

# Digest: Resolving phylogenomic conflicts in characiform fishes<sup>†</sup>

Santiago Montero-Mendieta<sup>1,2,\*</sup>  and Arjun Dheer<sup>3,\*</sup> 

<sup>1</sup>Conservation and Evolutionary Genetics Group, Doñana Biological Station (EBD-CSIC), Seville, Spain

<sup>2</sup>E-mail: [santiago.montero@ebd.csic.es](mailto:santiago.montero@ebd.csic.es)

<sup>3</sup>Department of Evolutionary Ecology, Leibniz Institute for Zoo and Wildlife Research (IZW), Berlin, Germany

Received December 3, 2018

Accepted December 12, 2018

How can taxonomists best resolve the challenge of curating and analyzing large phylogenomic datasets that produce incongruent but highly supported topologies? Betancur-R et al. used a recently established hypothesis-testing procedure on a large dataset of genes and species to study the evolutionary relationships of characiform fishes, finding that past conclusions of non-monophyly may have been problematic and establishing monophyly with high confidence. The new findings highlight the importance of using dense taxon sampling to resolve conflicting relationships with phylogenomic data.

Unraveling the Tree of Life remains a challenging issue. For decades, systematists have attempted to classify taxa based on limited resources such as morphological characters or few DNA markers (Nei and Kumar 2000). However, stochastic errors can arise from such small datasets and produce misleading phylogenetic hypotheses (Som 2014). In more recent research, the use of large amounts of genomic data for phylogenomic purposes can reduce these errors and contribute to resolving taxonomic issues, for example among major bird lineages (Jarvis et al. 2014) and neotropical frogs (Heinicke et al. 2018). Still, the resulting phylogenies can show considerably contrasting topologies depending on the different filtering strategies, the number of taxa included in each clade, and the accuracy of gene tree estimation (Herrando-Moraira et al. 2018).

In this issue, Betancur-R et al. (2018) applied a recently developed hypothesis-testing approach, Gene Genealogy Interrogation (GGI), to reconcile conflicting phylogenomic inferences and to estimate the effect of taxon sampling. The authors focused on fishes of the order Characiformes, a clade with a longstanding taxonomic controversy (see Chen et al. 2013) that has been

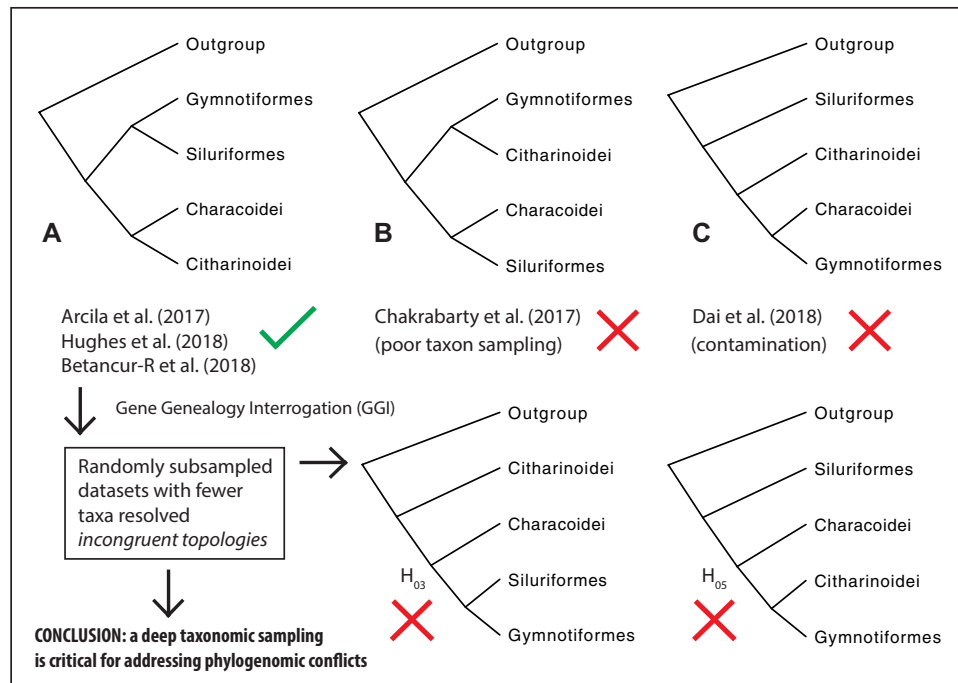
resolved as either monophyletic or non-monophyletic by recent phylogenomic studies (Fig. 1). An extended exon dataset with up to 1051 exons and 206 species (including 23 of 24 characiform families) was also assembled to infer the most comprehensive phylogeny for this group to date.

The study shows how top-ranking gene trees from GGI tests (GGI gene trees) can be used as input for summary coalescent analyses such as ASTRAL (Mirarab and Warnow 2015) to evaluate conflicting topologies. The monophyletic hypothesis was consistently confirmed across all the tests, including new tests based on an extended exonic dataset. The non-monophyletic hypothesis was only obtained when using randomly subsampled datasets with fewer taxa. The authors postulate that the non-monophyly from previous studies is a spurious result arising from limited taxonomic sampling in one study (Chakrabarty et al. 2017) and cross-contamination in another (Dai et al. 2018). The monophyly for Characiformes was also achieved when using randomly assembled subsets of gene trees. At least in this case, the use of denser taxon coverage appears to be more important than sequencing a large number of genes to resolve phylogenomic conflicts.

