METHODS AND RESULTS FROM ““INTEGRATIVE PHYLOGENOMICS” INTEGRATES ERROR”.

Nathan V. Whelan[1,2*]

1: Museum of Biological Diversity, The Ohio State University, Columbus, OH USA.
2: Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH USA.

*Whelan.105@osu.edu

Methods

Dataset download and assembly

Full datasets from Steenwyk and King (SK) were downloaded (*1, 2*). The reduced datasets of SK (*1*) were generated using metrics reported by SK (*2*). SK did not report metrics for some datasets (*2*), and I did not attempt to recreate datasets when not enough information was provided by SK to do so. Single gene alignments for each dataset (i.e., full datasets and reduced datasets) were concatenated with FASconCAT-G (*3*).

I did not analyze the recoded datasets generated by SK for two reasons. First, SK did not fully report gene trees and other information for recoded datasets (*2*). This was apparently because IQ-TREE (*4*) returned errors when Steenwyk and King (*1*) first tried analyzing them (J. Steenwyk, *pers. comm.*). Given that my intention was to assess the suitability of SK’s methods for testing the root of the animal tree of life, I saw no reason to try to analyze datasets that SK failed to fully analyze and were therefore fundamentally flawed. Second, recoding amino acid datasets has not been shown to improve phylogenetic inference and likely results in less accurate phylogenetic reconstruction (*5*). Moreover, the RCV-reduced datasets are designed to reduce compositional heterogeneity without removing information content in each locus, so recoded analyses are, at best, redundant (*6, 7*).

Coalescent-based Species Tree Inference and Quartet Score Calculation

SK did not calculate species trees in their study (*1*). They instead used mostly unresolved trees that are unsuited for calculating quartet scores. Unresolved relationships, especially artificially unresolved relationships as used by SK, provide no information that can inform quartet scores, rendering the analysis of SK deficient. The approach of SK also differs from that of Shen et al. (*8*), which provided the supposed theoretical basis for what SK repurposed as “Integrative Phylogenomics” (*1*). Shen et al. (*8*) used fully resolved

species trees inferred with gene trees of each locus. I am aware of no theoretical basis for calculating quartet scores on trees where the only resolved relationships are the three branches that define the relationship being examined, in this case sponges or ctenophores as the sister lineage to all other animals.

Therefore, I calculated coalescent-based species trees with weighted-ASTRAL in ASTER (*9*), one of the most widely used and accurate species tree methods available. Species tree inference was only done on the full datasets as quartet scores for each gene tree are calculated independently from ASTRAL species trees inference. Input gene trees were those inferred by SK (*2*). The ASTRAL trees calculated with the Dunn et al. (*10*), Philippe et al. (*11*), and Nosenko et al. (*12*) datasets resulted in unrealistic relationships that have been widely rejected, including in the original studies (e.g., paraphyletic sponges). Given that the inferred ASTRAL trees were likely inaccurate, I did not examine those datasets further. I also did not examine any of the reduced datasets SK created from the original Ryan et al. (2013) dataset because they did not report the metrics they used for creating them (*1, 2*). Focusing on more recently generated datasets is also appropriate because more recent datasets have greater taxon sampling. Thus, they are better suited for testing hypotheses about metazoan evolution (*13, 14*). Alternative topologies for calculating quartet scores were generated in Mesquite (*15*) by swapping the ctenophore and sponge branch in Mesquite's tree editor tool.

For each full and reduced dataset, normalized quartet scores were calculated for each gene tree under the two alternative sister-lineage hypotheses and the score under the sponge-sister topology was subtracted from the score under the ctenophore-sister topology. This was done using code from Shen et al. (*8*) because no code was provided by SK.

Maximum likelihood tree inference and likelihood calculations

Concatenated maximum likelihood trees for full and reduced datasets were inferred with IQ-TREE 3 and 1,000 ultrafast bootstrap replicates (*16, 17*). I did not analyze datasets from Dunn et al. (*10*), Phillipe et al. (*11*), and Nosenko et al. (*12*) in a concatenated maximum likelihood framework because unrealistic species trees rendered comparisons between concatenation and summary coalescent methods problematic (see above). Once each concatenated tree was inferred, an alternative hypothesis tree was generated in Mesquite as above. Log-likelihood values for each site were calculated as in SK (*1*), except fully resolved maximum likelihood trees and the alternative tree for each dataset was used. This approach for the reference trees is much more robust than using unrealistic and almost completely unresolved trees as in SK. It is also the approach used by Shen et al. (*8*), which was the source of the method SK repurposed as "Integrative Phylogenomics" (*1*). Site-wise log-likelihoods were added together to determine whether each gene supported

the ctenophore-sister or sponge-sister hypothesis, as in SK (*1*). The exact code used for calculating log-likelihood values was not provided by SK, so I used the code from Shen et al. (*8*).

