimelodidae	<i>Sorubim</i>	<i>elongatus</i>	AUM 3107
Siluriformes	Plotosidae	<i>Plotosus</i>	<i>canius</i>	INHS 93712
Siluriformes	Pseudopimelodidae	<i>Batrochoglanis</i>	<i>villosus</i>	AUM 3169
Siluriformes	Pseudopimelodidae	<i>Cephalosilurus</i>	<i>sp</i>	AUM 3918
Siluriformes	Pseudopimelodidae	<i>Microglanis</i>	<i>poecilus</i>	AUM 3173
Siluriformes	Schilbeidae	<i>Clupisoma</i>	<i>garua</i>	AUM 3821
Siluriformes	Schilbeidae	<i>Eutropiichthys</i>	<i>murius</i>	AUM 3826

Siluriformes	Siluridae	<i>Kryptopterus</i>	<i>cryptopterus</i>	AUM 5050
Siluriformes	Siluridae	<i>Phalacronotus</i>	<i>bleekeri</i>	AUM 5058
Siluriformes	Siluridae	<i>Wallago</i>	<i>attu</i>	AUM 3766
Siluriformes	Sisoridae	<i>Gagata</i>	<i>cenia</i>	AUM 3780
Siluriformes	Sisoridae	<i>Gogangra</i>	<i>viridescens</i>	AUM 3788
Siluriformes	Trichomycteridae	<i>Haemomaster</i>	<i>venezuelae</i>	AUM 4223
Siluriformes	Trichomycteridae	<i>Henonemus</i>	<i>sp</i>	AUM 4034
Siluriformes	Trichomycteridae	<i>Ituglanis</i>	<i>amazonicus</i>	AUM 4205
Siluriformes	Trichomycteridae	<i>Paracanthopoma</i>	<i>parva</i>	AUM 4222
Siluriformes	Trichomycteridae	<i>Pseudostegophilus</i>	<i>nemurus</i>	AUM 4023
Siluriformes	Trichomycteridae	<i>Trichomycteridae</i>	<i>sp</i>	AUM 4217
Siluriformes	Trichomycteridae	<i>Vandellia</i>	<i>sp</i>	AUM 4228

- 468 AMNH: American Museum of Natural History, New York, USA  
 469 ANSP: Academy of Natural Sciences, Philadelphia, Pennsylvania, USA  
 470 AUM: Auburn University Natural History Museum, Auburn, Alabama, USA  
 471 GO: The George Washington University, Washington, DC, USA  
 472 INHS: Illinois Natural History Survey, University of Illinois, Champaign, Illinois, USA  
 473 MUSM: Museo de Historia Natural "Javier Prado" de la Universidad Nacional Mayor de San  
 474 Marcos, Lima, Peru  
 475 ROM: Royal Ontario Museum, Toronto, Canada  
 476 UFRGS: Universidade Federal do Rio Grande do Sul, Departamento de Zoologia, Brazil



505 **Table S6.** Description of species sampled for the pilot study and data processing.

Family	Species	Total number of reads	Reads map to reference	Loci map to reference	Single-copy loci (similarity > 97%)
Adrianichthyidae	<i>Oryzias latipes</i>	11,284,58	416,815	669	551
Amiidae	<i>Amia calva</i>	7,958,122	516,985	707	583
Aplocheilidae	<i>Pachypanchax sakaramyi</i>	3,208,756	191,962	589	465
Apteronotidae	<i>Apteronotus albifrons</i>	2,506,082	558,652	1090	913
Ariidae	<i>Brustarius solidus</i>	12,301,320	456,444	906	762
Carangidae	<i>Caranx hippos</i>	6,049,418	346,062	820	713
Chanidae	<i>Chanos chanos</i>	3,711,118	283,111	876	724
Characidae	<i>Astyanax mexicanus</i>	32,037,054	1,874,071	1486	1,270
Clupeidae	<i>Pellona flavipinnis</i>	14,744,898	868,083	955	797
Cyprinidae	<i>Danio rerio</i>	27,109,090	11,124,272	3947	3,667
Kneriidae	<i>Kneria sp.</i>	12,034,058	597,480	828	697
Osteoglossidae	<i>Osteoglossum bicirrhosum</i>	2,372,852	108,917	462	362
Rivulidae	<i>Nematolebias whitei</i>	10,553,528	443,216	641	526
Cyprinidae	<i>Tanichthys sp.</i>	18,603,974	1,936,994	3198	2,936

506 **Table S7.** Age of branch stems subtending each major clade in Otophysi (in MY).

