The Journal of Open Source Software

REQ: assessing branch supports of a distance-based phylogenetic tree with the rate of elementary quartets

Alexis Criscuolo[1]

1 Institut Pasteur – Bioinformatics and Biostatistics Hub – C3BI, USR 3756 IP CNRS – 25-28 Rue du Docteur Roux, 75015 Paris, France

DOI: 10.21105/joss.00824

Software

- Review
- Repository
- Archive

Submitted: 13 July 2018
Published: 16 July 2018

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

REQ is a program for quickly estimating a confidence value at each branch of a distance-based phylogenetic tree. Branch support assessment is commonly based on bootstrap procedures (Felsenstein 1985; Makarenkov et al. 2010; Lemoine et al. 2018). Unfortunately, as they are based on numerous resampling of aligned characters, such procedures require long running times, despite some recent advances (Minh, Nguyen, and Haeseler 2013; DT Hoang et al. 2018; DY Hoang et al. 2018). In fact, direct branch support methods were already developed for character-based approaches that optimize maximum-parsimony or maximum-likelihood criteria, in order to achieve faster running times (Bremer 1988, 1994; Anisimova and Gascuel 2006; Anisimova et al. 2011). However, to our knowledge, no practical implementation of direct branch support methods is currently available for distance-based approaches.

Distance-based approaches proceed in two steps: a pairwise evolutionary distance is estimated between each pair of (biological) objects, and, next, an algorithm is used to infer the tree with branch lengths that best fits the evolutionary distance matrix (Pardi and Gascuel 2016). Because of their speed, distance-based methods are widely used for inferring phylogenetic trees. Moreover, as such algorithms only need a distance matrix, they allow phylogenetic analyses to be carried out from a wide range of data types, e.g. DNA-DNA hybridization experiments (Krajewski and Dickerman 1990), gene orders (Wang et al. 2006; House, Pellegrini, and Fitz-Gibbon 2014), gene content (Spencer, Bryant, and Susko 2007), or unaligned genome sequences (Chapus et al. 2005; Henz et al. 2005; Cohen and Chor 2012; Garcia-Hermoso et al. 2018). Nevertheless, in such cases, standard bootstrap-based methods can not be used for estimating branch confidence values.

In order to fill this void, the program *REQ* was developed. This tool estimates the rate of elementary quartets (REQ) for each branch of a given phylogenetic tree from the associated distance matrix, as described by (Guénoche and Garreta 2001). This method simply computes the proportion of four-leaf subtrees (i.e. quartets) induced by every internal branch that are supported by the four-point condition applied to the six corresponding pairwise evolutionary distances (Zaretskii 1965; Buneman 1971). Therefore, this measure is not based on a random sampling (such as bootstrap-based confidence supports). The closer this measure is to 1, the more the corresponding branch is fully supported by the pairwise evolutionary distances.

The program *REQ* is available on GitLab under the licence GNU GPLv3. Implemented in Java, *REQ* could be used on every operating system with a simple command line. *REQ* only needs two input files: a distance matrix file in either PHYLIP lower-triangular or square format, and a phylogenetic tree file in NEWICK format created from the distance matrix by any standard phylogenetic tree reconstruction method, e.g. neighbor-joining

(Saitou and Nei 1987; Studier and Kepler 1988), BioNJ (Gascuel 1997), FastME (Desper and Gascuel 2002). Although computing the REQ value for every branch of a phylogenetic tree on n leaves requires $O(n5)$ time complexity, REQ running time is quite fast (e.g. ~5 seconds with $n = 500$ on a standard computer) and could therefore be used with large phylogenetic trees.

References

Anisimova, M, and O Gascuel. 2006. "Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative." *Systematic Biology* 55 (4):860–921. https://doi.org/10.1080/10635150600755453.

Anisimova, M, M Gil, J-F Dufayard, C Dessimoz, and O Gascuel. 2011. "Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-Based Approximation Schemes." *Systematic Biology* 60 (5):685–99. https://doi.org/10.1093/sysbio/syr041.

Bremer, K. 1988. "The Limits of Amino Acid Sequence Data in Angiosperm Phylogenetic Reconstruction." *Evolution* 42 (4):795–803. https://doi.org/10.1111/j.1558-5646.1988.tb02497.x.

———. 1994. "Branch Support and Tree Stability." *Cladistics* 10 (3):295–304. https://doi.org/10.1111/j.1096-0031.1994.tb00179.x.

Buneman, P. 1971. "The Recovery of Trees from Measures of Dissimilarity." In *Mathematics in Archaeological and Historical Sciences*, edited by FR Hodson, DG Kendall, and P Tautu, 387–95. Edimburgh: Edinburgh University Press. http://homepages.inf.ed.ac.uk/opb/homepagefiles/phylogeny-scans/manuscripts.pdf.

Chapus, C, C Dufraigne, S Edwards, A Giron, B Fertil, and PJ Deschavanne. 2005. "Exploration of Phylogenetic Data Using a Global Sequence Analysis Method." *BMC Evolutionary Biology* 5:63. https://doi.org/10.1186/1471-2148-5-63.

