Running head: COMPARING PHYLOGENETIC METHODS

Bayesian and parsimony approaches reconstruct equally informative trees from simulated morphological datasets

Martin R. Smith

*Department of Earth Sciences, Lower Mount Joy, Durham University, Durham, DH1 3LE, UK; martin.smith@durham.ac.uk*

Tree topologies are a primary output of phylogenetic analysis. As such, it is important to select a method of tree reconstruction that produces useful and instructive tree topologies. Morphological systematists generally advocate either parsimony methods, using either equal or implied weighting, or Bayesian methods, which employ an explicit probabilistic model of evolution. The performance of these methods has been evaluated by simulating morphological datasets from a known tree topology, and calculating how accurately each method reconstructs the generative tree.

This focus on accuracy risks neglecting another aspect in which trees can be informative: precision. Considering both these aspects of information leads to quite different recommendations for phylogenetic practice. Parsimony methods with implied weighting can rival the performance of Bayesian approaches, whereas trees identified using ‘equal weights’ are less informative and more likely to be misleading. Summary or consensus trees generated by any method can usually be made informative be collapsing poorly supported nodes.

*Keywords*.—phylogenetic inference; parsimony analysis; equal weights; implied weights; Bayesian phylogenetic methods; information content

Evolutionary history can be reconstructed from morphological data using parsimony-based or probabilistic approaches. It is difficult to determine which approach better reconstructs phylogenetic relationships, which are typically unknown. A pragmatic approach to this question asks which method can best reconstruct a known tree from which a dataset has been artificially simulated. Such studies generally find probabilistic approaches to be most accurate, followed by parsimony under equal and implied weights in turn [1–6] – though it is difficult to demonstrate that simulated data realistically replicate morphological data [7,8].

Accuracy alone, however, is not the only goal when reconstructing trees: no tree is more accurate than a single polytomy, for a total absence of relationship information guarantees that no relationship is incorrectly resolved. An emphasis on accuracy therefore disadvantages methods that produce highly-resolved trees [9]. Resolution (= precision) is another aspect of a tree’s information content, which can be bartered for greater accuracy [10]. We should favour methods that recover as much information as possible about the true tree, striking a balance between accuracy and precision.

The quartet dissimilarity [11,12] and normalised symmetric difference [13] metrics allow information-based comparisons between a generated tree and a reference tree. The former considers whether each possible quartet (a set of four taxa) is resolved the same way (*s*) or a different way (*d*) on each tree; resolved in tree 1 only (*r*1); resolved in tree 2 only (*r*2); or unresolved in both trees (*u*) [12,14]. The latter replaces quartets with edges (i.e. bipartitions or splits) as its unit of measurement. Dissimilarity is given by (2*d* + *r*1 + *r*2) / 2*E*, where *E* is the total number of quartets or edges. By measuring the number of quartet or edge modifications taken to change from one tree into the other, this metric integrates both accuracy and precision into a single information-based measure akin to the Kullback-Leibler divergence [15] – though as neither quartets nor edges are independent, the metric does not satisfy the statistical properties of Shannon-Weiner information.

The normalised symmetric difference is subject to two major shortcomings. Firstly, its resolution is restricted by the small range of possible values that it can take. Secondly, its value is sometimes counterintuitive: moving a single tip to a particular location can generate a higher distance metric than moving both that tip and its immediate neighbour to the same point; and its maximum value can be reached in a pair of trees that only differ in the position of a single tip [11]. These issues do not affect the quartet dissimilarity measure, which therefore produces more useful results.

I have used quartet dissimilarity to determine whether the conclusions of two simulation studies change when precision is considered alongside accuracy. Congreve and Lamsdell [5; CL hereafter] simulated 55-character matrices from a bifurcating 22-tip tree using a Markov *k­*-state 1 parameter model with a gamma parameter. O’Reilly et al. [2; OR hereafter] simulated character matrices containing 100, 350 and 1000 characters from a bifurcating 75-tip tree using a modified HKY85 model. I used TNT v1.1 [16] to conduct parsimony searches on each of these matrices under equal and implied weights, using the parsimony ratchet and sectorial search heuristics (search options: xmult:hits 20 level 4 chklevel 5 rat10 drift10). I took a strict consensus of all optimal trees obtained under equal weights, and under implied weights [17] at the concavity constants used in each respective study (CL: *k* = 1, 2, 3, 5 and 10; OR: *k* = 2, 3, 5, 10, 20 and 200). For each dataset I generated a further strict consensus of all trees that were optimal under any of the concavity constants, excluding the unreasonable value of *k* = 1.

I also generated majority-rule consensus trees in MrBayes 3.2.2 [18] using an Mk model, with rates distributed according to a gamma parameter. I combined results from four runs of four chains, sampling every 10 000 generations for 10 000 000 generations, then discarding the first 40% of samples as burn-in (topology parameter: 0.999 < PSRF < 1.001; ESS > 400). Scripts are provided in the Supplementary Information.

