N/A, , 1–5
doi:

piqtree: A Python Package for Seamless Phylogenetic Inference with IQ-TREE

Robert Neil McArthur,[1,*] Thomas King-Fung Wong,[1,2,3] Yapeng Lang,[1,] Richard Andrew Morris,[1,] Katherine Caley,[1,] Vijini Mallawaarachchi,[4,] Bui Quang Minh[2,*] and Gavin Huttley[1,*]

[1]Research School of Biology, Australian National University, [2]School of Computing, Australian National University, [3]Mathematical Sciences Institute, Australian National University and [4]Flinders Accelerator for Microbiome Exploration, Flinders University

*Corresponding author. robert-nm-correspondence@outlook.com Gavin.Huttley@anu.edu.au, M.Bui@anu.edu.au

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

`piqtree` is an easy to use, open-source Python package that directly exposes IQ-TREE's phylogenetic inference engine. It offers Python functions for performing many of IQ-TREE's capabilities including phylogenetic reconstruction, ultrafast bootstrapping, branch length optimisation, ModelFinder, rapid neighbour-joining, and more. By exposing IQ-TREE's algorithms within Python, `piqtree` greatly simplifies the development of new phylogenetic workflows through seamless interoperability with other Python libraries and tools mediated by the `cogent3` package. It also enables users to perform interactive analyses with IQ-TREE through, for instance, Jupyter notebooks. We present the key features available in the `piqtree` library and a small case study that showcases its interoperability. The `piqtree` library can be installed with `pip install piqtree`, with the documentation available at `https://piqtree.readthedocs.io` and source at `https://github.com/iqtree/piqtree`.

Key words: bioinformatics, phylogenetics, gene tree, non-reversible Markov process, Python, Jupyter, cogent3

Introduction

We introduce `piqtree`, an open-source Python package that offers direct access to the maximum-likelihood phylogenetic inference capabilities of IQ-TREE on all major operating systems. It supports interactive analyses and simplifies the integration of IQ-TREE into workflows via `cogent3` functions. Bespoke phylogenetic analyses that integrate IQ-TREE's advanced algorithms can now be done with ease in Python.

IQ-TREE (Nguyen et al., 2015; Minh et al., 2020; Wong et al., 2025) offers a command-line interface to phylogenetic inference using maximum likelihood. Beyond direct usage by researchers, IQ-TREE has been integrated into other applications and pipelines. These include the workflow platform Galaxy (The Galaxy Community, 2024), the phylogenetics gateway CIPRES (Miller et al., 2010), the microbiome analysis application QIIME2 (Bolyen et al., 2019), the orthology inference OrthoFinder (Emms and Kelly, 2019), and the real-time tracking of pathogen evolution Nextstrain (Hadfield et al., 2018). Given the wide popularity of Python combined with the extensive ecosystem of scientific packages available in it, we have developed `piqtree` to directly expose the core functionality of IQ-TREE. `piqtree` is not intended as a replacement for the IQ-TREE application, but rather as a modular Python interface to its efficient algorithms. This integration is mediated through `cogent3` (COmparative GENomics Toolkit 3 Huttley et al., 2025b). While `cogent3` also implements robust algorithms and distinctive molecular evolutionary analyses (e.g. Schranz et al., 2008; Yap et al., 2010; Lin et al., 2012; Verbyla et al., 2013; Kaehler et al., 2015, 2017), its primary role for `piqtree` is to provide a library of fundamental biological data objects and to simplify the integration of `piqtree` with other tools.

We introduce the key features of `piqtree` and present a small case study emphasising the ease of exploiting data-level parallelism and generating workflows via the `cogent3` app ecosystem. `piqtree` can be installed from the Python Package Index. The data and code used for this work is available at DOI 10.5281/zenodo.15875241. An archive of the `piqtree` version is DOI 10.5281/zenodo.15860649. The source code and documentation are available at `https://github.com/iqtree/piqtree` and `https://piqtree.readthedocs.io` respectively.

© The Author.