Clade	This tree*	Chen et al. <sup>15</sup>	Nakatani et al. <sup>16</sup>
Gymnotiformes	61.0	65.2	39.0
Siluriformes	19.6	22.2	35.5
Characoidei	23.0	26.1	23.8
Citharinoidei	46.4	49.6	61.7

507 \*Otophysi tree grafted into the Fish Tree of Life (FToL)<sup>11,27</sup> and time-calibrated using TreePL  
508 and secondary calibrations (based on 61 primary fossil calibrations across all fishes; details in  
509 Betancur-R. et al.<sup>11,27</sup>). We did not use the original FToL<sup>11</sup> because *Distichodus*, the only  
510 representative examined of Citharinoidei in that study was later shown to be contaminated with  
511 Ctenoluciidae, a Characoidei (see<sup>15</sup>).

532 **Table S8.** Estimates of subtending branch lengths in coalescent time units for subclades in three  
 533 of the five datasets analyzed here. Stem ages are estimated from fossil-calibrated phylogenies (no  
 534 data available for metazoans and yeasts). Stem lengths in coalescent units shown in bold are  
 535 below the minimum theoretical values (5.3–8.3) required for the vast majority of genes in a  
 536 genome (99.99–99.999%) to achieve monophyly<sup>13</sup>. Note that although low values (or lack of  
 537 estimates) may be indicative of possible violations of the underlying assumption of subclade  
 538 monophyly, the modified GGI-based coalescent analyses that use a combination of  
 539 unconstrained and constrained gene trees obtain results identical to those using the constrained-  
 540 only version (Figs. 5, S6, S7).

Subclade	Stem length (in MY)	Generation time (years)	Effective population size (Ne)	Stem length (coalescent time units)
<b>Otophysi</b>				
Gymnotiformes	39.0-61.0 <sup>a</sup>	2.5-5.0 <sup>18</sup>	10000-100000 <sup>b</sup>	20-122
Siluriformes	19.6-35.5 <sup>a</sup>	2.5-5.0 <sup>18</sup>	10000-100000 <sup>b</sup>	20-71
Characoidei	23.0-26.1 <sup>a</sup>	2.5-5.0 <sup>18</sup>	10000-100000 <sup>b</sup>	23-52
Citharinoidei	46.4-61.7 <sup>a</sup>	2.5-5.0 <sup>18</sup>	10000-100000 <sup>b</sup>	46-123
<b>Eutherian mammals</b>				
Rodentia	9.2 <sup>28,29</sup>	0.2-0.3 <sup>30</sup>	400000-500000 <sup>31</sup>	30.6-57.2
Lagomorpha	26.2 <sup>28</sup>	0.2-0.3 <sup>30</sup>	400000-500000 <sup>31</sup>	87.3-163.7
Primates	1.8-5.2 <sup>28,29</sup>	8.0-15 <sup>30</sup>	10400-25200 <sup>31</sup>	<b>2.3</b> -31.2
<b>Neoavian birds</b>				
Coraciimorphae	0.88-0.99 <sup>29,32</sup>	1.0-12 <sup>33</sup>	10000-100000 <sup>b</sup>	<b>0.4</b> -49.5
Owl	12.9-14.5 <sup>29,32</sup>	1.0-12 <sup>33</sup>	10000-100000 <sup>b</sup>	5.3-725.0
Australaves	1.5-3.89 <sup>29,32</sup>	1.0-12 <sup>33</sup>	10000-100000 <sup>b</sup>	6.2-194.5
Accipitrimorphae	3.02-5.5 <sup>29,32</sup>	1.0-12 <sup>33</sup>	10000-100000 <sup>b</sup>	<b>1.25</b> -275

560

561 <sup>a</sup>averaged from values presented in Table S7.

562 <sup>b</sup>conservative range of values for freshwater fishes and birds.