To explore the barter between precision and accuracy, I generated 20 further trees for each dataset and each analytical configuration by progressively lowering the resolution of the most resolved tree. Under the Mk model, I collapsed clades with a posterior probability of < *x*%, with *x* varying uniformly from 50–100. In parsimony analyses, strict consensus trees were produced from forests of trees that were suboptimal by up to *x* steps (TNT command subopt *x*; bbreak;), with *x* corresponding to the integers 1..20 for equal weights, and drawn from a logarithmic distribution (0.730...19, 2.5×10−3→1×100) for implied weights.

Quartet dissimilarity was calculated using the new R package SlowQuartet, using explicit enumeration where trees contain polytomies, and the faster tqDist algorithm [19] where they do not. For comparison with previous results, the normalized symmetric distance was also calculated, using phangorn [20]. To summarise results, *s*, *d*, and *r2* were calculated for each individual tree relative to the relevant generative tree (bifurcating; thus *r*1 = *u* = 0), and the mean of each of parameter was calculated for each analysis at each value of *x*.

To avoid autocorrelation when graphing results [as encountered in ,2–4], I plotted the proportion of quartets or partitions that are the same in both trees (*s*), different in both trees (*d*), and unresolved in at least one tree (*r­*1 + *r*2 + *u*) on ternary plots using the Ternary package [21], oriented such that the *y* direction corresponds to tree dissimilarity, and the *x* direction to precision.

With the Congreve and Lamsdell datasets, there is no significant difference (at *p* = 0.01) between the quartet dissimilarity of the best trees generated by the *Mk* model or implied weights (*k* ∈ {2, 3, 5, 10}), but the best trees generated by equal weights, or implied weights with *k* = 1, are significantly worse than those produced by the other methods (Figure 1a).

The high precision of implied weights can be bartered for increased accuracy by collapsing the least-supported nodes [per ,7,22], leading to a trivial increase in the overall informativeness of the tree. After a point, the gain in accuracy no longer offsets the information lost, and collapsing further nodes increases the dissimilarity between the tree and the reference tree. Because equal weights and Bayesian approaches result in less precise trees, their most resolved trees cannot generally be improved by collapsing nodes.

These results hold even if the (problematic) symmetric difference metric is employed (Figure 1b), though once relatively more nodes must be collapsed before trees reach their maximal information content by this metric – indicating that the edge-based measure emphasizes accuracy over precision. The results do not meaningfully change when datasets with a low consistency index are excluded.

Similar results are observed in the O’Reilly *et al.* datasets (Figure 2): at any given level of precision, the best trees generated by the *Mk* model are similar in accuracy to those generated under implied weights (except with very small values of *k*), but are more accurate than those generated using equal weights.

In the simulation studies analysed herein, the results of parsimony analysis are neither more nor less useful than those of probabilistic analyses in reconstructing evolutionary history, particularly if the precision of parsimony trees is reduced to that of the 50% posterior probability tree obtained from the same dataset by Bayesian analysis. The caveat is that parsimony analysis must employ a sensible weighting scheme; implied weighting should be employed with a moderate concavity (*k*) value. Extremely low values (*k* < 3), or extremely high values (i.e. equally weighted parsimony, which is mathematically equivalent to implied weights with an infinite concavity constant) are likely to yield results that are less informative about evolutionary history, with this lack of information representing a combination of imprecision and inaccuracy; results arising under such values do not merit biological interpretation.

Previous results favouring Bayesian methods over parsimony [2], or equal weights over implied weights [5], have arisen because precision has been overlooked as a component of a tree’s utility; future simulation studies should employ metrics that consider both aspects of tree similarity, rather than accuracy alone. As (at least in the analyses examined herein) neither Bayesian nor parsimony analyses generate consistently superior results, researchers may wish to explicitly compare the results of both methods when conducting phylogenetic analysis; observations common to both methods are particularly likely to be well supported by underlying data.

# Figure legends

**Figure 1.** **Status of quartets and nodes in trees recovered from Congreve and Lamsdell datasets.** Points denote the average number of quartets (a) or partitions (b, c) that are the same as the generative tree (*Same*), resolved differently to the generative tree (*Different*), or not resolved (*Unresolved*). Each series indicates the effect of progressively collapsing the least-supported nodes in trees generated by analysis of datasets simulated by Congreve and Lamsdell (2016) under the specified analytical parameters. An increase in the *x* direction corresponds to a decrease in quartet dissimilarity or normalized symmetric difference (i.e. more informative trees); an increase in the *y* direction corresponds to a decrease in precision (less resolved trees).

**Figure 2. Status of quartets in trees recovered from O’Reilly *et al.* datasets.** Columns correspond to datasets with 100, 350 or 1000 characters simulated by O’Reilly *et al.* (2016).

