Molecular Biology and Evolution, 2026, 43, 1–6
https://doi.org/10.1093/molbev/msag117
Published: 5 May 2026
Resources

MBE OXFORD

IQ-TREE 3: phylogenomic inference software using complex evolutionary models

Thomas K.F. Wong [1,2†], Nhan Ly-Trong [1†], Huaiyan Ren [1], Piyumal Demotte [1], Hector Baños [3,4], Andrew J. Roger [4], Edward Susko [5], Chris Bielow [6], Nicola De Maio [7], Nick Goldman [7], Matthew W. Hahn [8], Mario dos Reis [9], Le Sy Vinh [10], Gavin Huttley [2], Robert Lanfear [2]*, Bui Quang Minh [1]*

[1]School of Computing, College of Systems and Society, Australian National University, Canberra, ACT, Australia
[2]Research School of Biology, College of Science and Medicine, Australian National University, Canberra, ACT, Australia
[3]Department of Mathematics, California State University San Bernardino, San Bernardino, CA, USA
[4]Department of Biochemistry and Molecular Biology, Faculty of Medicine, Dalhousie University, Halifax, NS, Canada
[5]Department of Mathematics and Statistics, Faculty of Science, Dalhousie University, Halifax, NS, Canada
[6]Bioinformatics Solution Center, Freie Universität Berlin, Berlin 14195, Germany
[7]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridgeshire, UK
[8]Department of Biology and Department of Computer Science, Indiana University, Bloomington, IN 47405, USA
[9]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4NS, UK
[10]University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
***Corresponding author:** Email: m.bui@anu.edu.au; rob.lanfear@anu.edu.au
[†]Co-first authors.
Associate editor: Koichiro Tamura

Abstract

IQ-TREE (https://iqtree.github.io/) is a widely used open-source software tool for efficiently inferring phylogenetic trees under maximum likelihood. Here, we present IQ-TREE version 3, the third major release of the software. IQ-TREE 3 significantly extends version 2 with new features, including mixture models as an alternative to partitioned models, gene and site concordance factors to quantify discordance between genomic regions, integration with phylogenomic divergence time estimation, and a fully featured sequence simulator. The IQ-TREE 3 source code is available at https://github.com/iqtree/iqtree3.

Keywords phylogenetic software, phylogenomic software, maximum likelihood, mixture model, concordance factor, sequence simulator

Introduction

IQ-TREE (Nguyen et al. 2015; Minh et al. 2020b) is a popular open-source software package for phylogenetic and phylogenomic analysis using a likelihood framework. IQ-TREE has been used by thousands of researchers and practitioners to understand evolution and biodiversity across many species and time scales. These studies range from inferring ancient relationships within Bacteria (Parks et al. 2018; Wehbi et al. 2024), Archaea (Alves et al. 2018; Dombrowski et al. 2020), and eukaryotes (Burki et al. 2016; Williamson et al. 2025); to more recent divergences among plants (Puttick et al. 2018), animals (Redmond and McLysaght 2021), birds (Stiller et al. 2024), gum trees (Crisp et al. 2024), and primates (Vanderpool et al. 2020); and even viral divergences such as SARS-CoV-2 virus (Rambaut et al. 2020; Holmes et al. 2021). The two major releases of IQ-TREE (Nguyen et al. 2015; Minh et al. 2020b) have been cited more than 44,000 times (Google Scholar; 17th March 2026) and downloaded more than 1.3 million times (combined data from the GitHub repository and Bioconda package). IQ-TREE is also an integral component of many open-source biomedical applications such as CIPRES (Miller et al. 2010), GTDB (Parks et al. 2022), QIIME2 (Bolyen et al. 2019), OrthoFinder (Emms and Kelly 2019), Nextstrain (Hadfield et al. 2018), Galaxy (Galaxy Community 2024), PhyloSuite (Zhang et al. 2020; Xiang et al. 2023), TBtools (Chen et al. 2023), and Get_PhyloMarkers (Vinuesa et al. 2018).

