

# An Experimental Study of Quartets MaxCut and Other Supertree Methods

M. Shel Swenson<sup>1,2</sup>, Rahul Suri<sup>1</sup>, C. Randal Linder<sup>3</sup>, and Tandy Warnow<sup>1</sup>

<sup>1</sup> Department of Computer Science, The University of Texas at Austin

<sup>2</sup> Department of Mathematics, The University of Texas at Austin

<sup>3</sup> Section of Integrative Biology, The University of Texas at Austin

**Abstract.** Although many supertree methods have been developed in the last few decades, none has been shown to produce more accurate trees than the popular Matrix Representation with Parsimony (MRP) method. In this paper, we evaluate the performance of several supertree methods based upon the Quartets MaxCut method of Snir and Rao. We show that two of these methods usually outperform MRP and all other supertree methods we studied under many realistic model conditions. In addition, we show that the popular criterion of minimizing the total topological distance to the source trees is only weakly correlated with topological accuracy, and therefore that evaluating supertree methods on biological datasets is problematic.

## 1 Introduction

Supertree methods comprise one approach to reconstructing large molecular phylogenies given a set (called a *profile*) of estimated trees (called *source trees*) for overlapping subsets of the entire set of taxa. Source trees are combined into a single supertree on the full set of taxa using various algorithmic techniques. Because of the computational difficulties in estimating large phylogenies, many computational biologists think that the only feasible strategy to estimating the Tree of Life will involve a divide-and-conquer approach where trees are estimated on subsets of taxa and a supertree method is used to assemble a tree on the entire taxon set from the source trees. While there are many supertree methods, only MRP is used regularly in supertree constructions on biological datasets (4); furthermore, no other supertree method has been shown to produce trees that are comparable in accuracy to MRP under the standard bipartition metric (5).

One version of the supertree estimation problem uses quartet amalgamation methods. Each estimated source tree is encoded by an *appropriately chosen subset* of its induced quartet trees, and the set of quartets (the union of the chosen subsets for each source tree) is used to estimate a supertree. Quartet amalgamation methods can thus be used to assemble supertrees from source trees of arbitrary size.

The Maximum Quartet Consistency (MQC) problem is a natural optimization problem, in which the input is a set of quartet trees and a supertree is sought that *displays* the maximum number of quartet trees. MQC is NP-hard, and generally hard to approximate except in special cases (3; 15; 6; 11). Theoretical results and heuristics for the special case where the input set contains a tree on every quartet appear in (24; 20; 14; 16; 22). In a recent paper (21), Snir and Rao presented *Quartets MaxCut* (QMC), a heuristic for MQC that can be applied to arbitrary sets of quartet trees (i.e., ones that may not contain a tree on every quartet). Snir and Rao showed that by encoding the source trees as quartet trees, QMC could be used as a generic supertree method. They then constructed supertrees using this QMC-based supertree method for a number of biological supertree profiles. Since the true supertree was not known, they could not evaluate the topological accuracy of the supertrees they constructed; instead, they compared the QMC supertree to the source trees to produce two different average similarity measures for each supertree. A comparison between QMC-based supertrees and MRP supertrees showed that QMC had higher average similarity to the source trees under one criterion, and lower average similarity with respect to another. QMC's failure to outperform MRP as a supertree method with respect to the average similarity to the source trees should not be considered a serious limitation for two reasons. First, average similarity to the source trees is not the same as accuracy with respect to the true tree (a phenomenon we investigate directly in this paper). Second, QMC depends critically upon the specific technique used to encode each source tree as a set of quartet trees. In other words, QMC might be producing highly accurate trees even though the average similarity is lower than MRP, and it might produce more accurate trees if other encodings of the source trees were used.

In this paper, we report results from a study in which we employ several encodings of the source trees by quartet trees and apply QMC to the resultant sets of quartet trees. We compare the accuracy of QMC using different encodings to MRP and five other supertree methods: Robinson-Foulds Supertrees (1), Q-Imputation (13), MinFlip (8; 7; 9), SFIT (10), and PhySIC (19). We find that the topological accuracy of QMC supertrees computed on different encodings varies substantially. Two QMC-based supertree methods, QMC(All) and QMC(Exp+TSQ) (differing only in how the source trees are encoded), perform similarly and *outperform* all the other supertree methods under many realistic model conditions, and have comparable accuracy under most others. However, MRP outperforms all QMC methods on the largest (1000-taxon) datasets. Finally, we find that using topological similarity to source trees as a proxy for topological accuracy with respect to the true tree is of limited use, and can be misleading. Thus, evaluating supertree methods on biological datasets is prob-

lematic, and supertree methods that seek to minimize topological distance to source trees may not have the best accuracy.