Key Features

Phylogenetic reconstruction

A maximum-likelihood phylogenetic tree can be constructed from a `cogent3` alignment object under a chosen substitution model with the `build_tree` function. Users can further customise the model by specifying how state frequencies are chosen (either empirical, equal, or optimised through maximum likelihood). Further, rate heterogeneity across sites can be incorporated by allowing for a proportion of invariable sites and by using the discrete Gamma or FreeRate models with a custom number of rate categories (Yang, 1994b; Gu et al., 1995; Yang, 1995; Soubrier et al., 2012). Ultrafast bootstrapping can be performed by specifying the number of bootstrap replicates (Minh et al., 2013; Hoang et al., 2018). The number of threads to parallelise across can be specified for more efficient computation, and a random seed may be set for reproducibility.

`piqtree` currently supports all IQ-TREE DNA models (including Lie Markov models, Woodhams et al., 2015), and amino-acid exchange rate matrices. The full listing of the available models can be found in the documentation.

Fitting branch lengths to a tree topology

The `fit_tree` function enables branch lengths to be fitted to a fixed tree topology according to a `cogent3` alignment object and a chosen substitution model. The function supports the same models as `build_tree`. Nucleotide or amino acid probabilities and rate parameters are annotated as appropriate along the resulting `cogent3` tree object's edges. The number of threads to parallelise across may also be set for faster computation.

Selecting an evolutionary model

The `model_finder` function exposes the model finding capabilities of IQ-TREE. The resulting object provides convenient interfaces to the models selected by the different information measures. The model selected according to the user preferred information criteria (e.g. BIC or AIC, Posada and Buckley, 2004) can be easily applied to either `build_tree` or `fit_tree`.

Estimating pairwise distances

The `jc_distances` function computes pairwise genetic distances from a `cogent3` alignment object using the JC69 model (Jukes and Cantor, 1969), returning a `cogent3` distance matrix object. The user can specify the number of threads to use for the computations.

Estimate a phylogeny using RapidNJ

From a `cogent3` distance matrix object, the `nj_tree` function uses the efficient Rapid Neighbour-Joining tree algorithm used internally by IQ-TREE (Simonsen et al., 2011). When a user has `piqtree` installed, methods on `cogent3` alignments and distance matrix objects will use this implementation rather than the native `cogent3` methods if the `piqtree` hook is specified. By default, negative branch lengths resulting from the method are set to 0.

Simulate a random tree

The `random_tree` function can be used to make a randomly generated tree. The number of taxa and strategy used to generate the tree can be specified. Options to generate include balanced trees, birth-death trees, caterpillar trees, star trees, uniform trees, or Yule-Harding trees. A random seed can be specified to ensure reproducibility.

Constructing a consensus tree

The `consensus_tree` function can be used to construct a consensus tree from a collection of trees on the same taxa set. By default, the majority-rule consensus tree is constructed (Margush and McMorris, 1981). Additionally, a minimum support value (0-1) can be specified, representing a proportion of trees. Consequently, a clade must appear in more trees than this proportion in order to appear in the resulting consensus tree. If a minimum support of 1 is specified, the algorithm constructs the strict consensus tree (Sokal and Rohlf, 1981). If a minimum support of 0 is specified, the algorithm constructs an extended majority-rule consensus tree (Felsenstein, 1985).

Case study – combining a non-stationary Markov process and rooted supertree method

In this case study, we demonstrate the interoperability of `piqtree` with other Python packages and how `cogent3` greatly simplifies data-level parallelisation of phylogenetic analyses with so called "apps". We take advantage of the ability of IQ-TREE to estimate the placement of the root on the phylogenies (Naser-Khdour et al., 2022) and combine it with a fast and accurate rooted supertree method (McArthur et al., 2024).

Chiari et al. (2012) applied time-reversible nucleotide substitution models and a range of rate heterogeneity strategies to a concatenated alignment derived from 248 protein-coding genes. Their analyses assessed the relative positions of turtles, crocodilians, and birds. Illustrating the patchiness of genome representation in that data, only five genes had all species present. For the concatenated analysis, when a gene was missing from a taxon, it was represented as missing data.