Since the release of IQ-TREE version 2 in 2020, we have substantially expanded its capabilities, allowing users to perform data analyses under additional models and scenarios. Here, we introduce IQ-TREE version 3 and highlight its key developments, including mixture models, concordance factors, sequence simulation, protein model estimation, divergence time estimation, and pathogen tree inference.

Received: April 3, 2025. **Revised:** April 2, 2026. **Accepted:** April 23, 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026

Mixture models

Phylogenetic analyses are often conducted with partitioned models, which assign different substitution models to different a priori defined groups of sites to account for heterogeneous evolutionary processes along genomes and/or within genes (e.g. among codon positions). Partitioned models are available in RAxML (Stamatakis 2014), RAxML-NG (Kozlov et al. 2019), and IQ-TREE 2 (Minh et al. 2020b); and tools like PartitionFinder (Lanfear et al. 2017) can merge partitions to reduce over-parameterization.

IQ-TREE 2 implemented mixture models (Le et al. 2008), which assume a user-defined number (k) of substitution models (or classes), like partition models. However, unlike partition models, users do not need to specify how to partition the data in mixture models. Rather, model-based probabilities of sites are conditional upon classes, and each site has a certain probability (or weight) of belonging to each class. The site likelihood is calculated as the weighted average of the conditional probabilities, giving the marginal probability of the site pattern. Models of rate heterogeneity among sites based on a Gamma distribution (Yang 1994) or free-rates (Soubrier et al. 2012; Kalyaanamoorthy et al. 2017) are versions of mixture models. Mixture models in IQ-TREE 2 have been used to address many deep phylogenetic questions (e.g. Redmond and McLysaght 2021; Williamson et al. 2025). In particular, such models help to avoid artefacts such as long-branch attraction (Felsenstein 1978; Wang et al. 2018; Leuchtenberger et al. 2020), which can be exacerbated with simpler models.

IQ-TREE 3 substantially improves the IQ-TREE 2 implementation by automatically determining the number of classes (k) and the substitution models for each class via the MixtureFinder algorithm (Ren et al. 2025) for DNA data. In this way, the usage of DNA mixture models is now even simpler than partitioned models because no partition file is needed. This also removes user-subjectivity in determining the initial partitioning scheme, although it does come at the cost of additional computational resources: the amount of time and memory required by mixture models increases linearly with k. For protein data, in addition to profile mixtures (Wang et al. 2008) and structure-based mixture models (Le et al. 2008) available in version 2, IQ-TREE 3 allows users to estimate amino-acid exchangeability matrices under complex profile mixture models, the so-called GTRpmix model (Baños et al. 2024). Notably, IQ-TREE 3 can also relax the assumption of a single tree common to many phylogenetic analyses by using the Mixtures Across Sites and Trees (MAST) model (Wong et al. 2024), which generalizes the General Heterogeneous evolution On a Single Topology (GHOST) mixture model of branch lengths (Crotty et al. 2020). The MAST model can be useful to examine controversies between competing phylogenetic hypotheses and to pinpoint which sites in an alignment are best explained by which topologies; currently, it requires a user-provided set of input tree topologies.

Concordance factors

While the MAST model can summarize the contributions of a limited set of alternative tree topologies, IQ-TREE 3 also allows users to summarize discordance on all branches of a reference tree (e.g. an estimate of the species tree). We previously introduced methods to calculate two measures, called the gene concordance factor (gCF) and the site concordance factor (sCF) (Minh et al. 2020a), which quantify the level of concordance between each branch of the reference tree and the genes/loci (for the gCF) and sites (for the sCF). As the original sCF is based on parsimony and can be affected by homoplasy (similarities due to processes other than common ancestry) and taxon sampling, we introduced a likelihood-based version of the sCF (Mo et al. 2023) that partially overcomes these issues. All of these methods and related summary statistics are available in IQ-TREE 3.