Identifying consistent genes

We first attempted to identify "consistent" genes (i.e., genes that supported the same topology in both maximum likelihood and summary coalescent frameworks) using the same criteria as SK (*1*). However, no gene in any dataset had an absolute normalized quartet score difference of greater than 0.1. Lowering the threshold to 0.02 produced a small minority of genes could be assessed for consistency between likelihood and summary coalescent methods for most, but not all datasets (Table 1). For datasets with no consistent genes at the 0.02 threshold, we used a quartet score threshold of 0.000001 (i.e., effectively no threshold). Significance for the number of genes supporting one hypothesis over another was assessed following SK (*1*).

Results

All maximum likelihood trees inferred with full and reduced datasets, including those with more complex models, recovered ctenophores as the sister lineage to all other Metazoa. Species tree inference with ASTRAL recovered ctenophores as the sister lineage to all other animals for every dataset except Dunn et al. (*10*), Philippe et al. (*11*), Nosenko et al. (*12*) and the two Whelan et al. (*18*) datasets. As stated above, species trees generated from these datasets, other than Whelan et al. (*18*), had relationships that have been widely rejected and were not analyzed further given dataset deficiencies for summary-coalescent methods. The Whelan et al. (*18*) species trees only differed from most species trees by placing sponges as the sister lineage to all other animals.

In every dataset, there were more consistent genes that supported the ctenophore-sister hypothesis than genes supporting the sponge-sister hypothesis, including for the Whelan et al. (*18*) datasets. The difference was significant for most datasets; datasets with non-significant differences had very few genes with absolute differences in normalized quartet scores above 0.02, at least partly because 0.02 is an arbitrary threshold and quartet scores can vary for no other reason than the number of taxa in a dataset. The threshold of 0.1 used by SK was equally arbitrary and problematic for a universal threshold (see above). When using a quartet score threshold of 0.000001, all datasets had significantly more consistent genes supporting ctenophores sister.

R code for the statistical analyses has been uploaded to GitHub (*19*). Inferred maximum likelihood and species trees in newick format have been deposited on GitHub. Normalized quartet scores and log-likelihood values for each locus can also be found on

GitHub . Metrics used to create reduced datasets, sequence alignments, and individual gene trees can be found in SK (*1, 2*).

References

1. J. L. Steenwyk, N. King, Integrative phylogenomics positions sponges at the root of the animal tree. *Science* **390**, 751–756 (2025).
2. J. L. Steenwyk, Data for: integrative phylogenomics positions sponges at the root of the animal tree. *FigShare*, (2025).
3. P. Kück, G. C. Longo, FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology* **11**, 81 (2014).
4. B. Q. Minh *et al.*, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
5. A. M. Hernandez, J. F. Ryan, Six-state amino acid recoding is not an effective strategy to offset compositional heterogeneity and saturation in phylogenetic analyses. *Syst. Biol.* **70**, 1200–1212 (2021).
6. M. J. Phillips, D. Penny, The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogen. Evol.* **28**, 171–185 (2003).
7. J. L. Steenwyk *et al.*, PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* **37**, 2325–2331 (2021).
8. X.-X. Shen, J. L. Steenwyk, A. Rokas, Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data. *Syst. Biol.* **70**, 997–1014 (2021).
9. C. Zhang, R. Nielsen, S. Mirarab, ASTER: A Package for Large-Scale Phylogenomic Reconstructions. *Mol. Biol. Evol.* **42**, msaf172 (2025).
10. C. W. Dunn *et al.*, Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
11. H. Philippe *et al.*, Phylogenomics Revives Traditional Views on Deep Animal Relationships. *Curr. Biol.* **19**, 706–712 (2009).
12. T. Nosenko *et al.*, Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogen. Evol.* **67**, 223–233 (2013).
13. T. A. Heath, S. M. Hedtke, D. M. Hillis, Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* **46**, 239–257 (2008).
14. T. A. Heath, D. J. Zwickl, J. Kim, D. M. Hillis, Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst. Biol.* **57**, 160–166 (2008).
15. W. P. Maddison, D. R. Maddison. (http://mesquiteproject.org, 2017).
16. T. K. F. Wong *et al.*, IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. *EcoEvoRxiv*, (2025).
17. D. T. Hoang, O. Chernomor, A. Von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

18. N. V. Whelan, K. M. Kocot, L. L. Moroz, K. M. Halanych, Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5773–2778 (2015).
19. N. V. Whelan, Data from "Integrative Phylogenomics" integrates errors, GitHub, (2025); https://doi.org/10.5281/zenodo.17883124