`piqtree` allows using stationary but non-reversible models implemented in IQ-TREE. We utilise the most general stationary non-reversible substitution model for nucleotides (Yang, 1994a). For individual alignments, we exclude taxa for which a gene was not present. This produces a collection of alignments with substantial but incomplete overlap in their taxa. The rooted phylogenies estimated from these analyses are then a suitable input to the recently published spectral cluster supertree method which also exposes its methods as a `cogent3` app (McArthur et al., 2024).

`cogent3` provides an "app" mechanism for validating the interoperability of different functions. An app is a single purpose Python function that defines input and return data types. `cogent3` leverages these type definitions to allow a series of apps to be added together to generate a "composed pipeline" that can effectively utilise data-level parallelism. Additionally, the alignment file from which a phylogeny was derived is automatically tracked while using data-level parallelism in the `cogent3` ecosystem. `piqtree` has exposed many IQ-TREE functions as apps to take advantage of these capabilities. All app names take the form `piq_<func_name>`, for example the app for `build_tree` is named `piq_build_tree`.

The eight statements in Figure 1 constitute the parallelised algorithm to produce a rooted supertree from the 248 gene alignments. Figure 2 shows the alternative serial version using the `piqtree` functions directly. In the parallelised instance, we used the `piq_build_tree` app for estimation of a rooted phylogenetic tree with the UNREST model (Yang, 1994a). The app is applied

```python
from cogent3 import get_app, open_data_store
from sc_supertree import construct_supertree

loader = get_app("load_aligned", moltype="dna") ①

tree_builder = get_app("piq_build_tree",
                       model="UNREST") ②

app = loader + tree_builder ③

dstore = open_data_store(TURTLES_PATH,
                         suffix="fa") ⑤

trees = list(app.as_completed(dstore,
                        parallel=True)) ④

supertree = construct_supertree(trees,
    pcg_weighting="branch") ⑥
```

Fig. 1: Combining `piqtree` with other apps to construct a rooted supertree. A sequence loader ① and `piqtree` ② are added together to produce a composed app ③ and applied in parallel ④ to a directory of gene alignment files (file names end with `.fa`) ⑤ to produce a list of `cogent3` tree objects. Spectral Cluster Supertree ⑥ is then applied to construct the supertree from these trees.

```python
from pathlib import Path

from cogent3 import load_aligned_seqs
from piqtree import build_tree
from sc_supertree import construct_supertree

aln_dir = Path("turtle_partitions/") ①

trees = []
for aln_file in aln_dir.glob("*.fa"):
    aln = load_aligned_seqs(aln_file,
                            moltype="dna") ②
    trees.append(build_tree(aln, "UNREST")) ③

supertree = construct_supertree(trees,
    pcg_weighting="branch") ④
```

Fig. 2: Using `piqtree` to construct a rooted supertree. The FASTA files for each gene are extracted from the data directory ①. Each alignment is loaded into Python ② before `build_tree` is used to reconstruct the corresponding gene tree under the UNREST model ③. Spectral Cluster Supertree ④ is then applied to construct the supertree from these gene trees. There is no parallelisation of the computations.

in parallel to all FASTA files in the `TURTLES_PATH` directory. The result of this operation is a list of `cogent3` tree objects, each of which has an attribute that identifies the path to the alignment file that produced it. We then use the Spectral Cluster Supertree (McArthur et al., 2024) algorithm to combine the resulting gene trees into the single supertree displayed in Figure 3.

The result of the analysis substantively recapitulates that previously reported. It should first be noted that the birds (Gallus and Taeniopygia) are generally accepted to be the sister group of crocodilians (Alligator and Caiman) (Chiari et al., 2012). However, in the Chiari et al. (2012) analysis, when a consensus tree was formed from the set of gene trees for the 248 genes in the nucleotide dataset, the bird grouping was placed one taxonomical level higher. Our analysis aggregating the collection of calculated gene trees reproduced the same tree topology. The root in the supertree is two edges away from the expected position on the Protopterus (lungfish) branch. This is unsurprising as it is long established that stationary non-reversible models have low statistical power to detect the root position (Yap and Speed, 2005).