## 2 Basics

*Supertree Datasets* Because of the taxon sampling strategies used by biologists, source trees tend to be focused either on intensively sampled, smaller subgroups, like big cats, or on larger, sparsely sampled groups, like all vertebrates. The first type is called a *clade* source tree, and the second type is called a *scaffold*. Supertree profiles include scaffolds to ensure sufficient overlap among the clade trees.

The input to the supertree problem is a set of source trees,  $\{t_1, t_2, \dots, t_k\}$ , on subsets of a set  $S$  of taxa. Source trees are often estimated using biomolecular sequence datasets. Each source tree is estimated on its aligned sequence dataset using computationally intensive methods—e.g., maximum parsimony or maximum likelihood heuristics like RAxML (23). A supertree method combines the source trees into a tree on the full dataset.

*Matrix Representation with Parsimony* Matrix representation with parsimony (MRP) (2; 18) is currently the most widely used supertree method. It encodes source trees as a matrix of *partial binary characters*: all entries in the matrix are 0, 1, or ?, with each column in the matrix defined by a single edge in a source tree. The matrix is then analyzed using a heuristic for the NP-hard maximum parsimony problem (12).

*Quartets MaxCut (QMC)* QMC is a quartet amalgamation method, operating in polynomial time and providing no guarantees with respect to its optimization problem, MQC. The source trees are encoded by sets of quartet trees, and QMC is applied to the union of these sets.

*Quartet Encodings of Source Trees* Here, we explore several techniques for representing source trees by sets of quartet trees. Two of these techniques use random sampling strategies (21), which are based upon computation of the topological distance between leaves in the source tree. The *topological diameter* of a quartet tree  $q$  with respect to a source tree  $t$  is the maximum of its leaf-to-leaf topological distances within the source tree and is denoted  $diam_t(q)$ . The quartet encoding strategies used in (21) also include calculation of the *Topologically-Short Quartet* (TSQ) trees, defined as follows: For each edge in a source tree, pick the topologically nearest leaves in each of the subtrees around the edge. If two or more leaves within a subtree have the same topological distance to the edge, pick all such leaves. The set of quartet trees formed by picking one such

leaf from each subtree forms the TSQs around that edge. The union of all these is the set of TSQ trees.

We tested five strategies for encoding a source tree  $t$  by a set of quartet trees:

**All quartets:** include all induced four-taxon trees.

**k-short:** a generalization of the TSQs: for each edge in a source tree, pick the  $k$  topologically nearest leaves in each of the subtrees around that edge. The (approximately)  $k^4$  quartets of leaves are the  $k$ -short quartet trees around that edge, and the set of all such  $k$ -short quartet trees (unioning over all the internal edges) forms the set  $k$ -short. In this study, we let  $k = 5$  and  $k = 25$ .

**Geo+TSQ:** include a quartet  $q$  with probability  $d^{-3}$  where  $d = \text{diam}_t(q)$ , and add the TSQ trees (this was studied in (21)).

**Exp+TSQ:** compute the topological distance between every pair of leaves, include a quartet with probability  $1.5^{-d}$  where  $d = \text{diam}_t(q)$ , and add the TSQ trees (this was also studied in (21)).

### 3 Performance Study

We performed a study using simulated datasets to evaluate QMC-based supertree methods in comparison to MRP and other supertree methods. Simulations are used to evaluate phylogeny estimation methods, because the true tree is known exactly. For our simulations, we used the SMIDGen (25) methodology, and used datasets with 100, 500 and 1000 taxa. We used SMIDGen to produce supertree datasets of *mixed* source trees, consisting of one scaffold dataset (produced by a random selection of taxa from the entire dataset) and many clade-based datasets (focused, dense taxon sampling within a rooted subtree).

