

Supplementary Material: A Maximum-Likelihood Method to Construct Phylogenetic Trees Using Low-Homoplasy Markers: New Insights into Neoaves Phylogeny

Liang Gao¹ and Stefan Grünewald¹

¹Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, 200031, Shanghai, China

Correspondence to be sent to: stefan@picb.ac.cn

This supplementary material contains a series of tables and figures that support and extend the findings presented in the main text.

- Table S1 summarizes the additional number of markers required to reconstruct specific clades, along with the bird taxa included in each corresponding cluster.
- Figure S1 shows the true tree used in the simulation experiments, serving as a reference for evaluating the performance of different phylogenetic reconstruction methods.
- Figure S2 uses the dataset from [5] and compares the reconstruction results obtained by two different methods.
- Figure S3 presents simulation results designed to test the robustness of our approach under controlled conditions.
- Figures S4–S11 display branch lengths inferred by MLRE on several previously published Neoaves phylogenies, allowing direct comparison with existing topologies.
- Figure S12 shows an ASTRAL-BP tree with outgroup taxa included, highlighting the impact of rooting on tree topology.
- Figure S13 illustrates branch lengths inferred by MLRE on the tree constructed using the Dolpenny method [1].
- Figure S14 represents the number of congruent markers supporting each branch of the MLRE tree.
- Figure S15 shows a simplified Neoaves tree containing only 13 taxa.

Together, these figures offer additional validation, comparison, and context for the phylogenetic inferences discussed in the main manuscript.

References

- [1] J. Felsenstein. Phylip-phylogeny inference package (version 3.2). *Cladistics*, 5:164–166, 1989.
- [2] J. Gatesy and M. S. Springer. Phylogenomic coalescent analyses of avian retroelements infer zero-length branches at the base of neoaves, emergent support for controversial clades, and ancient introgressive hybridization in afroaves. *Genes*, 13(7):1167, 2022.
- [3] E. D. Jarvis, S. Mirarab, A. J. Aberer, B. Li, P. Houde, C. Li, S. Y. Ho, B. C. Faircloth, B. Nabholz, J. T. Howard, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346(6215):1320–1331, 2014.
- [4] H. Kuhl, C. Frankl-Vilches, A. Bakker, G. Mayr, G. Nikolaus, S. Boerno, S. Klages, B. Timmermann, and M. Gahr. An unbiased molecular approach using 3'-utrs resolves the avian family-level tree of life. *Molecular Biology and Evolution*, 38:108 – 127, 2020.

- [5] E. K. Molloy, J. Gatesy, and M. S. Springer. Theoretical and practical considerations when using retroelement insertions to estimate species trees in the anomaly zone. *Systematic Biology*, 71(3):721–740, 2022.
- [6] R. O. Prum, J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon, and A. R. Lemmon. A comprehensive phylogeny of birds (aves) using targeted next-generation dna sequencing. *Nature*, 526(7574): 569–573, 2015.
- [7] S. Reddy, R. T. Kimball, A. Pandey, P. A. Hosner, M. J. Braun, S. J. Hackett, K.-L. Han, J. Harshman, C. J. Huddleston, S. Kingston, et al. Why do phylogenomic data sets yield conflicting trees? data type influences the avian tree of life more than taxon sampling. *Systematic biology*, 66(5):857–879, 2017.
- [8] J. Stiller, S. Feng, A.-A. Chowdhury, I. Rivas-González, D. A. Duchêne, Q. Fang, Y. Deng, A. M. Kozlov, A. Stamatakis, S. Claramunt, et al. Complexity of avian evolution revealed by family-level genomes. *Nature*, 629:851 – 860, 2024.
- [9] A. Suh, L. Smeds, and H. Ellegren. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS biology*, 13(8):e1002224, 2015.
- [10] S. Wu, F. E. Rheindt, J. Zhang, J. Wang, L. Zhang, C. Quan, Z. Li, M. Wang, F. Wu, Y. Qu, S. V. Edwards, Z. Zhou, and L. Liu. Genomes, fossils, and the concurrent rise of modern birds and flowering plants in the late cretaceous. *Proc Natl Acad Sci*, 121, 2024.
- [11] G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8, 2017.