The phylogenetic Python data science ecosystem

Users of `piqtree` will be able to take advantage of Python's sophisticated support for interactive exploratory data analyses and other Python phylogenetic packages that complement it. For instance, while `cogent3` has its own built-in tree visualisation capabilities (e.g. Figure 3) the ETE3 toolkit provides more comprehensive capabilities (Huerta-Cepas et al., 2016). A `cogent3` app `cogent3-ete3` can be used to transform the `cogent3` trees into ETE3. The `diverse-seq` package (Huttley et al., 2025a) for selecting representative sets of sequences for the development of phylogenetic workflows also provides `cogent3` apps. The `phylim` package (Lang and Huttley, 2025) is explicitly designed to work with `piqtree` and check that a fitted phylogenetic model is valid, which can address uncertainties regarding the reliability of inference when site-saturation is a concern.

Conclusion

`piqtree` combines the algorithmic sophistication of IQ-TREE with the accessibility and flexibility of data analysis using Python. This combination enables exploratory analyses and is expected to facilitate development of new phylogeny based methodologies.

Acknowledgments

Chan-Zuckerberg Initiative EOSS5-0000000223 awarded to GAH, VM and BQM and EOSS4-0000000312 to BQM. National Science Foundation 2333243 awarded to Joanna Masel with a subcontract to BQM.

References

Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, Eric J. Alm, Manimozhiyan Arumugam, Francesco Asnicar, Yang Bai, Jordan E. Bisanz, Kyle Bittinger, Asker Brejnrod, Colin J. Brislawn, C. Titus Brown, Benjamin J. Callahan, Andrés Mauricio Caraballo-Rodríguez, John Chase, Emily K. Cope, Ricardo Da Silva, Christian Diener, Pieter C. Dorrestein, Gavin M. Douglas, Daniel M. Durall, Claire Duvallet, Christian F. Edwardson, Madeleine Ernst, Mehrbod Estaki, Jennifer Fouquier, Julia M. Gauglitz, Sean M. Gibbons, Deanna L. Gibson, Antonio Gonzalez, Kestrel Gorlick, Jiarong Guo, Benjamin Hillmann, Susan Holmes, Hannes Holste, Curtis Huttenhower, Gavin A. Huttley, Stefan Janssen, Alan K. Jarmusch, Lingjing Jiang, Benjamin D. Kaehler, Kyo Bin Kang, Christopher R. Keefe, Paul Keim, Scott T. Kelley, Dan Knights, Irina Koester, Tomasz Kosciolek, Jorden Kreps,

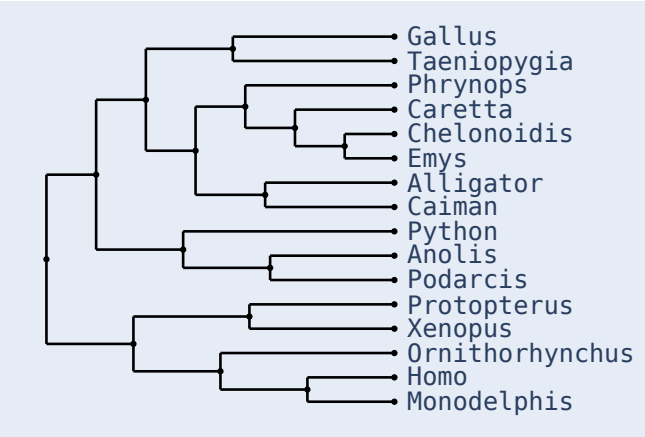

Fig. 3: The supertree formed from the gene trees of the turtles dataset (Chiari et al., 2012). The image was created using the `cogent3` built-in dendrogram algorithms which use the Plotly interactive graphics library (Plotly Technologies Inc, 2015).