# References

1 Wright, A. M. & Hillis, D. M. 2014 Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One* **9**, e109210. (doi:10.1371/journal.pone.0109210)

2 O’Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., Pisani, D. & Donoghue, P. C. J. 2016 Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biol. Lett.* **12**, 20160081. (doi:10.1098/rsbl.2016.0081)

3 Puttick, M. N., O’Reilly, J. E., Pisani, D. & Donoghue, P. C. J. 2018 Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* (doi:10.1111/pala.12388)

4 Puttick, M. N. et al. 2017 Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B* **284**, 20162290. (doi:10.1098/rspb.2016.2290)

5 Congreve, C. R. & Lamsdell, J. C. 2016 Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices. *Palaeontology* **59**, 447–462. (doi:10.1111/pala.12236)

6 O’Reilly, J. E., Puttick, M. N., Pisani, D. & Donoghue, P. C. J. 2018 Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* **61**, 105–118. (doi:10.1111/pala.12330)

7 Goloboff, P. A., Torres, A. & Arias, J. S. 2018 Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* **34**, 407–437. (doi:10.1111/cla.12205)

8 Goloboff, P. A., Torres Galvis, A. & Arias, J. S. 2018 Parsimony and model-based phylogenetic methods for morphological data: comments on O’Reilly *et al*. *Palaeontology* **61**, 625–630. (doi:10.1111/pala.12353)

9 Brown, J. W., Parins-Fukuchi, C., Stull, G. W., Vargas, O. M. & Smith, S. A. 2017 Bayesian and likelihood phylogenetic reconstructions of morphological traits are not discordant when taking uncertainty into consideration: a comment on Puttick *et al*. *Proc. R. Soc. B* **284**, 20170986. (doi:10.1098/rspb.2017.0986)

10 Mackay, D. M. 1953 Quantal aspects of scientific information. *IEEE Trans. Inf. Theory* **1**, 60–80. (doi:10.1109/TIT.1953.1188569)

11 Steel, M. A. & Penny, D. 1993 Distributions of tree comparison metrics—some new results. *Syst. Biol.* **42**, 126–141. (doi:10.1093/sysbio/42.2.126)

12 Estabrook, G. F., McMorris, F. R. & Meacham, C. A. 1985 Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Syst. Zool.* **34**, 193–200. (doi:10.2307/sysbio/34.2.193)

13 Penny, D. & Hendy, M. 1985 The use of tree comparison metrics. *Syst. Zool.* **34**, 75–82. (doi:10.2307/2413347)

14 Bandelt, H. J. & Dress, A. 1986 Reconstructing the shape of a tree from observed dissimilarity data. *Adv. Appl. Math.* **7**, 309–343. (doi:10.1016/0196-8858(86)90038-2)

15 Kullback, S. & Leibler, R. A. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86. (doi:10.1214/aoms/1177729694)

16 Goloboff, P. A., Farris, J. S. & Nixon, K. C. 2008 TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786. (doi:10.1111/j.1096-0031.2008.00217.x)

17 Goloboff, P. A. 1993 Estimating character weights during tree search. *Cladistics* **9**, 83–91. (doi:10.1111/j.1096-0031.1993.tb00209.x)

18 Huelsenbeck, J. P. & Ronquist, F. 2001 MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755. (doi:10.1093/bioinformatics/17.8.754)

19 Sand, A., Holt, M. K., Johansen, J., Brodal, G. S., Mailund, T. & Pedersen, C. N. S. 2014 tqDist: a library for computing the quartet and triplet distances between binary or general trees. *Bioinformatics* **30**, 2079–2080. (doi:10.1093/bioinformatics/btu157)

20 Schliep, K. P. 2011 phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593. (doi:10.1093/bioinformatics/btq706)

21 Smith, M. R. 2017 Ternary: an R package to generate ternary plots. *The Comprehensive R Archive Network*, doi:10.5281/zenodo.1068997.

22 Goloboff, P. A. & Szumik, C. A. 2015 Identifying unstable taxa: efficient implementation of triplet-based measures of stability, and comparison with Phyutility and RogueNaRok. *Mol. Phylo. Evol.* **88**, 93–104. (doi:10.1016/j.ympev.2015.04.003)

# Acknowledgements

The TNT software is supported by the Willi Hennig Society.

# Author contributions

MS is the sole author.

# Data accessibility

The SlowQuartet package will be submitted to the CRAN repository when the review process is complete. Its vignettes and provide detailed examples of situations where existing tree distance measures produce undesirable results, and depict analytical results for each individual Congreve & Lamsdell dataset.

Reviewers can access the SlowQuartet package at https://github.com/ms609/SlowQuartet

Supplementary data files have been uploaded to FigShare, 10.6084/m9.figshare.5659195

Reviewers can access this repository at: https://figshare.com/s/46744779f750495e527b

# Funding

None to report.

# Competing interests

None to report.

# Ethical statement

No ethical approval was required to conduct this research.