*Simulation Study Design:* For this study, we used simulated datasets generated for another study (25), and, therefore, describe the methodology only in brief. The simulated datasets are produced by simulating evolution under a GTR+Gamma+I process, down pure-birth model trees, deviated from a clock, and containing up to 1000 leaves. We generated 30 replicates for each 100- and 500-taxon model condition, and 10 replicates for each 1000-taxon model condition. Each model condition is indicated by the density of the scaffold dataset, which is the percentage of the entire taxon set in the scaffold dataset, with scaffold densities ranging from 20% to 100%. We used RAxML (23) to estimate phylogenetic trees. We performed the MP search in the MRP analyses, using a very effective heuristic search technique called *the Ratchet* (17), and computed a greedy consensus (gMRP) tree for the set of most parsimonious trees found during this search. We also computed supertrees based upon five ways of encoding the source trees as sets of quartet trees and then applying QMC, as

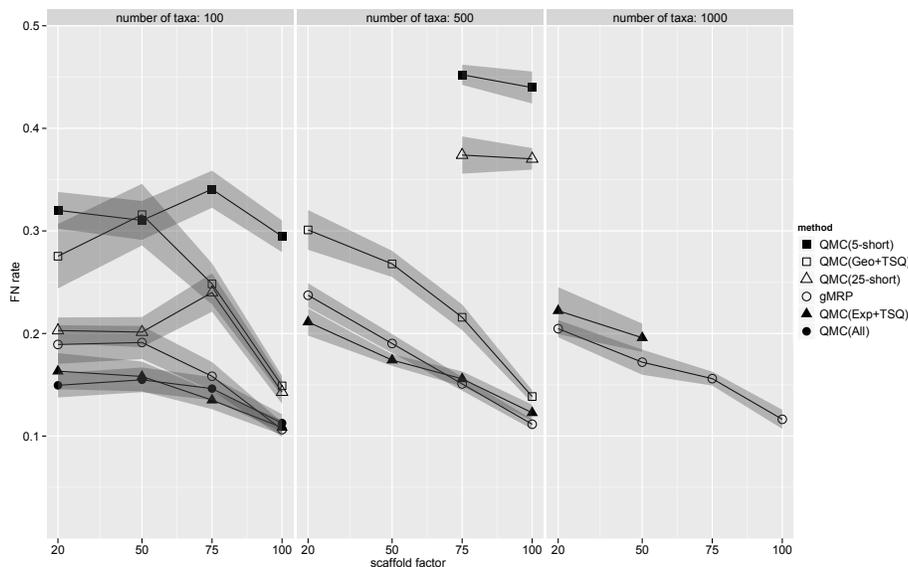
described above. Finally, we computed supertrees using several other methods, including Q-imputation (Q-Imp), Robinson-Foulds Supertrees (RFS), MinFlip, SFIT, and PhysIC, all in their default settings. For RFS, MinFlip, and PhysIC, methods that require rooted trees, we used mid-point rooting to root the source trees, a method commonly used to root unrooted trees and particularly appropriate because our source trees were not strongly deviated from ultrametricity. We computed three types of topological error rates for each estimated supertree when compared with the model tree: false positive rates, false negative rates, and Robinson-Foulds rates. We also computed the total topological distance of each supertree to the estimated source trees, using FN (false negative), FP (false positive) and RF (bipartition distance) errors modified so that we could handle trees on different taxon datasets. We restricted the supertree to the subset of taxa for the source tree, and then compute the topological distances between the two trees. We note that the bipartition distance, also known as the “Robinson-Foulds” (RF) distance, is the standard metric used in most studies. In our study, we show both FN and FP as well, thus providing a more nuanced description of error. Because QMC failed to return trees on some inputs, we restricted our results to datasets for which all the reported methods returned trees. This reduced the number of replicates for some model conditions. We also recorded the running time of each method on each dataset. Because the analyses were run under Condor (a distributed software environment (26)), running times (for the larger datasets, especially) are inexact and are larger than if they had been run on a dedicated processor. Running times are, therefore, an approximation of the time needed to perform these analyses.

## 4 Results

### 4.1 Exploring QMC under Different Source Tree Encodings

We compared the performance of the QMC variants and gMRP (Fig. 1). For a given model condition, we include only those methods that successfully completed on at least one third of the replicates, and display results for only those replicates on which every selected method successfully completed. We report performance with respect to FN rates, but the performance with respect to FP and RF rates is almost identical.

