

New Approaches to Compare Phylogenetic Search Heuristics

Seung-Jin Sul, Suzanne Matthews, and Tiffani L. Williams
 Department of Computer Science
 Texas A&M University
 College Station, TX 77843-3112
 {sulsj, sjm, tlw}@cs.tamu.edu

Abstract

We present new and novel insights into the behavior of two maximum parsimony heuristics for building evolutionary trees of different sizes. First, our results show that the heuristics find different classes of good-scoring trees, where the different classes of trees may have significant evolutionary implications. Secondly, we develop a new entropy-based measure to quantify the diversity among the evolutionary trees found by the heuristics. Overall, topological distance measures such as the Robinson-Foulds distance identify more diversity among a collection of trees than parsimony scores, which implies more powerful heuristics could be designed that use a combination of parsimony scores and topological distances. Thus, by understanding phylogenetic heuristic behavior, better heuristics could be designed, which ultimately leads to more accurate evolutionary trees.

Keywords: phylogenetic trees, maximum parsimony, phylogenetic heuristics, performance analysis

1 Introduction

Phylogenetics is concerned with inferring the genealogical relationships between a group of organisms (or taxa). These evolutionary relationships are typically depicted in a binary tree, where leaves represent the organisms of interest and edges represent the evolutionary relationships. Phylogenetic trees have been used successfully in designing more effective drugs, tracing the transmission of deadly viruses, and guiding conservation and biodiversity efforts [1], [7]. However, inferring evolutionary trees is not a trivial task. Since it is impossible to know the true evolutionary history for a set of organisms, the problem is often reformulated as an NP-hard optimization problem. Here, trees are given a score, where trees with better scores are believed to be better approximations of the truth. Given the exponential

number of potential hypotheses (or trees) for a set of taxa, an exhaustive exploration of the tree space is not possible. Instead, phylogenetic inference relies on effective heuristics for obtaining good-scoring trees.

In this paper, we develop novel approaches to compare two well-known maximum parsimony (MP) search algorithms, Parsimony Ratchet [9] and Recursive-Iterative DCM3 (Rec-I-DCM3) [12] on three molecular datasets of 60, 174, and 500 taxa. Our parsimony ratchet algorithm is called Paupratsince we used PAUP* [14] to implement it. Our work centers around the following two questions.

1. What value (if any) do slower heuristics provide?
2. How effective are parsimony scores in distinguishing between different tree topologies?

Traditional techniques for comparing phylogenetic heuristics use convergence plots to show how the best score improves over time. Under this measure, the fastest heuristic is desired. Given that different tree topologies may have identical tree scores, preference of good-scoring trees found by fast heuristics may result in overlooking potentially more accurate evolutionary histories that can be found by slower approaches.

Our first observation is that there are benefits to considering different speed heuristic implementations of a MP phylogenetic analysis. In general, Pauprat is a slower heuristic than Rec-I-DCM3. Since we were curious of the merits of a heuristic, time constraints were removed from consideration in this study. However, both Pauprat and Rec-I-DCM3 find different trees with the same best parsimony scores. These diverse best-scoring trees denote that the heuristics are visiting different areas of the exponentially-sized tree space. We note that although TNT [4] has a faster implementation of parsimony ratchet than PAUP*, TNT does not have the capability to return to the user the set of trees found during each iterative step of the parsimony ratchet algorithm (see Section 2.1). The Pauprat implementation of parsimony ratchet provides this capability. Moreover, the

Rec-I-DCM3 implementation also provides users with the trees found during each step of the algorithm.

Secondly, although different trees are found with the same parsimony score, it’s interesting to consider whether maximum parsimony is effectively distinguishing between the trees, which has significant implications for understanding evolution. By using a measure called *relative entropy*, we show for a given collection of trees that parsimony scores have less information content than topological distance measures such as the Robinson-Foulds (RF) distance [11]. In other words, for a collection of trees, parsimony scores identify fewer unique trees—which increases the potential of being stuck in a local optimum and producing less accurate phylogenies—than topological distance measures. Thus, more powerful search strategies could be designed that use a combination of score and topological distance to guide the search into fruitful areas of the exponentially-sized tree space.