Table S1. Number of markers added

Cluster	Taxa in Cluster	flexible c	constant c
Otididae [3]	Otidimorphae+Strisores	3	3
Otidimorphae [3]	cuckoos, bustards, turacos	0	1
Columbea [3]	Columbimorphae+Mirandornithes	8	4
Columbimorphae [3]	pigeons, mesites, sandgrouse	2	2
Gruae [3]	Opisthocomiformes+Gruimorphae	3	4
Phaethoquornithes [3]	Eurypygimorphae+Aequornithes	1	1
Inopinaves [6]	Opisthocomiformes+Telluraves	6	3
Gruimorphae [8]	Gruiformes+Charadriiformes	3	3
Columbaves [6]	Columbimorphae+Otidimorphae	9	9
Litusilvanae [10]	Strisores+Gruimorphae	3	5
Elementaves [8]	Gruae+Strisores+Phaethoquornithes	4	>10
Aequorlitorntithes [10]	Mirandornithes+Opisthocomiformes+Phaethoquornithes	>10	>10

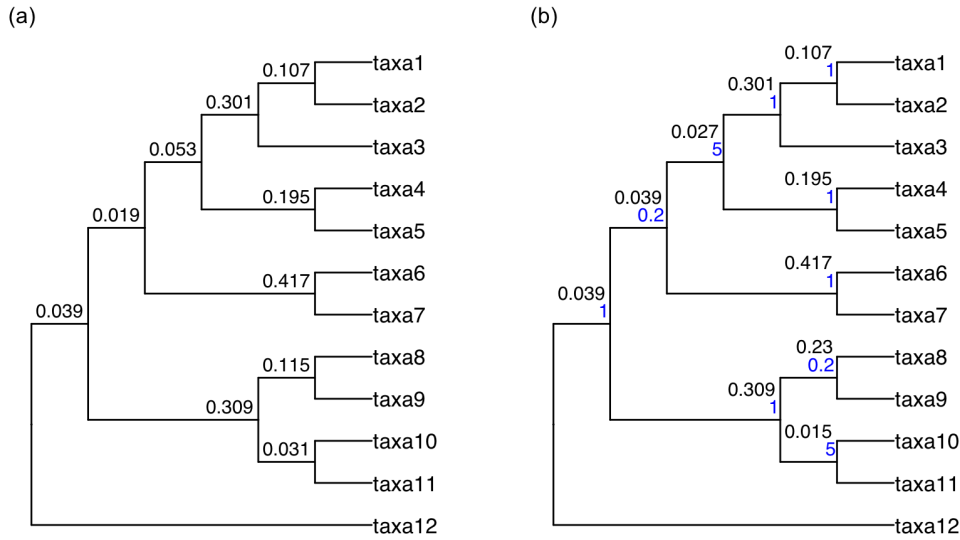


Fig. S1. Simulated tree from adjusted palaeognathae. (a) The tree used for simulations under the original model, (b) The tree used for simulations with variable insertion rates. Black numbers indicate the branch lengths, and blue numbers indicate insertion rates. The phylogenetic trees in this supplementary material are all plotted using the R package ggtree [11].

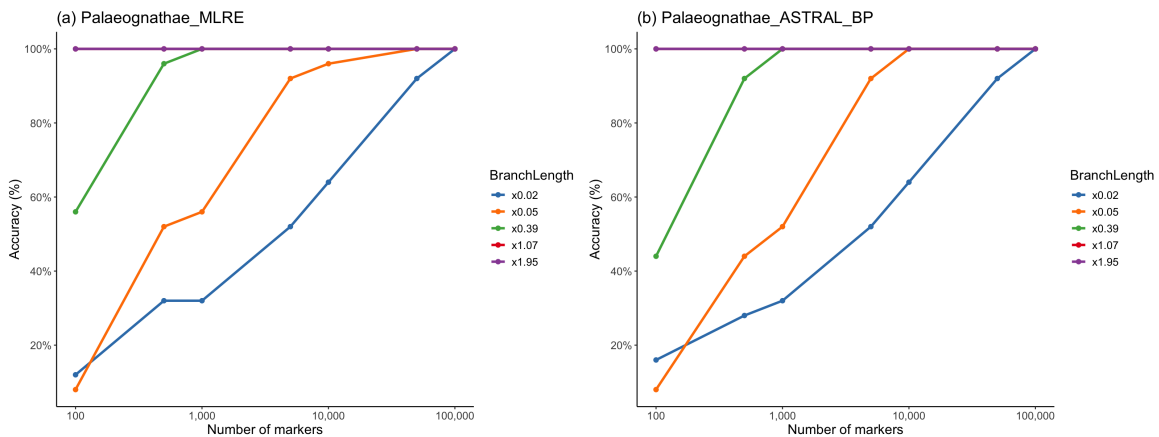


Fig. S2. Performance comparison on simulated data [5] generated by ms. (a) Results from the MLRE method, (b) Results by ASTRAL_BP. Different colors represent different branch lengths. The x-axis indicates the data size, and the y-axis represents the percentage of correctly recovered branches.