On the mixed 100-taxon datasets, QMC(All) and QMC(Exp+TSQ) were essentially tied as the best methods, followed by gMRP. Furthermore, QMC(All) and QMC(Exp+TSQ) had the greatest advantage over gMRP for the sparse scaffold cases. The other QMC variants had worse accuracy. On a large number of the 500- and 1000-taxon datasets, many of the QMC variants failed to complete, indicating that computational requirements can limit QMC’s utility. On



**Fig. 1.** Average topological error (False Negative (FN) rates) with standard error regions on *mixed* source-tree datasets. We use shaded regions in place of standard error bars as it better demonstrates overlap; however, the shading between data points for a method is not intended as an interpolation of error for scaffold factors not tested. Results are reported for the QMC variants and gMRP, as a function of the scaffold factor and by number of taxa. Points are graphed for a method if it had at least six datasets that completed in common with all other methods.

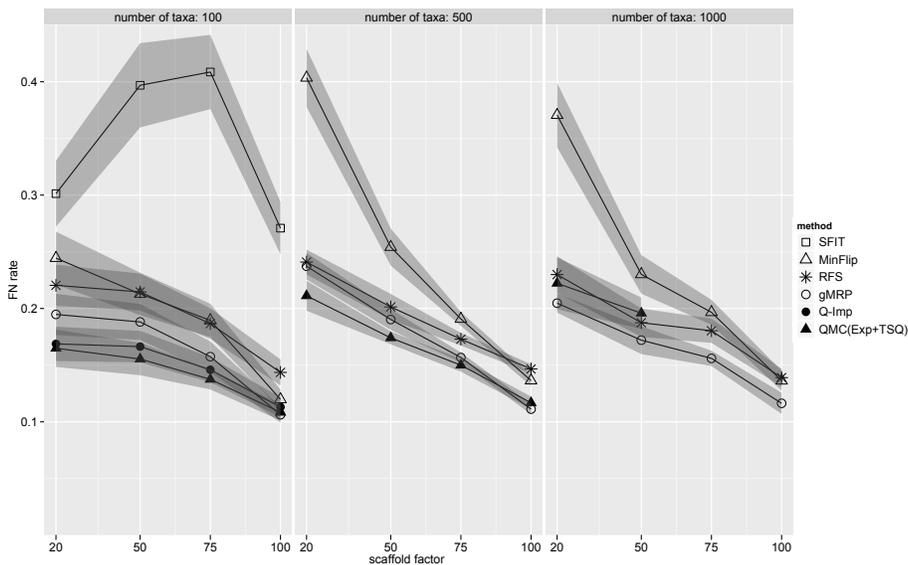
the 500-taxon datasets for which QMC(Exp+TSQ) could be run, it produced topologically more accurate trees than gMRP, giving the biggest advantage on the sparse scaffold datasets. For the 1000-taxon datasets, gMRP outperformed all the QMC variants that completed. However, most QMC variants failed to return trees on most inputs.

## 4.2 Comparing QMC(Exp+TSQ) to Other Supertree Methods

We compared QMC(Exp+TSQ) to six other supertree methods: gMRP, Q-Imp, SFIT, MinFlip, PhysIC, and Robinson-Foulds Supertrees (RFS).

All of these methods could be run on the 100-taxon datasets, but some failed to run on the larger datasets. For this reason, we obtained results for all seven methods on the 100-taxon datasets, but only five methods on the 500-taxon datasets (where SFIT and Q-Imp failed to run, due to computational limitations), and only four methods on the 1000-taxon datasets (where we did not try to run PhysIC, since it was computationally intensive for the 500-taxon datasets). In

addition, QMC(Exp+TSQ) failed to run on some datasets; we therefore only report results for those datasets on which all reported methods were able to run. PhySIC gives by far the worst results, producing completely unresolved trees except when the scaffold density is 100%, at which point it produces results that are still worse than the other methods. Because of it is not competitive with other methods, we omit PhySIC from our graphs.



**Fig. 2.** We report False Negative (FN) rates (means with standard error regions) for QMC(Exp+TSQ), gMRP, SFIT, MinFlip, RFS, and Q-Imp, as a function of the scaffold factor, for 100-, 500- and 1000-taxon model conditions.