Betancur-R et al. (2018) also show that currently accepted methods for analyzing large datasets are outpaced by the sheer growth of data. Hence, presently applied techniques, when scaled up, struggle to handle phylogenomic analyses due to amplified biases and altered signal–noise ratios. Here, hypothesis-testing procedures applied on a gene-by-gene basis have proven useful

\*Both authors contributed equally to this work

<sup>†</sup>This article corresponds to Betancur-R., R., Arcila, D., Vari, R. P., Hughes, L. C., Oliveira, C., Sabaj, M. H., and Ortí, G. (2018) Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: the monophyly of characiform fishes. *Evolution*. <https://doi.org/10.1111/evo.13649>. <https://onlinelibrary.wiley.com/doi/10.1111/evo.13649>



**Figure 1.** Four recent phylogenomic studies (Arcila et al. 2017; Chakrabarty et al. 2017; Dai et al. 2018; Hughes et al. 2018) supported three different topologies (A–C) for the major otophysan lineages (Gymnotiformes, Siluriformes, Characoidei, and Citharinoidei). In this issue, by using Gene Genealogy Interrogation (GGI) in combination with a comprehensive exonic dataset, Betancur-R et al. (2018) resolved with high confidence the long-debated monophyly of the fish order Characiformes (including suborders Characoidei and Citharinoidei; Topology “A”). The GGI approach applied to randomly-assembled subsets of gene trees produced the correct monophyletic topology, while the same approach on subsampled datasets with fewer taxa frequently resolved incongruent, non-monophyletic topologies (H<sub>03</sub> and H<sub>05</sub>).

for clarifying evolutionary relationships with phylogenomic data.

Under particular circumstances, such as in the case of incomplete lineage sorting, some branches in the phylogeny can hold a large proportion of gene tree topologies that are incongruent with the underlying species tree (i.e., the so called “anomaly zone”; Degnan and Rosenberg 2006). In this situation, the GGI approach (GGI gene trees) in combination with species tree methods can simultaneously account for gene tree error and anomaly zone issues. Yet, it should remain critical for taxonomy to include evidence from multiple sources (not only molecular, but also ecological, behavioral, and morphological characters; Padial et al. 2010) in order to confidently resolve conflicting phylogenies. This study serves as a guideline for future projects aiming to resolve the evolutionary structure of intractable clades, and is a crucial step toward unraveling of the Tree of Life.

#### LITERATURE CITED

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

Betancur-R., R. D. Arcila, R. P. Vari, L. C. Hughes, C. Oliveira, M. H. Sabaj, and G. Ortí. 2018. Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: the monophyly of Characiform fishes. *Evolution* <https://doi.org/10.1111/evo.13649>.

Chakrabarty, P., B. C. Faircloth, F. Alda, W. B. Ludd, C. D. McMahan, T. J. Near, A. Dornburg, J. S. Albert, J. Arroyave, M. L. J. Stiassny et al. 2017. Phylogenomic systematics of Ostariophysan fishes: ultraconserved elements support the surprising non-monophyly of Characiformes. *Syst. Biol.* 66:881–895.

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

Dai, W., M. Zou, L. Yang, L. Du, W. Chen, Y. Shen, R. L. Mayden, and S. He. 2018. Phylogenomic perspective on the relationships and evolutionary history of the major Otocephalan lineages. *Sci. Rep.* 8:205.

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

Heinicke, M. P., A. R. Lemmon, E. M. Lemmon, K. McGrath, and S. B. Hedges. 2018. Phylogenomic support for evolutionary relationships of New World direct-developing frogs (Anura: Terraranae). *Mol. Phylogenet Evol.* 118:145–155.

Herrando-Moraira, S., The Cardueae Radiations Group. 2018. Exploring data processing strategies in NGS target enrichment to disentangle radiations in the tribe Cardueae (Compositae). *Mol. Phylogenet Evol.* 128:69–87.

Hughes, L. C., G. Ortí, Y. Huang, Y. Sun, C. C. Baldwin, A. W. Thompson, D. Arcila, R. Betancur-R, C. Li, L. Becker, et al. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on

- transcriptomic and genomic data. *Proc. Natl. Acad. Sci. USA* 115:6249–6254.
- Jarvis, E. D., S. Mirarab, A. J. Aberer, B. Li, P. Houde, . . . G., Zhang. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320-1331.
- Mirarab, S., and T. Warnow. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44-52.
- Nei, M., and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford, U.K.
- Padial, J. M., A. Miralles, I. De la Riva, and M. Vences. 2010. The integrative future of taxonomy. *Front. Zool.* 7:16.
- Som, A. 2014. Causes, consequences and solutions of phylogenetic incongruence. *Brief. Bioinform.* 16:536–548.

Associate Editor: K. Moore

Handling Editor: Mohamed A. F. Noor

#### **SUBMIT A DIGEST**

Digests are short (~500 word), news articles about selected original research included in the journal, written by students or postdocs. These digests are published online and linked to their corresponding original research articles. For instructions on Digests preparation and submission, please visit the following link: <https://sites.duke.edu/evodigests/>.