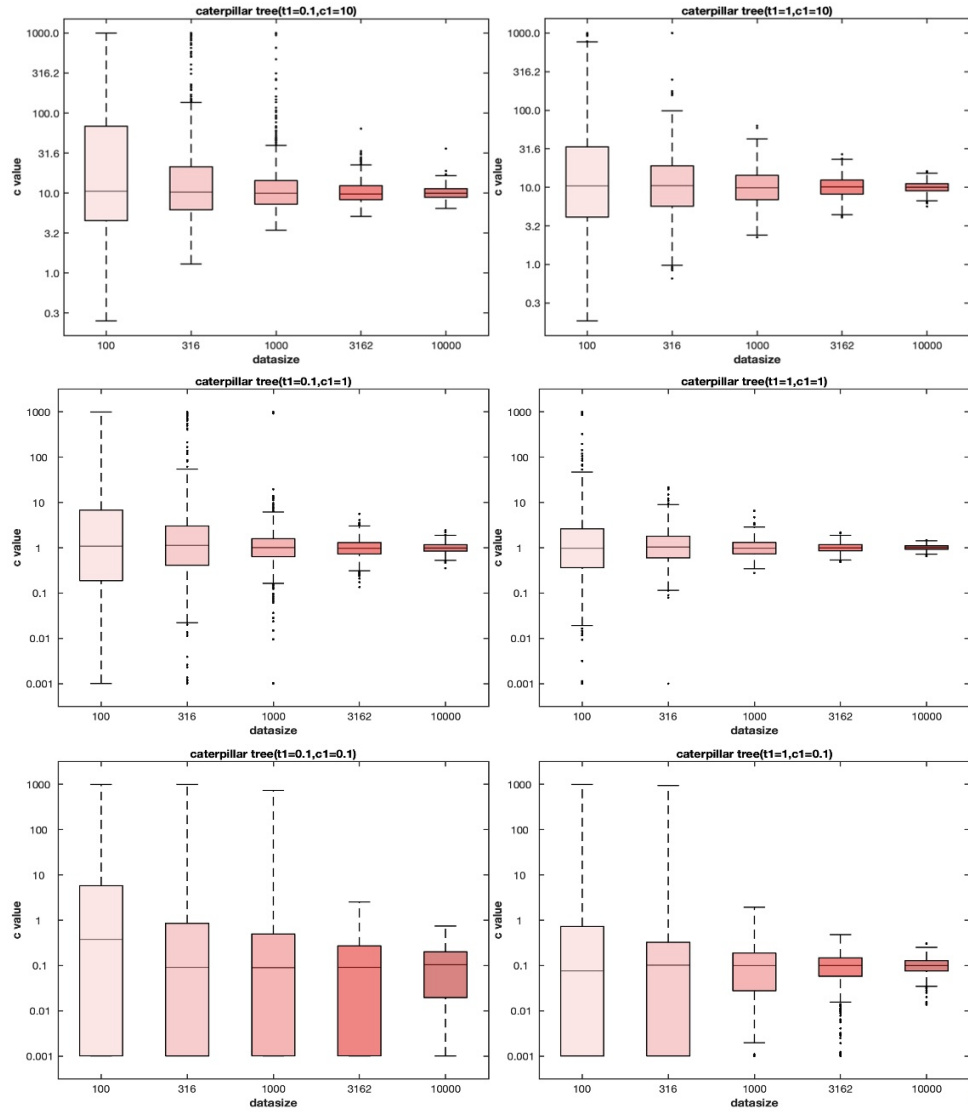


Fig. S3. Distribution of estimates of insertion rate c for different data sizes. Left and right represent simulated trees with branch lengths of 0.1 and 1, respectively. Top, middle, and bottom represent simulated values of 10, 1, and 0.1, respectively, for the insertion rate c . The x-axis is the number of markers ranging from 100, 316, 1,000, and 3,162 to 10,000; and the y-axis is the distribution of the estimated value of the insertion rate c , which is constrained to the range between (0.001, 1,000). Note that the y-axis is scaled differently.