The experiments show that three methods—QMC(Exp+TSQ), Q-Imp, and gMRP—generally outperform the remaining methods with respect to topological accuracy (Fig. 2). As with Fig. 1, in Fig. 2 we only include results for replicates for which all displayed methods were able to complete. Since Fig. 2 includes a different collection of methods, the results for a different collection of replicates are used. On the 100-taxon datasets, QMC(Exp+TSQ) and Q-Imp both gave higher accuracy than gMRP and all other methods (except on the 100% scaffold datasets, where they were equal to gMRP). On the 500-taxon datasets with sparse scaffolds, QMC(Exp+TSQ) performed better than all methods, with only a slight advantage over gMRP. On the 500-taxon datasets with dense (75%

and 100%) scaffolds, QMC(Exp+TSQ) and gMRP were the most accurate, and had essentially the same accuracy. On the 1000-taxon datasets, gMRP had an advantage over QMC(Exp+TSQ) and other methods, and QMC(Exp+TSQ) failed to run on the dense scaffold datasets (QMC fails to run on profiles with large source trees, due to computational reasons). The remaining methods—PhySIC, SFIT, MinFlip, and RFS—are generally less accurate than QMC(Exp+TSQ), QImp, and gMRP, and some (i.e., PhySIC and SFIT) cannot be run on large datasets. Interestingly, RFS outperforms QMC(Exp+TSQ) on the 1000 taxon datasets, where it matches the accuracy on the sparse scaffold datasets and (unlike QMC(Exp+TSQ)) is able to run on the dense scaffold datasets.

### 4.3 Evaluating Supertree Methods on Biological Datasets

For biological datasets, the true tree is not available, so evaluations of accuracy have tended to use average or total topological distance to the source trees (for example, (1; 21)). To test whether this is a good proxy for the quality of the supertree, we computed three distances for each supertree  $T$  to the profile  $\mathcal{T}$  of source trees:

- **SumFN** is defined as follows:  $\text{SumFN}(T, \mathcal{T}) = \frac{\sum_{t \in \mathcal{T}} (\text{FN}(T, t))}{M}$ , where  $\text{FN}(T, t)$  is the number of edges in  $t$  that do not appear in  $T$ , and  $M = \sum_{t \in \mathcal{T}} m_t$ , where  $m_t$  is the number of internal edges in  $t$ .
- **SumFP** and **SumRF** are defined similarly, with  $\text{FP}(T, t)$  and  $\text{RF}(T, t)$  replacing  $\text{FN}(T, t)$ , respectively. Here, FP denotes the false positive distance and RF denotes the Robinson-Foulds (“bipartition”) distance. Each distance is normalized to produce a value between 0 and 1. The false positive distance between a supertree  $T$  and a source tree  $t$  in the profile  $\mathcal{T}$  is the number of edges in  $T$  that do not appear in  $t$ , and the Robinson-Foulds distance is the total number of missing and false positive edges.

Note that if the supertree and all source trees are binary, then for each source tree  $t$ ,  $\text{RF}(T, t) = 2\text{FN}(T, t) = 2\text{FP}(T, t)$ , and after normalization all three distances are equal.

We examined how closely measurements of this sort are correlated to actual topological accuracy, that is, how closely SumFN, SumFP, or SumRF are correlated to the FN, FP or RF distance to the true tree. We found the correlations to be largely independent of the choice of topological distance to source trees (SumFN, SumFP, or SumRF) or topological error (FN, FP or RF). The reason for this was that the true supertree was fully resolved or nearly so, and all the computed supertrees were either fully resolved or nearly so. We therefore present results focusing on the correlation between SumFN (topological distance to the source trees) and FN (topological distance to the true tree).

To assess whether SumFN, SumFP or SumRF is a good optimality criterion, we calculated Spearman rank-correlations for each of the 100-taxon simulated datasets for the six supertree methods that consistently perform reasonably well (MinFlip, gMRP, Q-Imp, QMC(All), QMC(Exp+TSQ), and RFS). Correlations were calculated for each of these measures of distance to source trees and each of FN, FP and RF (calculated by comparing the supertree estimated by each of the methods with the true tree). The statistics were calculated this way to test whether the rank-order of the topological distances to source trees correlated strongly with the true rank-order of the supertrees, in terms of topological accuracy with respect to the true tree.

The results (Table 1) show clearly that attempting to optimize the total distance to the source trees is of limited use in producing accurate supertrees. None of the optimality criteria averaged better than 60% correlation with measures of true accuracy for a given scaffold factor, and for some datasets, the criteria were negatively correlated with the true quality of the supertrees that were estimated.