Morgan G. I. Langille, Joslynn Lee, Ruth Ley, Yong-Xin Liu, Erikka Loftfield, Catherine Lozupone, Massoud Maher, Clarisse Marotz, Bryan D. Martin, Daniel McDonald, Lauren J. McIver, Alexey V. Melnik, Jessica L. Metcalf, Sydney C. Morgan, Jamie T. Morton, Ahmad Turan Naimey, Jose A. Navas-Molina, Louis Felix Nothias, Stephanie B. Orchanian, Talima Pearson, Samuel L. Peoples, Daniel Petras, Mary Lai Preuss, Elmar Pruesse, Lasse Buur Rasmussen, Adam Rivers, Michael S. Robeson, Patrick Rosenthal, Nicola Segata, Michael Shaffer, Arron Shiffer, Rashmi Sinha, Se Jin Song, John R. Spear, Austin D. Swafford, Luke R. Thompson, Pedro J. Torres, Pauline Trinh, Anupriya Tripathi, Peter J. Turnbaugh, Sabah Ul-Hasan, Justin J. J. van der Hooft, Fernando Vargas, Yoshiki Vázquez-Baeza, Emily Vogtmann, Max von Hippel, William Walters, Yunhu Wan, Mingxun Wang, Jonathan Warren, Kyle C. Weber, Charles H. D. Williamson, Amy D. Willis, Zhenjiang Zech Xu, Jesse R. Zaneveld, Yilong Zhang, Qiyun Zhu, Rob Knight, and J. Gregory Caporaso. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8): 852–857, August 2019. ISSN 1546-1696. doi: 10.1038/s41587-019-0209-9. URL https://www.nature.com/articles/s41587-019-0209-9. Publisher: Nature Publishing Group.

Ylenia Chiari, Vincent Cahais, Nicolas Galtier, and Frédéric Delsuc. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biology*, 10(1):65, July 2012. ISSN 1741-7007. doi: 10.1186/1741-7007-10-65. URL https://doi.org/10.1186/1741-7007-10-65.

David M. Emms and Steven Kelly. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1):1–14, November 2019. ISSN 1474-760X. doi: 10.1186/s13059-019-1832-y. URL https://doi.org/10.1186/s13059-019-1832-y.

Joseph Felsenstein. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. *Evolution*, 39(4):783–791, July 1985. ISSN 0014-3820. doi: 10.1111/j.1558-5646.1985.tb00420.x. URL https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

X Gu, Y X Fu, and W H Li. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution*, 12(4):546–557, July 1995. ISSN 0737-4038. doi: 10.1093/oxfordjournals.molbev.a040235. URL https://doi.org/10.1093/oxfordjournals.molbev.a040235.

James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, December 2018. ISSN 1367-4803. doi: 10.1093/bioinformatics/bty407. URL https://doi.org/10.1093/bioinformatics/bty407.

Diep Thi Hoang, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, and Le Sy Vinh. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2):518–522, February 2018. ISSN 0737-4038. doi: 10.1093/molbev/msx281. URL https://doi.org/10.1093/molbev/msx281.

Jaime Huerta-Cepas, François Serra, and Peer Bork. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, 33(6):1635–1638, June 2016. ISSN 0737-4038. doi: 10.1093/molbev/msw046. URL https://doi.org/10.1093/molbev/msw046.

Gavin Huttley, Katherine Caley, and Robert McArthur. diverse-seq: an application for alignment-free selecting and clustering biological sequences. *Journal of Open Source Software*, 10(110): 7765, June 2025a. ISSN 2475-9066. doi: 10.21105/joss.07765. URL https://joss.theoj.org/papers/10.21105/joss.07765.

Gavin Huttley, Katherine Caley, Nick-Foto, Stephen Ka-Wah Ma, Moses Koh, richard morris, Robert McArthur, Daniel McDonald, Fred Jaya, Peter Maxwell, Mike Robeson, James Martini, thomas la, Jai Ram Rideout, justin212k, Kyle Bittinger, tla6677658, Yapeng Lang, Benjamin Millar, jistombaugh, Jesse Zaneveld, Catherine Lozupone, Antonio Gonzalez, i000, Jens Reeder, Ryan Liu, Alex Iliadis, Tony, and Yixuan567. cogent3: making sense of sequence, May 2025b. URL https://zenodo.org/records/15500183.