Stiller et al, logL = -9986.852414461573

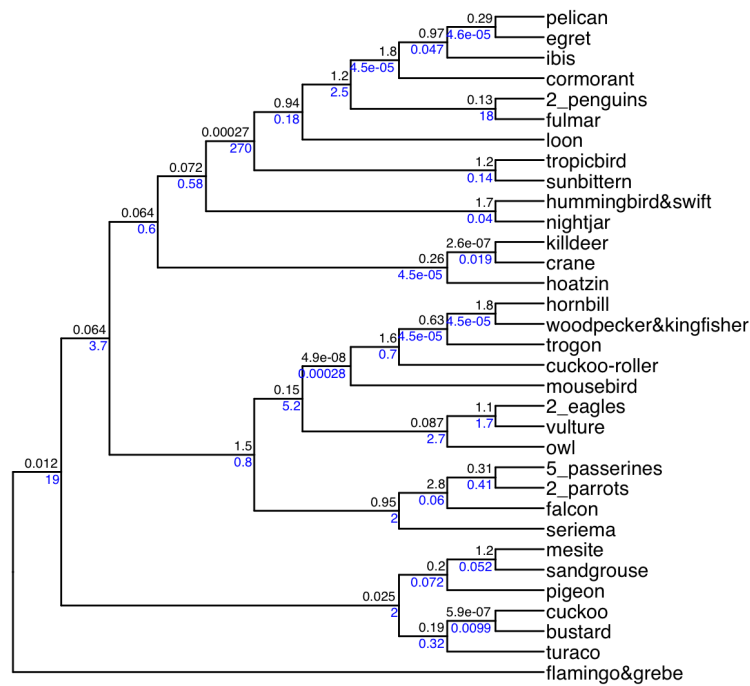


Fig. S4. The tree topology is from [8], with branch lengths inferred using the MLRE method.

Wu et al, logL = -10068.39202245297

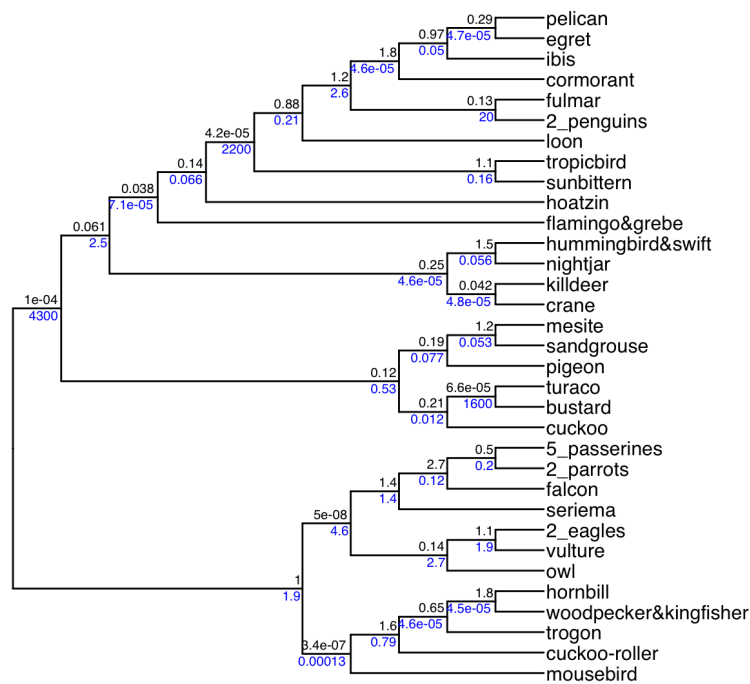
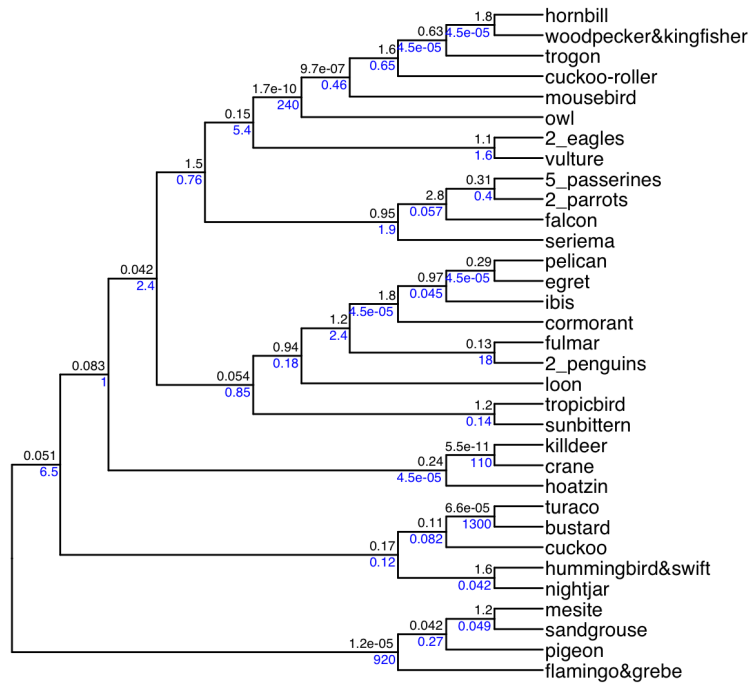
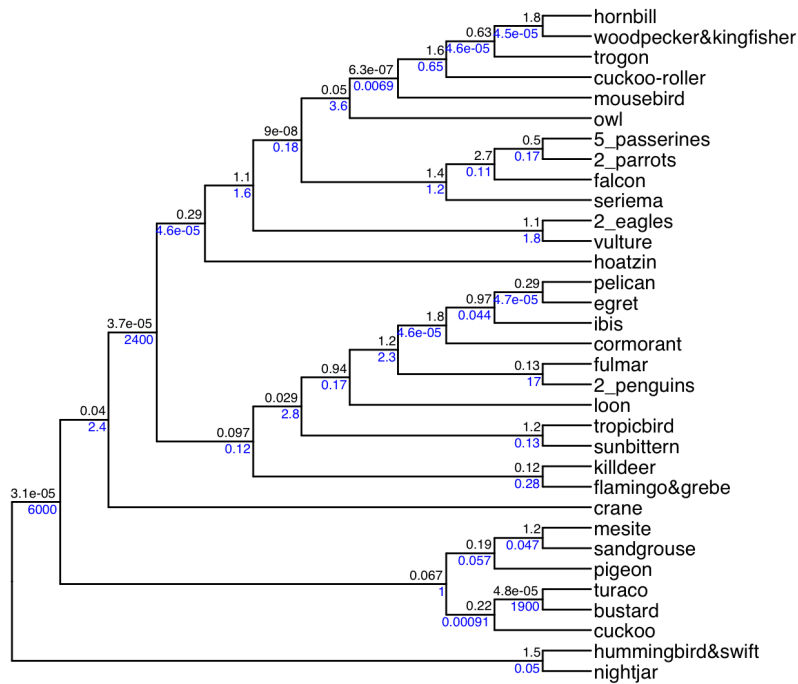