**Table 1.** Results of Spearman rank-order correlations of SumFN, SumFP, and SumRF with the true FN, FP, and RF measures of supertrees estimated using six supertree methods.

scaffold factor	optimality criterion	FN		FP		RF	
		mean	range	mean	range	mean	range
25	SumFN	0.401	-0.890, 0.939	0.376	-0.890, 0.926	0.391	-0.890, 0.926
	SumFP	0.421	-0.890, 0.939	0.421	-0.890, 0.926	0.426	-0.890, 0.926
	SumRF	0.406	-0.890, 0.939	0.395	-0.890, 0.926	0.406	-0.890, 0.926
50	SumFN	0.544	-0.203, 1.000	0.536	-0.348, 0.971	0.541	-0.203, 0.971
	SumFP	0.546	-0.143, 1.000	0.539	-0.257, 0.971	0.543	-0.143, 0.971
	SumRF	0.546	-0.143, 1.000	0.539	-0.257, 0.971	0.543	-0.143, 0.971
75	SumFN	0.593	-1.000, 0.986	0.589	-1.000, 0.986	0.591	-1.000, 0.986
	SumFP	0.593	-1.000, 0.986	0.589	-1.000, 0.986	0.591	-1.000, 0.986
	SumRF	0.593	-1.000, 0.986	0.589	-1.000, 0.986	0.591	-1.000, 0.986
100	SumFN	0.447	-0.789, 1.000	0.447	-0.789, 1.000	0.447	-0.789, 1.000
	SumFP	0.447	-0.789, 1.000	0.447	-0.789, 1.000	0.447	-0.789, 1.000
	SumRF	0.447	-0.789, 1.000	0.447	-0.789, 1.000	0.447	-0.789, 1.000

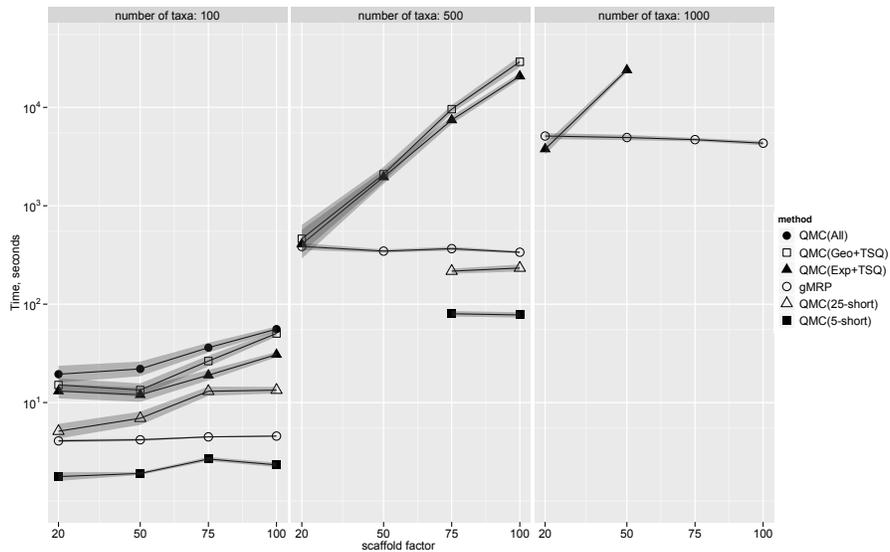
Thus, the correlation between topological distance to source trees and topological error (i.e., distance to the true tree) tends to be only weakly positive, so that while, in general, supertrees with smaller topological distance to the source trees are more accurate, there can be more accurate supertrees with higher topological distance to the source trees. These results suggest that the highest accuracy supertrees may not optimize SumFN (or any other topological distance to source trees).

This observation has two consequences for supertree analyses. First, directly trying to optimize the topological distance is not likely to produce the most accurate trees, since better trees are being produced through other means. Secondly, because the true tree is not known for biological supertree datasets, it is difficult to evaluate supertree methods using biological datasets.