Sequence simulations

Sequence simulations are used in many scenarios, such as evaluating new methods, testing hypotheses with parametric bootstrap (Giacomelli et al. 2025), performing approximate Bayesian computation (Csilléry et al. 2010), and generating training and testing data to help with the development of machine learning applications (Smith and Hahn 2023, but see Trost et al. 2024). However, existing tools like Seq-Gen (Rambaut and Grassly 1997), Dawg (Cartwright 2005), and INDELible (Fletcher and Yang 2009) lack support for mixture and other complex models. IQ-TREE 3 implements a built-in alignment simulator called AliSim (Ly-Trong et al. 2022) to take advantage of all advanced models available in IQ-TREE. To optimize speed, AliSim provides two different algorithms, one to simulate sequences of low divergence and another for high divergence, and can adaptively determine which algorithm to use. AliSim substantially outperforms Seq-Gen, Dawg, and INDELible in terms of both runtimes and memory usage. We also provided a high-performance version, AliSim-HPC (Ly-Trong et al. 2023), which can parallelize the simulations over multicore CPUs and multinode clusters.

New protein models

Protein data have been used to address many phylogenetic questions, especially at deep time scales. In many cases, analyses rely on empirical protein models like JTT (Jones et al. 1992), WAG (Whelan and Goldman 2001), and LG (Le and Gascuel 2008), which were estimated from large collections of alignments spanning many clades on the tree of life. The scale of modern phylogenomic datasets and powerful computers present an opportunity to estimate models tailored specifically to individual clades. To facilitate this, IQ-TREE 3 provides a new maximum likelihood tool called QMaker (Minh et al. 2021) to estimate a protein substitution matrix for a user-provided set of protein alignments and introduces a general matrix Q.pfam and a series of new clade-specific matrices—Q.bird, Q.insect, Q.mammal, Q.plant, Q.yeast—derived from large empirical datasets (Jarvis et al. 2014; Misof et al. 2014; Ran et al. 2018; Shen et al. 2018; Wu et al. 2018; El-Gebali et al. 2019). These matrices will be useful for phylogenetic analyses of these and related clades. IQ-TREE 3 also provides an extension of QMaker for nonreversible models (Dang et al. 2022) and a series of corresponding nonreversible Q matrices, as well as two new matrices, EAL (Eukaryotes and Archaea) and ELM (Eukaryotes) (Baños et al. 2024), to be used with profile mixture models. The nonreversible matrices allow IQ-TREE 3 to reconstruct rooted trees without the requirement of an outgroup (e.g. Naser-Khdour et al. 2022; Wang and Luo 2025), as well as to infer the confidence of the placement of the root via the rootstrap (Naser-Khdour et al. 2022).

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026

Pandemic-scale phylogenetics

IQ-TREE 3 integrates CMAPLE (De Maio et al. 2023; Ly-Trong et al. 2024), an ultra-fast method for inferring trees with millions of closely related sequences, such as those produced by global SARS-CoV-2 (the virus that caused the COVID-19 pandemic) sequencing efforts (McBroome et al. 2021). This enables IQ-TREE 3 to perform functions that are useful in tracking large pathogen outbreaks, such as online sample placement. Additionally, IQ-TREE 3 introduces a dedicated "–pathogen" option for dynamically selecting between the CMAPLE and IQ-TREE search algorithms to optimize runtime based on sequence divergence: CMAPLE is selected for highly similar sequences, and IQ-TREE is selected for more distantly related sequences.

Moreover, IQ-TREE 3 and CMAPLE implement subtree pruning and regrafting tree assessment (SPRTA), a fast and appropriate approach for assessing phylogenetic uncertainty in genomic epidemiology (De Maio et al. 2025). SPRTA can be invoked via the "–sprta" option.