Fig. S5. The tree topology is from [10], with branch lengths inferred using the MLRE method.

Jarvis et al, logL = -9971.5799796794

**Fig. S6.** The tree topology is from [3], with branch lengths inferred using the MLRE method.

Prum et al, logL = -10105.12198879041

**Fig. S7.** The tree topology is from [6], with branch lengths inferred using the MLRE method.

Reddy et al, logL = -10016.291319967757

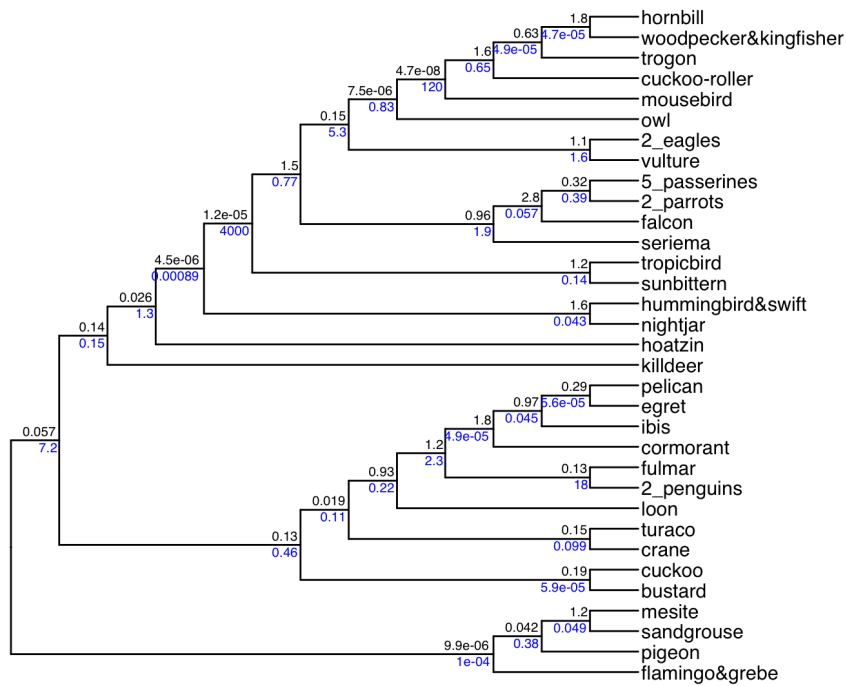


Fig. S8. The tree topology is from [7], with branch lengths inferred using the MLRE method.