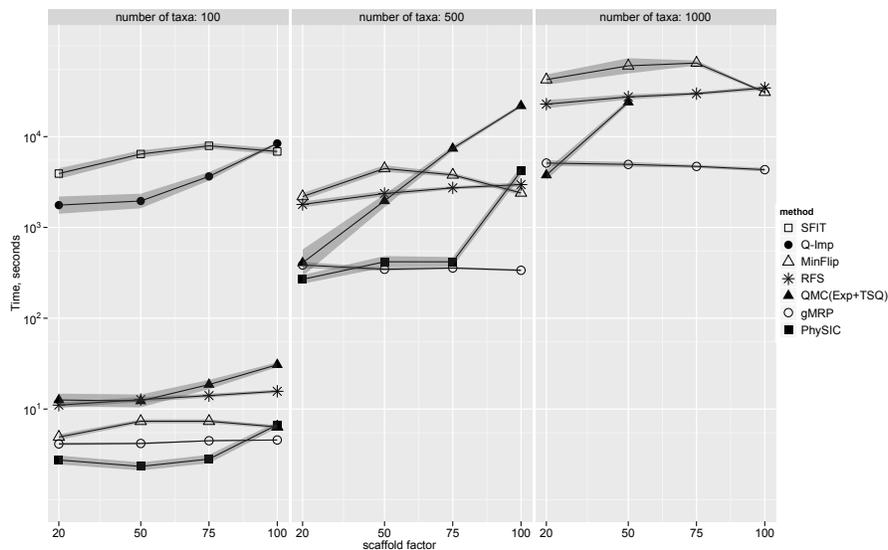
These conclusions are clearly based upon the conditions of this experiment, in which the source trees were reasonably, but not extremely, accurate. However, when source trees have no error at all, the true tree is guaranteed to minimize the distance to the source trees. Under this condition, MRP will also be guaranteed to return the true tree as one of the solutions. Thus, for very highly accurate source trees, both MRP and minimizing the total topological distance may be very good optimality criteria; the issue is how well supertree methods perform under more realistic conditions, where source trees have error.

#### 4.4 Scalability

We now discuss running time issues on simulated data. Fig. 3 gives the results for the QMC variants and gMRP, and Fig. 4 gives results for QMC(Exp+TSQ), gMRP, and the other (not QMC-based) supertree methods.



**Fig. 3.** Running times in seconds (means with standard error regions) of QMC supertree methods on mixed datasets; the y-axis is given with a logarithmic scale.



**Fig. 4.** Running times in seconds (means with standard error regions) of supertree methods on mixed datasets; the y-axis is given with a logarithmic scale.

Supertree methods on the simulated datasets showed some differences in running times. First, gMRP was faster than the accurate QMC variants for most of the model conditions, and the degree of improvement ranged from very small (a few seconds) to several hours. In general, we saw that profiles with large source trees were particularly difficult for QMC(Exp+TSQ) and QMC(All), and that for such datasets, gMRP had a running time advantage.

We note that the running times of QMC(Geo+TSQ), QMC(Exp+TSQ), and QMC(All) are directly impacted by the size of the source trees, since each four-tuple of taxa must be examined to produce the quartet trees. Thus, for large source trees, we expect these three methods to suffer computational limitations.

## 5 Conclusions

This study makes several important contributions. First, we show that while MRP is still the most accurate supertree method for the largest datasets, both QMC(Exp+TSQ) and Q-Imp produce more accurate supertrees than MRP and other supertree methods for the smaller (100- and 500-taxon) datasets. Therefore, an effort should be made to produce scalable and robust implementations of the quartet methods, QMC(Exp+TSQ) and Q-Imp. Each of these methods

produces, at some point, a quartet encoding of the source trees. Scalable implementations of these methods will require *not* using all the quartets in these encodings, as such approaches simply will fail on large datasets.

The second important contribution of this study is that the total topological distance to the source trees only provides limited information about topological accuracy, and that reliable comparisons can only be made between supertrees that have very different total topological distances. Consequently, previous studies that have explored performance of supertree methods using total topological distance to the source trees need to be revisited.

### **Acknowledgments**

We thank Sagi Snir for assistance with using the QMC code and for providing the software for generating the quartet encodings Exp+TSQ, Geo+TSQ, and AllQuartets. This work was supported by the US National Science Foundation ITR-0331453, for the CIPRES project.