Cohen, E, and B Chor. 2012. "Detecting Phylogenetic Signals in Eukaryotic Whole Genome Sequences." *Journal of Computational Biology* 19 (8):945–56. https://doi.org/10.1089/cmb.2012.0122.

Desper, R, and O Gascuel. 2002. "Fast and Accurate Phylogeny Reconstruction Algorithms Based on the Minimum-Evolution Principle." *Journal of Computational Biology* 19 (5):687–705. https://doi.org/10.1089/106652702761034136.

Felsenstein, J. 1985. "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." *Evolution* 39 (4):783–91. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

Garcia-Hermoso, D, A Criscuolo, SC Lee, M Legrand, M Chaouat, B Denis, M Lafaurie, et al. 2018. "Outbreak of Invasive Wound Mucormycosis in a Burn Unit Due to Multiple Strains of Mucor Circinelloides F. Circinelloides Resolved by Whole-Genome Sequencing." *MBio* 9 (2):e00573–18. https://doi.org/10.1128/mBio.00573-18.

Gascuel, O. 1997. "BIONJ: An Improved Version of the Nj Algorithm Based on a Simple Model of Sequence Data." *Molecular Biology and Evolution* 14 (7):685–95. https://doi.org/10.1093/oxfordjournals.molbev.a025808.

Guénoche, A, and H Garreta. 2001. "Can We Have Confidence in a Tree Representation?" In *Computational Biology*, edited by O Gascuel and MF Sagot, 45–56. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45727-5_5.

Henz, SR, DH Huson, AF Auch, K Nieselt-Struwe, and SC Schuster. 2005. "Whole-Genome Prokaryotic Phylogeny." *Bioinformatics* 21 (10):2329–35. https://doi.org/10.1093/bioinformatics/bth324.

Hoang, DT, O Chernomor, A von Haeseler, BQ Minh, and LS Vinh. 2018. "UFBoot2: Improving the Ultrafast Bootstrap Approximation." *Molecular Biology and Evolution* 35 (2):518–22. https://doi.org/10.1093/molbev/msx281.

Hoang, DY, LS Vinh, T Flouri, A Stamatakis, A von Haeseler, and BQ Minh. 2018. "MPBoot: Fast Phylogenetic Maximum Parsimony Tree Inference and Bootstrap Approximation." *BMC Evolutionary Biology* 18 (1):11. https://doi.org/10.1186/s12862-018-1131-3.

House, CH, M Pellegrini, and ST Fitz-Gibbon. 2014. "Genome-Wide Gene Order Distances Support Clustering the Gram-Positive Bacteria." *Frontiers in Microbiology* 5:785. https://doi.org/10.3389/fmicb.2014.00785.

Krajewski, C, and AW Dickerman. 1990. "Bootstrap Analysis of Phylogenetic Trees Derived from Dna Hydridization Distances." *Systematic Zoology* 39 (4):383–90. https://doi.org/10.2307/2992358.

Lemoine, F, J-B Domelevo-Entfellner, E Wilkinson, D Correia, M Davila Felipe, T De Oliveira, and O Gascuel. 2018. "Renewing Felsenstein's Phylogenetic Bootstrap in the Era of Big Data." *Nature* 556 (7702):452–56. https://doi.org/10.1038/s41586-018-0043-0.

Makarenkov, V, A Boc, J Xie, P Peres-Neto, F-J Lapointe, and P Legendre. 2010. "Weighted Bootstrapping: A Correction Method for Assessing the Robustness of Phylogenetic Trees." *BMC Evolutionary Biology* 10:250. https://doi.org/10.1186/1471-2148-10-250.

Minh, BQ, MA Nguyen, and A von Haeseler. 2013. "Ultrafast Approximation for Phylogenetic Bootstrap." *Molecular Biology and Evolution* 30 (5):1188–95. https://doi.org/10.1093/molbev/mst024.

Pardi, F, and O Gascuel. 2016. "Distance-Based Methods in Phylogenetics." In *Encyclopedia of Evolutionary Biology*, edited by Kliman R, 458–65. Academic Press. https://hal-lirmm.ccsd.cnrs.fr/lirmm-01386569.

Saitou, N, and M Nei. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution* 4 (4):406–25. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Spencer, M, D Bryant, and E Susko. 2007. "Conditioned Genome Reconstruction: How to Avoid Choosing the Conditioning Genome." *Systematic Biology* 56 (1):25–43. https://doi.org/10.1080/10635150601156313.

Studier, JA, and KJ Kepler. 1988. "A Note on the Neighbour-Joining Method of Saitou and Nei." *Molecular Biology and Evolution* 5 (1):729–31. https://doi.org/10.1080/10635150601156313.

Wang, LS, T Warnow, BM Moret, RK Jansen, and LA Raubeson. 2006. "Distance-Based Genome Rearrangement Phylogeny." *Journal of Molecular Evolution* 63 (4):473–83. https://doi.org/10.1007/s00239-005-0216-y.

Zaretskii, ZA. 1965. "Constructing a Tree on the Basis of a Set of Distances Between the Hanging Vertices [in Russian]." *Uspekhi Matematicheskikh Nauk* 20 (6):94–96. http://mi.mathnet.ru/eng/umn6134.