IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models

Thomas K.F. Wong[1,2,$], Nhan Ly-Trong[1,$], Huaiyan Ren[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], Gavin Huttley[2], Robert Lanfear[2,*], Bui Quang Minh[1,*]

1 School of Computing, College of Systems and Society, Australian National University, Australia

2 Research School of Biology, College of Science and Medicine, Australian National University, 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, 14195 Berlin, Germany

7 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK

8 Department of Biology and Department of Computer Science, Indiana University, Bloomington, Indiana 47405, USA

$ Co-first authors

* Co-corresponding authors: Bui Quang Minh (m.bui@anu.edu.au) and Robert Lanfear (rob.lanfear@anu.edu.au)

Abstract

IQ-TREE (http://www.iqtree.org) 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, 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, Schmidt, et al. 2020) 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); 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 divergences within the last decade in the SARS-CoV-2 virus (Rambaut et al. 2020; Holmes et al. 2021). The two major releases of IQ-TREE, version 1 (Nguyen et al. 2015) and 2 (Minh, Schmidt, et al. 2020), have been cited more than 20,000 and 10,000 times, respectively (Google Scholar; 3rd March 2025) and downloaded more than 1.1 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), 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, and pathogen tree inference.

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, Schmidt, et al. 2020); 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 shown to be crucial in answering many deep phylogenetic questions (e.g., Williamson et al. 2025). In particular, such models

help to avoid artefacts such as long-branch attraction (Felsenstein 1978; Wang et al. 2018) 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 partition models because no partition file is needed. This also removes user-subjectivity in determining the initial partitioning scheme. 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 GHOST (General Heterogeneous evolution On a Single Topology) 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, Hahn, et al. 2020), 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 version 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 data for machine learning applications (Smith and Hahn 2023). 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 significantly 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 multi-core CPUs and multi-node clusters.

Other advances

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 all estimated from large collections of alignments spanning many clades on the tree of life. The scale of modern phylogenomic datasets and modern 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 series of new taxon-specific matrices—Q.pfam, 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 non-reversible models (Dang et al. 2022) and a series of corresponding non-reversible 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.

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 taxa, such as those produced by global pathogen sequencing efforts (McBroome et al. 2021). This enables IQ-TREE 3 to perform functions which 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.

Finally, IQ-TREE 3 also includes a fast reimplementation of distance-based methods such as Neighbor-Joining (Saitou and Nei 1987) via the DecentTree library (Wang et al. 2023).

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 http://www.iqtree.org/doc/. User questions, issue reports, and feature requests can be submitted at the GitHub repository.

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 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.), an Australian Research Council Discovery Project (DP200103151 to R.L., B.Q.M., and M.W.H.), Natural Sciences and Engineering Research Council of Canada Discovery Grants (to A.J.R. and E.S.) and the European Molecular Biology Laboratory and by MRC-NIHR Better Methods, Better Research grant (MR/Z503526/1 to N.D.M. and N.G.). Computational resources were provided by the Australian Government through the National Computational Infrastructure (NCI) under the ANU Merit Allocation Scheme and the Center for Integrative Bioinformatics Vienna, Austria.

Acknowledgements

We thank James Barbetti for the DecentTree library and code optimization; Joshua Measure-Hughes for ARM architecture code; Cornelius Roemer, Cuong Cao Dang, Diep Thi Hoang, Justin Deneau, Martin Grigorov, Stefan Flaumberg for code contributions.

References

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

Baños H, Wong TKF, Daneau J, Susko E, Minh BQ, Lanfear R, Brown MW, Eme L, Roger AJ. 2024. GTRpmix: A linked general time-reversible model for profile mixture models. *Mol. Biol. Evol.* 41:msae174.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm

EJ, Arumugam M, Asnicar F, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37:852–857.

Burki F, Kaplan M, Tikhonenkov DV, Zlatogursky V, Minh BQ, Radaykina LV, Smirnov A, Mylnikov AP, Keeling PJ. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. Biol. Sci.* 283:20152802.

Cartwright RA. 2005. DNA assembly with gaps (Dawg): simulating sequence evolution. *Bioinformatics* 21 Suppl 3:iii31–iii38.

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, et al. 2023. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* 16:1733–1742.

Crisp MD, Minh BQ, Choi B, Edwards RD, Hereward J, Kulheim C, Lin YP, Meusemann K, Thornhill AH, Toon A, et al. 2024. Perianth evolution and implications for generic delimitation in the eucalypts (Myrtaceae), including the description of the new genus, *Blakella*. *J. Syst. Evol.* 62:942–962.

Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, Haeseler AV. 2020. GHOST: Recovering historical signal from heterotachously evolved sequence alignments. *Syst. Biol.* 69:249–264.

Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25:410–418.

Dang CC, Minh BQ, McShea H, Masel J, James JE, Vinh LS, Lanfear R. 2022. NQMaker: Estimating time nonreversible amino acid substitution models. *Syst. Biol.* 71:1110–1123.

De Maio N, Kalaghatgi P, Turakhia Y, Corbett-Detig R, Minh BQ, Goldman N. 2023. Maximum likelihood pandemic-scale phylogenetics. *Nat. Genet.* 55:746–752.

Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, Rinke C, Spang A. 2020. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* 11:3939.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47:D427–D432.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238.

Felsenstein J. 1978. Cases in which Parsimony or Compatibility Methods will be Positively Misleading. *Syst. Biol.* 27:401–410.

Fletcher W, Yang Z. 2009. INDELible: A Flexible Simulator of Biological Sequence Evolution. *Mol. Biol. Evol.* 26:1879–1888.

Galaxy Community. 2024. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 52:W83–W94.

Giacomelli M, Vecchi M, Guidetti R, Rebecchi L, Donoghue PCJ, Lozano-Fernandez J, Pisani D. 2025. CAT-posterior mean site frequencies improves phylogenetic modeling under maximum likelihood and resolves Tardigrada as the sister of Arthropoda plus Onychophora. *Genome Biol. Evol.* 17:evae273.

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123.

Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, Wertheim JO, Anthony SJ, Barclay WS, Boni MF, Doherty PC, et al. 2021. The origins of SARS-CoV-2: A critical review. *Cell* 184:4848–4856.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.

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

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

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

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

Lanfear R, Hahn MW. 2024. The meaning and measure of concordance factors in phylogenomics. *Mol. Biol. Evol.* 41:msae214.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25:1307–1320.

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

Ly-Trong N, Barca GMJ, Minh BQ. 2023. AliSim-HPC: parallel sequence simulator for phylogenetics. *Bioinformatics* 39:btad540.

Ly-Trong N, Bielow C, De Maio N, Minh BQ. 2024. CMAPLE: Efficient phylogenetic inference in the pandemic era. *Mol. Biol. Evol.* 41:msae134.

Ly-Trong N, Naser-Khdour S, Lanfear R, Minh BQ. 2022. AliSim: A Fast and Versatile Phylogenetic Sequence Simulator for the Genomic Era. *Mol. Biol. Evol.* 39:msac092.

McBroome J, Thornlow B, Hinrichs AS, Kramer A, De Maio N, Goldman N, Haussler D, Corbett-Detig R, Turakhia Y. 2021. A daily-updated database and tools for comprehensive SARS-CoV-2 mutation-annotated trees. *Mol. Biol. Evol.* 38:5819–5824.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). IEEE. http://dx.doi.org/10.1109/gce.2010.5676129

Minh BQ, Dang CC, Vinh LS, Lanfear R. 2021. QMaker: Fast and accurate method to estimate empirical models of protein evolution. *Syst. Biol.* 70:1046–1060.

Minh BQ, Hahn MW, Lanfear R. 2020. New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* 37:2727–2733.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.

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

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

Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50:D785–D794.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36:996–1004.

Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D, et al. 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* 28:733–745.e2.

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

Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, du Plessis L, Pybus OG. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5:1403–1407.

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

Redmond AK, McLysaght A. 2021. Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat. Commun.* 12:1783.

Ren H, Wong TKF, Minh BQ, Lanfear R. 2025. MixtureFinder: Estimating DNA mixture models for phylogenetic analyses. *Mol. Biol. Evol.* 42:msae264.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.

Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, Haase MAB, Wisecaver JH, Wang M, Doering DT, et al. 2018. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell* 175:1533–1545.e20.

Smith ML, Hahn MW. 2023. Phylogenetic inference using generative adversarial networks. *Bioinformatics* 39:btad543.

Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates.

Mol. Biol. Evol. 29:3345–3358.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Stiller J, Feng S, Chowdhury A-A, Rivas-González I, Duchêne DA, Fang Q, Deng Y, Kozlov A, Stamatakis A, Claramunt S, et al. 2024. Complexity of avian evolution revealed by family-level genomes. *Nature* 629:851–860.

Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Hibbins MS, Williamson RJ, et al. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biol.* 18:e3000954.

Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B. 2018. 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.* 9:771.

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

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

Wang W, Barbetti J, Wong T, Thornlow B, Corbett-Detig R, Turakhia Y, Lanfear R, Minh BQ. 2023. DecentTree: scalable Neighbour-Joining for the genomic era. *Bioinformatics* 39:btad536.

Wehbi S, Wheeler A, Morel B, Manepalli N, Minh BQ, Lauretta DS, Masel J. 2024. 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.* 121:e2410311121.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18:691–699.

Williamson K, Eme L, Baños H, McCarthy CGP, Susko E, Kamikawa R, Orr RJS, Muñoz-Gómez SA, Minh BQ, Simpson AGB, et al. 2025. A robustly rooted tree of eukaryotes reveals their excavate ancestry. *Nature*. https://doi.org/10.1038/s41586-025-08709-5

Wong TKF, Cherryh C, Rodrigo AG, Hahn MW, Minh BQ, Lanfear R. 2024. MAST: Phylogenetic inference with mixtures across sites and trees. *Syst. Biol.* 73:375–391.

Wu S, Edwards S, Liu L. 2018. Genome-scale DNA sequence data and the evolutionary history of placental mammals. *Data Brief* 18:1972–1975.

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

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20:348–355.