Time tree estimation

IQ-TREE 3 provides two facilities for divergence time estimation. The first is the Least Squares Dating method, LSD2 (To et al. 2016), which is suitable for dating very large phylogenies with tip date information, such as viral phylogenies. The second is the new IQ2MC workflow (Demotte et al. 2025) to facilitate Bayesian divergence time estimation with MCMCTree (dos Reis and Yang 2011). In the IQ2MC workflow, IQ-TREE 3 is used to compute the gradient vector and Hessian matrix of the branch lengths on a fixed tree topology, which are then used for fast approximate Bayesian MCMC sampling with MCMCTree. IQ-TREE dramatically speeds up the precalculation steps needed by MCMCTree, particularly for large phylogenomic datasets, and allows the use of the wide range of IQ-TREE's complex substitution models (such as mixture models), which were previously unavailable for time tree inference, and which are essential for dating ancient phylogenies.

Other advances

Finally, IQ-TREE 3 also includes a reimplementation of fast distance-based methods such as Neighbor-Joining (Saitou and Nei 1987) via the DecentTree library (Wang et al. 2023). It also includes a Python interface through piqtree (McArthur et al. 2026), an open-source Python package that exposes IQ-TREE's phylogenetic inference engine as a library of Python functions, thus facilitating interactive analyses in environments such as Jupyter notebooks and simplifying the development of custom phylogenetic workflows.

Conclusions

IQ-TREE version 3 presents a joint collaborative effort among researchers from across the globe, which substantially extends version 2. Particularly, we would like to advocate for the use of mixture models instead of only using partition models, especially as the single-model assumption within a locus is often violated (Ren et al. 2025). In addition, if the single-tree assumption common to concatenation approaches is violated, then users should employ the MAST model and/or interrogate their data by examining concordance factors (Lanfear and Hahn 2024).

We strive to continuously support the user community. Every new feature is accompanied by a tutorial and the manual is frequently updated at https://iqtree.github.io/doc/. User questions, issue reports, and feature requests can be submitted at the GitHub repository.

Acknowledgments

We thank James Barbetti for the DecentTree library and code optimization; Thu-Hien To for supports in integrating least square dating; Joshua Measure-Hughes for ARM architecture code; Cornelius Roemer, Cuong Cao Dang, Diep Thi Hoang, Justin Deneau, Martin Grigorov, Stefan Flaumberg for code contributions.

Funding

This work was financially supported by Chan-Zuckerberg Initiative grants for essential open-source software for science (EOSS-0000000132 and EOSS4-0000000312 to B.Q.M. and R.L.; EOSS5-0000000223 to G.H. and B.Q.M.), Simons Foundation and Gordon and Betty Moore Foundation grant (https://doi.org/10.46714/735923LPI. to A.J.R., E.S., and B.Q.M.), U.S. National Science Foundation grants (DBI-2146866 to M.W.H.; DMS-2331660 to H.B.), Australian Research Council grant for Discovery Project (DP200103151 to R.L., B.Q.M., and M.W.H.), Natural Sciences and Engineering Research Council of Canada grant (RGPIN-2022-05430 to A.J.R. and E.S.), Australian Research Council grant for Training Centre for Accelerated Future Crops Development (IC210100047), Medical Research Council and National Institute for Health and Care Research grant for Better Methods, Better Research (MR/Z503526/1 to N.D.M. and N.G.), Biotechnology and Biological Sciences Research Council (BB/T01282X/1 to M.d.R), and the Vietnam National Foundation for Science and Technology Development (102.05-2025.65 to L.S.V.). Computational resources were provided by the National Computational Infrastructure Australia (ANU Merit Allocation Scheme) and the Universität Wien and Medizinische Universität Wien (Center for Integrative Bioinformatics Vienna), Austria by Arndt von Haeseler.

Conflicts of Interest

None declared.

Data availability

There are no data with this paper.

References

Alves RJE, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun*. 2018:9:1517. https://doi.org/10.1038/s41467-018-03861-1.