T Jukes and C Cantor. Mammalian Protein Metabolism. *Evolution of protein molecules*, pages 21–132, 1969.

Benjamin D Kaehler, Von Bing Yap, Rongli Zhang, and Gavin A Huttley. Genetic distance for a general non-stationary Markov substitution process. *Systematic Biology*, 64(2):281–93, March 2015. ISSN 1076-836X. doi: 10.1093/sysbio/syu106. URL http://www.ncbi.nlm.nih.gov/pubmed/25503772.

Benjamin D Kaehler, Von Bing Yap, and Gavin A Huttley. Standard Codon Substitution Models Overestimate Purifying Selection for Non-Stationary Data. *Genome Biology and Evolution*, 9(1):134–149, January 2017. doi: 10.1093/gbe/evw308. URL https://academic.oup.com/gbe/article-lookup/doi/10.1093/gbe/evw308. Publisher: Oxford University Press.

Yapeng Lang and Gavin Huttley. phylim: detecting the limits of phylogenetic inference, June 2025. URL https://zenodo.org/records/15717906.

Yu Lin, Vaibhav Rajan, and Bernard M E Moret. A metric for phylogenetic trees based on matching. *IEEE/ACM Trans Comput Biol Bioinform*, 9(4):1014–1022, 2012. doi: 10.1109/TCBB.2011.157. Place: United States ISBN: 1557-9964.

T. Margush and F. R. McMorris. Consensus$n$-trees. *Bulletin of Mathematical Biology*, 43(2):239–244, January 1981. ISSN 0092-8240. doi: 10.1016/S0092-8240(81)90019-7. URL https://www.sciencedirect.com/science/article/pii/S0092824081900197.

Robert N. McArthur, Ahad N. Zehmakan, Michael A. Charleston, Yu Lin, and Gavin Huttley. Spectral cluster supertree: fast and statistically robust merging of rooted phylogenetic trees. *Frontiers in Molecular Biosciences*, 11, October 2024. ISSN 2296-889X. doi: 10.3389/fmolb.2024.1432495. URL https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2024.1432495/full. Publisher: Frontiers.

Mark A. Miller, Wayne Pfeiffer, and Terri Schwartz. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pages 1–8, November 2010. doi: 10.1109/GCE.2010.5676129. URL https://ieeexplore.ieee.org/abstract/document/5676129. ISSN: 2152-1093.

Bui Quang Minh, Minh Anh Thi Nguyen, and Arndt von Haeseler. Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30(5):1188–1195, May 2013. ISSN 0737-4038. doi: 10.1093/molbev/mst024. URL https://doi.org/10.1093/molbev/mst024.

Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, May 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015. URL https://doi.org/10.1093/molbev/msaa015.

Suha Naser-Khdour, Bui Quang Minh, and Robert Lanfear. Assessing Confidence in Root Placement on Phylogenies: An Empirical Study Using Nonreversible Models for Mammals. *Systematic Biology*, 71(4):959–972, July 2022. ISSN 1063-5157. doi: 10.1093/sysbio/syab067. URL https://doi.org/10.1093/sysbio/syab067.

Lam-Tung Nguyen, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1):268–274, January 2015. ISSN 0737-4038. doi: 10.1093/molbev/msu300. URL https://doi.org/10.1093/molbev/msu300.

. Plotly Technologies Inc. Collaborative data science, 2015. URL https://plot.ly. Place: Montreal, QC Publisher: Plotly Technologies Inc.

David Posada and Thomas R Buckley. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol*, 53(5):793–808, 2004. doi: 10.1080/10635150490522304. ISBN: 1063-5157 (Print).