Kuhl et al, logL = -10019.763542383553

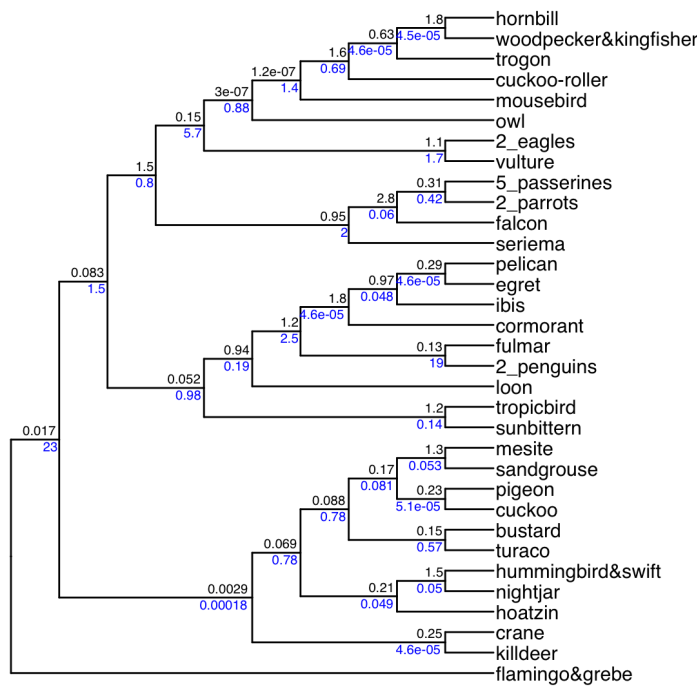
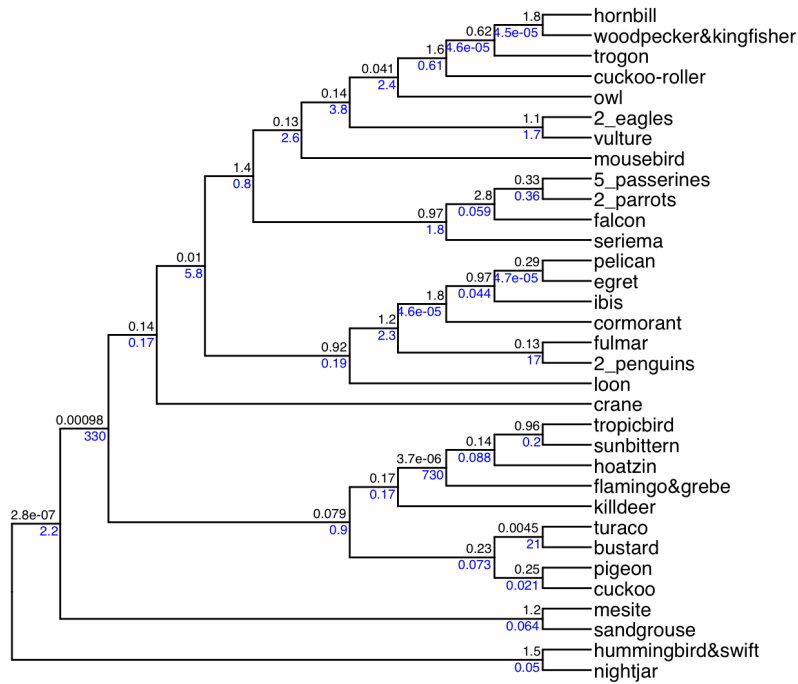
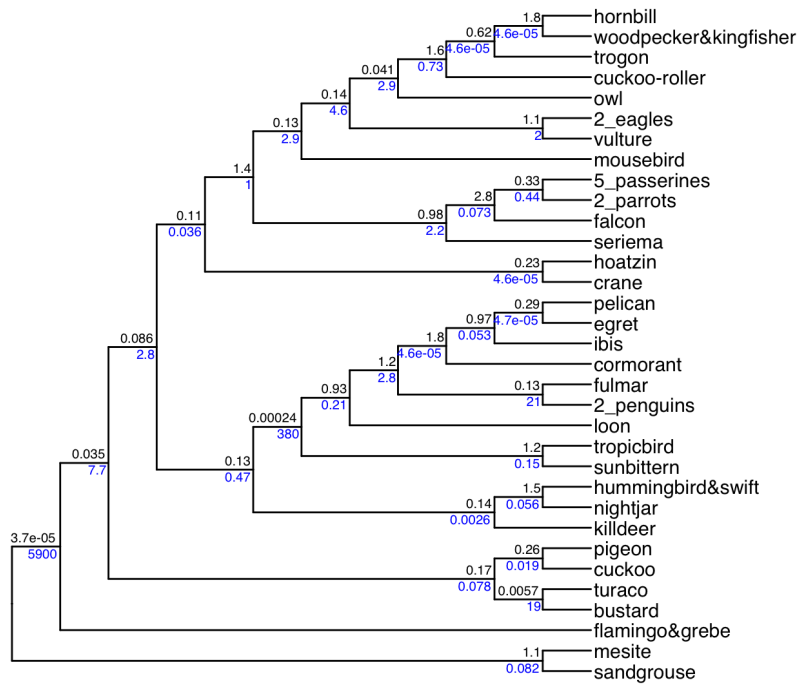


Fig. S9. The tree topology is from [4], with branch lengths inferred using the MLRE method.