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026

Baños H *et al.* GTRpmix: a linked general time-reversible model for profile mixture models. *Mol Biol Evol*. 2024:41:msae174. https://doi.org/10.1093/molbev/msae174.

Bolyen E *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019:37:852–857. https://doi.org/10.1038/s41587-019-0209-9.

Burki F *et al.* Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc Biol Sci*. 2016:283:20152802. https://doi.org/10.1098/rspb.2015.2802.

Cartwright RA. DNA assembly with gaps (Dawg): simulating sequence evolution. *Bioinformatics*. 2005:21:iii31–iii38. https://doi.org/10.1093/bioinformatics/bti1200.

Chen C *et al.* TBtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol Plant*. 2023:16:1733–1742. https://doi.org/10.1016/j.molp.2023.09.010.

Crisp MD *et al.* Perianth evolution and implications for generic delimitation in the eucalypts (Myrtaceae), including the description of the new genus, *Blakella*. *J Syst Evol*. 2024:62:942–962. https://doi.org/10.1111/jse.13047.

Crotty SM *et al.* GHOST: recovering historical signal from heterotachously evolved sequence alignments. *Syst Biol*. 2020:69:249–264. https://doi.org/10.1093/sysbio/syz051.

Csilléry K, Blum MGB, Gaggiotti OE, François O. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol*. 2010:25:410–418. https://doi.org/10.1016/j.tree.2010.04.001.

Dang CC *et al.* NQMaker: estimating time nonreversible amino acid substitution models. *Syst Biol*. 2022:71:1110–1123. https://doi.org/10.1093/sysbio/syac007.

De Maio N *et al.* Maximum likelihood pandemic-scale phylogenetics. *Nat Genet*. 2023:55:746–752. https://doi.org/10.1038/s41588-023-01368-0.

De Maio N, Ly-Trong N, Martin S, Minh BQ, Goldman N. Assessing phylogenetic confidence at pandemic scales. *Nature*. 2025:647:472–478. https://doi.org/10.1038/s41586-025-09567-x.

Demotte P *et al.* 2025. IQ2MC: A new framework to infer phylogenetic time trees using IQ-TREE 3 and MCMCTree with mixture models. EcoEvoRxiv [Internet]. https://doi.org/10.32942/X2CD2X.

Dombrowski N *et al.* Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat Commun*. 2020:11:3939. https://doi.org/10.1038/s41467-020-17408-w.

dos Reis M, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol*. 2011:28:2161–2172. https://doi.org/10.1093/molbev/msr045.

El-Gebali S *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019:47:D427–D432. https://doi.org/10.1093/nar/gky995.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.

Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Biol*. 1978:27:401–410. https://doi.org/10.1093/sysbio/27.4.401.

Fletcher W, Yang Z. INDELible: a flexible simulator of biological sequence evolution. *Mol Biol Evol*. 2009:26:1879–1888. https://doi.org/10.1093/molbev/msp098.

Galaxy Community. The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res*. 2024:52:W83–W94. https://doi.org/10.1093/nar/gkae410.

Giacomelli M *et al.* CAT-posterior mean site frequencies improves phylogenetic modeling under maximum likelihood and resolves Tardigrada as the sister of Arthropoda plus Onychophora. *Genome Biol Evol*. 2025:17:evae273. https://doi.org/10.1093/gbe/evae273.

Hadfield J *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018:34:4121–4123. https://doi.org/10.1093/bioinformatics/bty407.

Holmes EC *et al.* The origins of SARS-CoV-2: a critical review. *Cell*. 2021:184:4848–4856. https://doi.org/10.1016/j.cell.2021.08.017.

Jarvis ED *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014:346:1320–1331. https://doi.org/10.1126/science.1253451.

Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. 1992:8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017:14:587–589. https://doi.org/10.1038/nmeth.4285.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019:35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 2017:34:772–773. https://doi.org/10.1093/molbev/msw260.

