

Random tree-puzzle leads to the yule-harding distribution

Vinh L.S., Fuehrer A., Von Haeseler A.

Computer Science Department, University of Engineering and Technology, Vietnam National University
Hanoi, Cau Giay, Hanoi, Viet Nam; Center for Integrative Bioinformatics Vienna, Max F. Perutz
Laboratories, University of Vienna, Vienna, Austria

Abstract: Approaches to reconstruct phylogenies abound and are widely used in the study of molecular evolution. Partially through extensive simulations, we are beginning to understand the potential pitfalls as well as the advantages of different methods. However, little work has been done on possible biases introduced by the methods if the input data are random and do not carry any phylogenetic signal. Although Tree-Puzzle (Strimmer K, von Haeseler A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol.* 13:964-969; Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502-504) has become common in phylogenetics, the resulting distribution of labeled unrooted bifurcating trees when data do not carry any phylogenetic signal has not been investigated. Our note shows that the distribution converges to the well-known Yule-Harding distribution. However, the bias of the Yule-Harding distribution will be diminished by a tiny amount of phylogenetic information. maximum likelihood, phylogenetic reconstruction, Tree-Puzzle, tree distribution, Yule-Harding distribution. ?? 2010 The Author.

Author Keywords: maximumlikelihood; phylogenetic reconstruction; tree distribution; Tree-Puzzle; Yule-Harding distribution

Year: 2011

Source title: Molecular Biology and Evolution

Volume: 28

Issue: 2

Page : 873-877

Link: Scopus Link

Correspondence Address: Von Haeseler, A.; Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria; email: arndt.von.haeseler@univie.ac.at

ISSN: 7374038

CODEN: MBEVE

DOI: 10.1093/molbev/msq212

Language of Original Document: English

Abbreviated Source Title: Molecular Biology and Evolution

Document Type: Article

Source: Scopus

Authors with affiliations:

1. Vinh, L.S., Computer Science Department, University of Engineering and Technology, Vietnam National University Hanoi, Cau Giay, Hanoi, Viet Nam
2. Fuehrer, A., Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria
3. Von Haeseler, A., Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria

References:

1. Blum, M.G.B., François, O., Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance (2006) *Syst Biol*, 55, pp. 685-691
2. Brandley, M.C., Leache, A.D., Warren, D.L., McGuire, J.A., Are unequal clade priors problematic for Bayesian phylogenetics (2006) *Syst Biol*, 55, pp. 138-146
3. Dobson, A.J., Unrooted trees in numerical taxonomy (1974) *J Appl Probab*, 11, pp. 32-42
4. Felsenstein, J., (2003) *Inferring Phylogenies*, Sunderland (MA): Sinauer Associates
5. Harding, E.F., The probabilities of rooted tree-shapes generated by random bifurcation (1971) *Adv Appl Probab*, 3, pp. 44-77
6. Jaynes, E.T., (2003) *Probability Theory: The Logic of Science*, Cambridge (UK): Cambridge University Press
7. Kupczok, A., Consequences of different null models on the tree shape bias of supertree methods (2009) *Syst Biol*, Submitted
8. Mau, B., Newton, M.A., Larget, B., Bayesian phylogenetic inference via Markov chain Monte Carlo methods (1999) *Biometrics*, 55 (1), pp. 1-12
9. McKenzie, A., Steel, M., Distributions of cherries for two models of trees (2000) *Math Biosci*, 164, pp. 81-92
10. Nei, M., Sudhir, K., (2000) *Molecular Evolution and Phylogenetics*, New York: New York University Press
11. Pickett, K.M., Randle, C.P., Strange Bayes indeed: Uniform topological priors imply non-uniform clade priors (2005) *Mol Phylogenet Evol*, 34, pp. 203-211
12. Rambaut, A., Grassly, N.C., Seq-gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees (1997) *Computer Applications in the Biosciences*, 13 (3), pp. 235-238
13. Rannala, B., Yang, Z., Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference (1996) *J Mol Evol*, 43, pp. 304-311
14. Salemi, M., Vandamme, A.M., (2003) *The Phylogenetics Handbook: A Practical Approach to DNA and Protein Phylogeny*, Cambridge (UK): Cambridge University Press
15. Schmidt, H.A., Strimmer, K., Vingron, M., Von Haeseler, A., TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing (2002) *Bioinformatics*, 18, pp. 502-504
16. Semple, C., Steel, M., *Phylogenetics* (2003) Oxford Lecture Series in Mathematics and Its Applications, 24, Oxford: Oxford University Press
17. Steel, M., McKenzie, A., Properties of phylogenetic trees generated by yule-type speciation models (2001) *Math Biosci*, 170, pp. 91-112
18. Strimmer, K., Von Haeseler, A., Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies (1996) *Mol Biol Evol*, 13, pp. 964-969
19. Wilkinson, M., Cotton, J.A., Creevey, C., Eulenstein, O., Harris, S.R., Lapointe, F.J., Levasseur, C., Thorley, J.L., The shape of supertrees to come: Tree shape related properties of fourteen supertree methods (2005) *Syst Biol*, 54, pp. 419-431
20. Yang, Z., (2006) *Computational Molecular Evolution*, Oxford: Oxford University Press
21. Yule, G., A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis (1924) *Philos Trans R Soc London Ser B*, 213, pp. 21-87