## Bibliography

- [1] Bansal, M., Burleigh, J.G., Eulenstein, O., Fernández-Baca, D.: Robinson-foulds supertrees (2009)
- [2] Baum, B.R.: Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41, 3–10 (1992)
- [3] Ben-dor, A., Chor, B., Graur, D., Ophir, R., Pelleg, D.: Constructing phylogenies from quartets: Elucidation of eutherian superordinal relationships. *Journal of Computational Biology* 5(3), 377–390 (1998), earlier version appeared in RECOMB 1998
- [4] Bininda-Emonds, O.R.P.: The evolution of supertrees. *Trends in Ecology and Evolution* 19, 315–322 (2004)
- [5] Bininda-Emonds, O.R.P.: *Phylogenetic Supertrees: Combining Information To Reveal The Tree Of Life*. Computational Biology, Kluwer Academic, Dordrecht, the Netherlands (2004)
- [6] Bolaender, H., Fellows, M., Warnow, T.: Two strikes against perfect phylogeny. *Lecture Notes in Computer Science* 623, 273–283 (1992)
- [7] Burleigh, J.G., Eulenstein, O., Fernández-Baca, D., Sanderson, M.J.: MRF supertrees. In: Bininda-Emonds, O.R.P. (ed.) *Phylogenetic Supertrees: Combining Information To Reveal The Tree Of Life*. pp. 65–86. Kluwer Academic, Dordrecht, the Netherlands (2004)
- [8] Chen, D., Diao, L., Eulenstein, O., Fernández-Baca, D., Sanderson, M.J.: Flipping: a supertree construction method. In: *Bioconsensus. DIMACS: Series in Discrete Mathematics and Theoretical Computer Science*, vol. 61, pp. 135–160. American Mathematical Society-DIMACS, Providence, Rhode Island (2003)
- [9] Chen, D., Eulenstein, O., Fernández-Baca, D., Burleigh, J.G.: Improved heuristics for minimum-flip supertree construction. *Evol. Bioinform.* 2, 401–410 (2006)
- [10] Creevey, C.J., McInerney, J.O.: Clann: investigating phylogenetic information through supertree analyses. *Bioinformatics* 21(3), 390 – 392 (2005)
- [11] Dress, A., Steel, M.: Convex tree realizations of partitions. *Applied Mathematics Letters* 5(3), 3–6 (1992)
- [12] Foulds, L.R., Graham, R.L.: The steiner problem in phylogeny is NP-complete. *Adv. in Appl. Math.* 3(43-49), 299 (1982)
- [13] Holland, B., Conner, G., Huber, K., Moulton, V.: Imputing supertrees and supernetworks from quartets. *Syst. Biol.* 57(1), 299–308 (2007)
- [14] Jiang, T., Kearney, P., Li, M.: Orchestrating quartets: approximation and data correction. In: Motwani, R. (ed.) *Proceedings of the 39th IEEE Annual Symposium on Foundations of Computer Science*, pp. 416–425. Los Alamitos, CA. (1998)
- [15] Jiang, T., Kearney, P., Li, M.: A polynomial-time approximation scheme for inferring evolutionary trees from quartet topologies and its applications. *SIAM J. Comput.* 30(6), 1924–1961 (2001)
- [16] John, K.S., Warnow, T., Moret, B.M.E., Vawter, L.: Performance study of phylogenetic methods: (unweighted) quartet methods and neighbor-joining. *Journal of Algorithms* 48, 173–193 (2003)
- [17] Nixon, K.C.: The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15, 407–414 (1999)
- [18] Ragan, M.A.: Phylogenetic inference based on matrix representation of trees. *Mol. Phylo. Evol.* 1, 53–58 (1992)

- [19] Ranwez, V., Berry, V., Criscuolo, A., Fabre, P.H., Guillemot, S., Scornavacca, C., Douzery, E.J.P.: PhySIC: a veto supertree method with desirable properties. *Syst. Biol.* 56(5), 798 – 817 (2007)
- [20] Ranwez, V., Gascuel, O.: Quartet-Based phylogenetic inference: Improvements and limits. *Mol Biol Evol* 18(6), 1103–1116 (Jun 2001)
- [21] Snir, S., Rao, S.: Quartets MaxCut: a divide and conquer quartets algorithm. In: *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2008)
- [22] Snir, S., Warnow, T., Rao, S.: Short quartet puzzling: A new Quartet-Based phylogeny reconstruction algorithm. *J. Comput. Biol.* 15(1), 91–103 (2008)
- [23] Stamatakis, A.: RAxML-NI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690 (2006)
- [24] Strimmer, K., von Haeseler, A.: Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* 13(7), 964–969 (1996)
- [25] Swenson, M.S., Barbançon, F., Warnow, T., Linder, C.R.: A simulation study comparing supertree and combined analysis methods using SMIDGen. *Algorithms for Molecular Biology* 5, 8 (2010)
- [26] Thain, D., Tannenbaum, T., Livny, M.: Distributed computing in practice: the Condor experience. *Concurrency and Computation: Practice and Experience* 17, 323–356 (2005)