Lanfear R, Hahn MW. The meaning and measure of concordance factors in phylogenomics. *Mol Biol Evol*. 2024:41:msae214. https://doi.org/10.1093/molbev/msae214.

Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. 2008:25:1307–1320. https://doi.org/10.1093/molbev/msn067.

Le SQ, Lartillot N, Gascuel O. Phylogenetic mixture models for proteins. *Philos Trans R Soc Lond B Biol Sci*. 2008:363:3965–3976. https://doi.org/10.1098/rstb.2008.0180.

Leuchtenberger AF *et al.* Distinguishing Felsenstein zone from Farris zone using neural networks. *Mol Biol Evol*. 2020:37:3632–3641. https://doi.org/10.1093/molbev/msaa164.

Ly-Trong N, Barca GMJ, Minh BQ. AliSim-HPC: parallel sequence simulator for phylogenetics. *Bioinformatics*. 2023:39:btad540. https://doi.org/10.1093/bioinformatics/btad540.

Ly-Trong N, Bielow C, De Maio N, Minh BQ. CMAPLE: efficient phylogenetic inference in the pandemic era. *Mol Biol Evol*. 2024:41:msae134. https://doi.org/10.1093/molbev/msae134.

Ly-Trong N, Naser-Khdour S, Lanfear R, Minh BQ. AliSim: a fast and versatile phylogenetic sequence simulator for the genomic era. *Mol Biol Evol*. 2022:39:msac092. https://doi.org/10.1093/molbev/msac092.

McArthur RN *et al.* Piqtree: a python package for seamless phylogenetic inference with IQ-TREE. *Mol Biol Evol*. 2026:43:msag061. https://doi.org/10.1093/molbev/msag061.

McBroome J *et al.* A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. *Mol Biol*

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026

Evol. 2021:38:5819–5824. https://doi.org/10.1093/molbev/msab264.

Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: *2010 Gateway computing environments workshop (GCE)*. IEEE; 2010, p. 1–8. https://doi.org/10.1109/gce.2010.5676129.

Minh BQ *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020b:37:1530–1534. https://doi.org/10.1093/molbev/msaa015.

Minh BQ, Dang CC, Vinh LS, Lanfear R. QMaker: fast and accurate method to estimate empirical models of protein evolution. *Syst Biol*. 2021:70:1046–1060. https://doi.org/10.1093/sysbio/syab010.

Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol*. 2020a:37:2727–2733. https://doi.org/10.1093/molbev/msaa106.

Misof B *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014:346:763–767. https://doi.org/10.1126/science.1257570.

Mo YK, Lanfear R, Hahn MW, Minh BQ. Updated site concordance factors minimize effects of homoplasy and taxon sampling. *Bioinformatics*. 2023:39:btac741. https://doi.org/10.1093/bioinformatics/btac741.

Naser-Khdour S, Quang Minh B, Lanfear R. Assessing confidence in root placement on phylogenies: an empirical study using nonreversible models for mammals. *Syst Biol*. 2022:71:959–972. https://doi.org/10.1093/sysbio/syab067.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015:32:268–274. https://doi.org/10.1093/molbev/msu300.

Parks DH *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018:36:996–1004. https://doi.org/10.1038/nbt.4229.

Parks DH *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022:50:D785–D794. https://doi.org/10.1093/nar/gkab776.

Puttick MN *et al.* The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr Biol*. 2018:28:733–745.e2. https://doi.org/10.1016/j.cub.2018.01.063.

Rambaut A *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020:5:1403–1407. https://doi.org/10.1038/s41564-020-0770-5.

Rambaut A, Grassly NC. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci*. 1997:13:235–238. https://doi.org/10.1093/bioinformatics/13.3.235.

Ran J-H, Shen T-T, Wang M-M, Wang X-Q. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proc Biol Sci*. 2018:285:20181012. https://doi.org/10.1098/rspb.2018.1012.