Harold W Schranz, Von Bing Yap, Simon Easteal, Rob Knight, and Gavin A Huttley. Pathological rate matrices: from primates to pathogens. *BMC Bioinformatics*, 9:550, 2008. doi: 10.1186/1471-2105-9-550. URL https://www.ncbi.nlm.nih.gov/pubmed/19099591. Place: England ISBN: 1471-2105.

Martin Simonsen, Thomas Mailund, and Christian N. S. Pedersen. Inference of Large Phylogenies Using Neighbour-Joining. In Ana Fred, Joaquim Filipe, and Hugo Gamboa, editors, *Biomedical Engineering Systems and Technologies*, pages 334–344, Berlin, Heidelberg, 2011. Springer. ISBN 978-3-642-18472-7. doi: 10.1007/978-3-642-18472-7_26.

Robert R. Sokal and F. James Rohlf. Taxonomic Congruence in the Leptopodomorpha Re-Examined. *Systematic Zoology*, 30(3):309–325, 1981. ISSN 0039-7989. doi: 10.2307/2413252. URL https://www.jstor.org/stable/2413252. Publisher: [Oxford University Press, Society of Systematic Biologists, Taylor & Francis, Ltd.].

Julien Soubrier, Mike Steel, Michael S.Y. Lee, Clio Der Sarkissian, Stéphane Guindon, Simon Y.W. Ho, and Alan Cooper. The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution*, 29(11):3345–3358, November 2012. ISSN 0737-4038. doi: 10.1093/molbev/mss140. URL https://doi.org/10.1093/molbev/mss140.

The Galaxy Community. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, 52(W1):W83–W94, July 2024. ISSN 0305-1048. doi: 10.1093/nar/gkae410. URL https://doi.org/10.1093/nar/gkae410.

Klara L Verbyla, Von Bing Yap, Anuj Pahwa, Yunli Shao, and Gavin A Huttley. The Embedding Problem for Markov Models of Nucleotide Substitution. *PLoS ONE*, 8(7):e69187, July 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0069187. URL https://dx.plos.org/10.1371/journal.pone.0069187. Publisher: Public Library of Science ISBN: 1932-6203.

Thomas K. F. Wong, Nhan Ly-Trong, Huaiyan Ren, Hector Baños, Andrew J. Roger, Edward Susko, Chris Bielow, Nicola De Maio, Nick Goldman, Matthew W. Hahn, Gavin Huttley, Rob Lanfear, and Bui Quang Minh. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models, April 2025. URL https://ecoevorxiv.org/repository/view/8916/. Publisher: EcoEvoRxiv.

Michael D. Woodhams, Jesús Fernández-Sánchez, and Jeremy G. Sumner. A New Hierarchy of Phylogenetic Models Consistent with Heterogeneous Substitution Rates. *Systematic Biology*, 64(4):638–650, July 2015. ISSN 1063-5157. doi: 10.1093/sysbio/syv021. URL https://doi.org/10.1093/sysbio/syv021.

Z Yang. Estimating the pattern of nucleotide substitution. *J Mol Evol*, 39(1):105–111, 1994a. doi: 10.1007/BF00178256. ISBN: 0022-2844 (Print).

Z Yang. A space-time process model for the evolution of DNA sequences. *Genetics*, 139(2):993–1005, February 1995. ISSN 1943-2631. doi: 10.1093/genetics/139.2.993. URL https://doi.org/10.1093/genetics/139.2.993.

Ziheng Yang. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, 39(3):306–314, September 1994b. ISSN 1432-1432. doi: 10.1007/BF00160154. URL https://doi.org/10.1007/BF00160154.

Von Bing Yap and Terry Speed. Rooting a phylogenetic tree with nonreversible substitution models. *BMC Evol Biol*, 5(1):2, 2005. doi: 10.1186/1471-2148-5-2. ISBN: 1471-2148 (Electronic).

Von Bing Yap, Helen Lindsay, Simon Easteal, and Gavin Huttley. Estimates of the Effect of Natural Selection on Protein-Coding Content. *Molecular Biology and Evolution*, 27(3):726–734, March 2010. ISSN 0737-4038. doi: 10.1093/molbev/msp232. URL https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msp232.