Gatesy et al, logL = -9979.90883974701

**Fig. S10.** The tree topology is from [2], with branch lengths inferred using the MLRE method.

MPRE, logL = -9940.254815261942

**Fig. S11.** The tree topology is from [9], with branch lengths inferred using the MLRE method.

adjusted ASTRAL, logL = -9947.135874193296

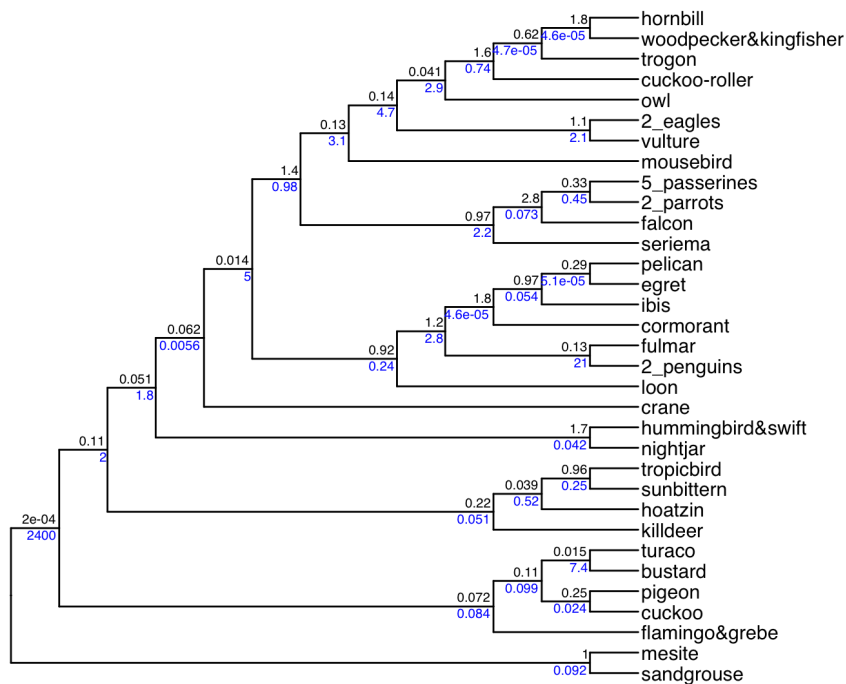


Fig. S12. ASTRAL_BP adjusted. To root the tree, an all-zero outgroup was added to the dataset of 48 Neoaves taxa, and the phylogeny was inferred using the ASTRAL_BP method.

Dolpenny, logL = -9936.345085416004

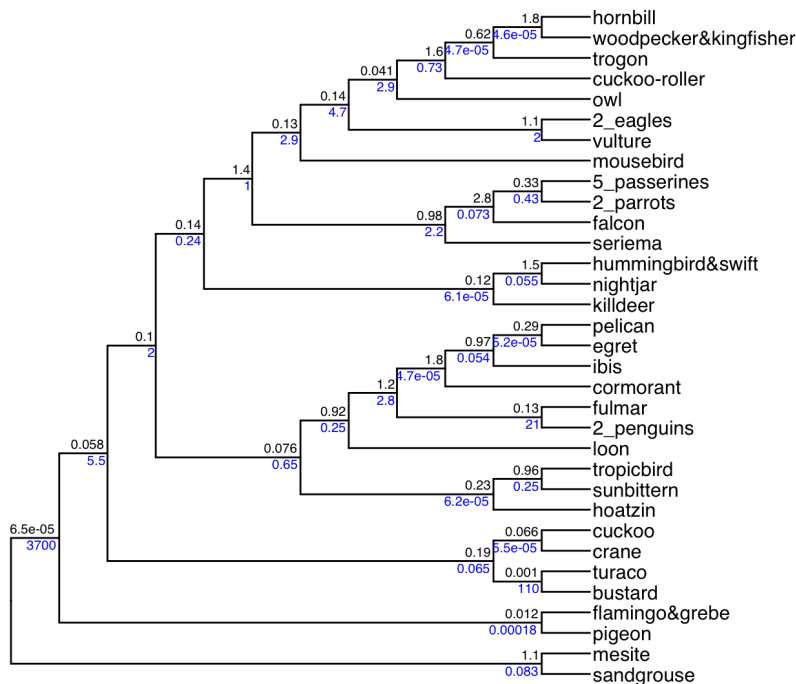


Fig. S13. The tree is generated from dolpenny method [1], with branch lengths inferred using the MLRE method.