Redmond AK, McLysaght A. Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat Commun*. 2021:12:1783. https://doi.org/10.1038/s41467-021-22074-7.

Ren H, Wong TKF, Minh BQ, Lanfear R. MixtureFinder: estimating DNA mixture models for phylogenetic analyses. *Mol Biol Evol*. 2025:42:msae264. https://doi.org/10.1093/molbev/msae264.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987:4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Shen X-X *et al.* Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*. 2018:175:1533–1545.e20. https://doi.org/10.1016/j.cell.2018.10.023.

Smith ML, Hahn MW. Phylogenetic inference using generative adversarial networks. *Bioinformatics*. 2023:39:btad543. https://doi.org/10.1093/bioinformatics/btad543.

Soubrier J *et al.* The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol*. 2012:29:3345–3358. https://doi.org/10.1093/molbev/mss140.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014:30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stiller J *et al.* Complexity of avian evolution revealed by family-level genomes. *Nature*. 2024:629:851–860. https://doi.org/10.1038/s41586-024-07323-1.

To T-H, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 2016:65:82–97. https://doi.org/10.1093/sysbio/syv068.

Trost J *et al.* Simulations of sequence evolution: How (Un)realistic they are and why. *Mol Biol Evol*. 2024:41:msad277. https://doi.org/10.1093/molbev/msad277.

Vanderpool D *et al.* Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol*. 2020:18:e3000954. https://doi.org/10.1371/journal.pbio.3000954.

Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B. GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical Geno-taxonomic revision of the genus Stenotrophomonas. *Front Microbiol*. 2018:9:771. https://doi.org/10.3389/fmicb.2018.00771.

Wang H-C, Li K, Susko E, Roger AJ. A class frequency mixture model that adjusts for site-specific amino acid frequencies and improves inference of protein phylogeny. *BMC Evol Biol*. 2008:8:331. https://doi.org/10.1186/1471-2148-8-331.

Wang H-C, Minh BQ, Susko E, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst Biol*. 2018:67:216–235. https://doi.org/10.1093/sysbio/syx068.

Wang S, Luo H. Dating the bacterial tree of life based on ancient symbiosis. *Syst Biol*. 2025:74:639–655. https://doi.org/10.1093/sysbio/syae071.

Wang W *et al.* DecentTree: scalable neighbour-joining for the genomic era. *Bioinformatics*. 2023:39:btad536. https://doi.org/10.1093/bioinformatics/btad536.

Wehbi S *et al.* Order of amino acid recruitment into the genetic code resolved by last universal common ancestor's protein domains. *Proc Natl Acad Sci U S A*. 2024:121:e2410311121. https://doi.org/10.1073/pnas.2410311121.

Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001:18:

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026

691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851.

Williamson K *et al.* A robustly rooted tree of eukaryotes reveals their excavate ancestry. *Nature*. 2025:640:974–981. https://doi.org/10.1038/s41586-025-08709-5.

Wong TKF *et al.* MAST: phylogenetic inference with mixtures across sites and trees. *Syst Biol*. 2024:73:375–391. https://doi.org/10.1093/sysbio/syae008.

Wu S, Edwards S, Liu L. Genome-scale DNA sequence data and the evolutionary history of placental mammals. *Data Brief*. 2018:18:1972–1975. https://doi.org/10.1016/j.dib.2018.04.094.

Xiang C-Y *et al.* Using PhyloSuite for molecular phylogeny and tree-based analyses. *Imeta*. 2023:2:e87. https://doi.org/10.1002/imt2.87.

Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol*. 1994:39:306–314. https://doi.org/10.1007/BF00160154.

Zhang D *et al.* PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020:20:348–355. https://doi.org/10.1111/1755-0998.13096.

Downloaded from https://academic.oup.com/mbe/article/43/5/msag117/8669857 by Wissenschaftskolleg zu Berlin user on 25 May 2026