Support markers in MLRE, total =1404

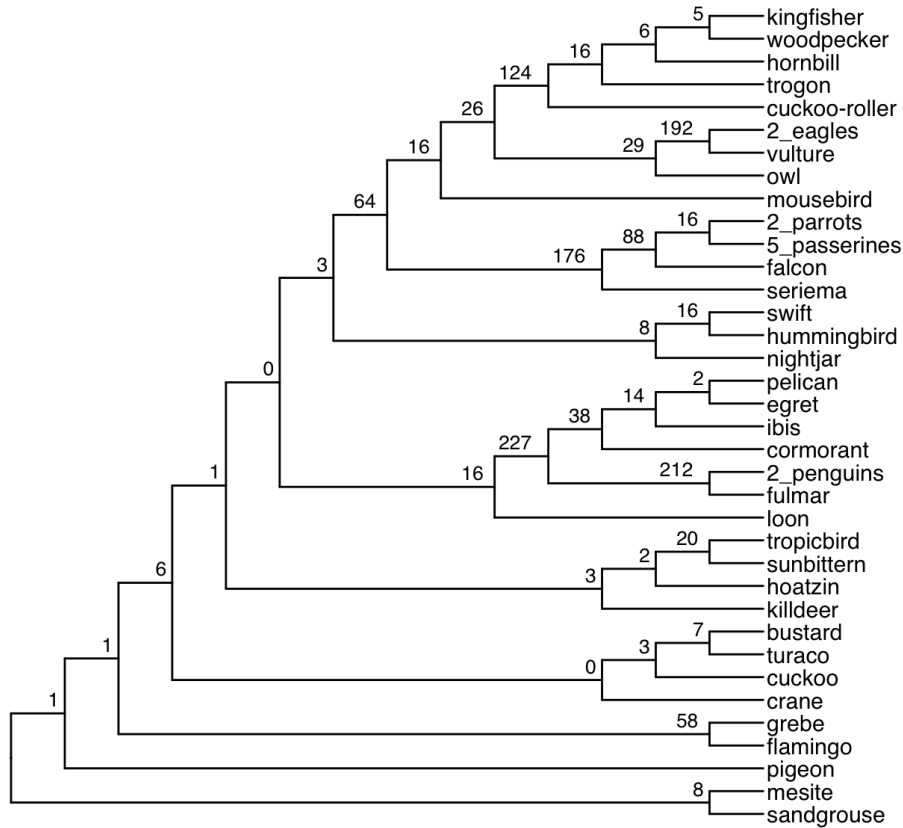


Fig. S14. Congruent markers with the MLRE tree. For each branch of the MLRE tree, we calculated the number of congruent markers. There are 1404 such markers compare to 1398 markers for MPRE method, MLRE consistently results in a higher number of congruent markers.

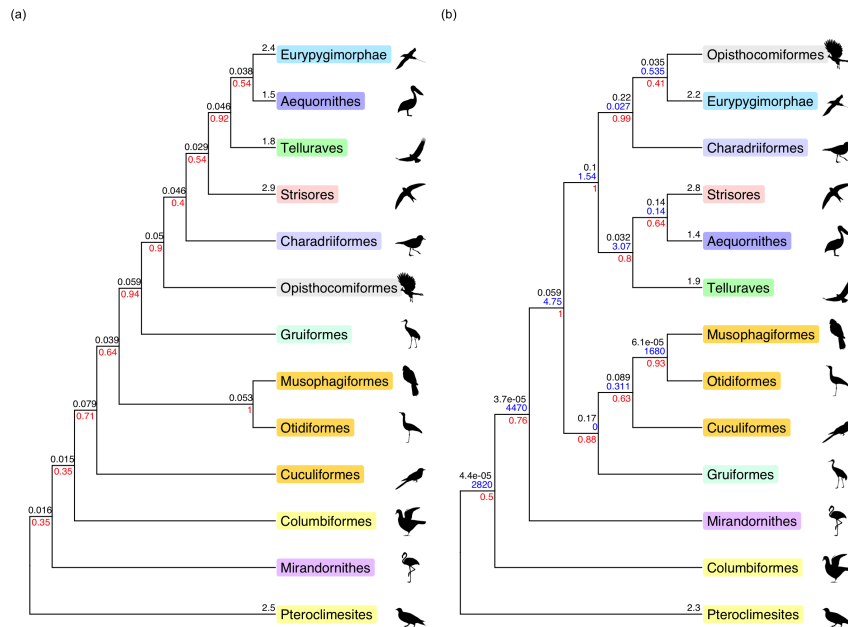


Fig. S15. MLRE tree for 13 Neoaves taxa. (a) The Neoaves tree under the MLRE(c=1), logL = -3142.83593; (b) The Neoaves tree with variable insertion rates, logL = -3114.30806. Black numbers indicate branch length, blue numbers indicate insertion rates, and red numbers indicate branch support. In particular, only the evolutionary branches corresponding to the group-taxa have finite leaf edge length.