2 Maximum Parsimony Heuristics

We study heuristics that use the maximum parsimony (MP) optimization criterion for inferring the evolutionary history of different taxa. Each of the taxa in the input is represented by a string of characters such as DNA or RNA molecular sequences. The molecular sequences for each of the taxa are put into a multiple alignment, so that they all have the same length. Maximum parsimony then seeks a tree, along with inferred ancestral sequences, so as to minimize the total number of evolutionary events by counting only point mutations.

2.1 Parsimony ratchet

Parsimony ratchet is a particular kind of phylogenetic search performed with alternating cycles of reweighting and Tree Bisection Recombination (TBR). The approach works as follows: starting with an initial tree, a few of the characters (between 5 – 25%) are sampled, and reweighted. It suffices to say here that reweighting of characters involves duplicating the characters so that each shows up twice (or more) in the resulting dataset. Then, using these reweighted characters, TBR search is performed until a new starting tree is reached using this subset of data. This new starting tree is then used with the original data set to repeat the phylogenetic search. Parsimony ratchet tries to refine the search by generating a tree from a small subset of the data and using it as a new starting point. If the new tree is better than the old one, then the new one is used as the new starting tree. Otherwise, the old one is kept.

2.2 Rec-I-DCM3

Recursive-Iteration DCM3 (Rec-I-DCM3) [12] implements a disk-covering method (DCM) [5], [6], [8] to improve the score of the trees it finds. A DCM is a divide-and-conquer technique that consists of four stages: divide, solve, merge, and refine. At a high level, these stages follow directly from DCM being a divide-and-conquer technique.

Rec-I-DCM3, involves all of the above DCM stages, but in addition, is both recursive and iterative. The recursive part concerns the divide stage of the DCM, where after dividing the input tree’s leaf nodes into overlapping subsets of taxa, or subproblems, the subproblems themselves may be further divided into smaller subproblems. This is an important enhancement to the DCM approach since for very large datasets, the subproblems remain too large for an immediate solution. Thanks to the recursion, the subproblems are eventually small enough that they may be solved directly using some chosen base method. At this point, Rec-I-DCM3 uses strict consensus merger to do the work of recombining the overlapping subtrees to form a single tree solution. The iterative part of Rec-I-DCM3 refers to the repetition of the entire process just described. That is, the resulting tree solution becomes the input tree for a subsequent iteration of Rec-I-DCM3.

3 Comparing Collections of Trees

3.1 Robinson-Foulds distance

In our experiments, we compare good-scoring trees found by the Pauprat and Rec-I-DCM3 algorithms. We use the Robinson-Foulds (RF) distance to measure the topological distance between two trees. The RF distance between two trees is the number of bipartitions that differ between them. It is useful to represent evolutionary trees in terms of *bipartitions*, in which removing an edge e from a tree separates the leaves on one side from the leaves on the other. The division of the leaves into two subsets is the bipartition B_i associated with edge e_i . Let $\Sigma(T)$ be the set of bipartitions defined by all edges in tree T . The RF distance between trees T_1 and T_2 is defined as

$$d_{RF}(T_1, T_2) = \frac{|\Sigma(T_1) - \Sigma(T_2)| + |\Sigma(T_2) - \Sigma(T_1)|}{2}$$

Our figures plot the *RF rate*, which is obtained by normalizing the RF distance by the number of internal edges and multiplying by 100. Assuming n is the number of taxa, there are $n - 3$ internal edges in a binary tree. Hence the maximum RF distance between two trees is $n - 3$, which results in an RF rate of 100%. The RF rate allows us to compare topological differences when the number of taxa is different. Thus, the RF rate varies between 0% and 100%

signifying that trees T_1 and T_2 are identical and maximally different, respectively.

3.2 All-pairs RF rate

The *All-pairs RF rate* measure takes the average RF distance between a collection of t trees. To do this, we first compute a $t \times t$ matrix of Robinson-Foulds (RF) rates between every pair of trees. Entry (i, j) in the RF matrix corresponds to the RF distance between tree T_i and tree T_j . Since the RF matrix is symmetric, we only sum the RF rates in the upper triangle and divide by $\frac{t(t-1)}{2}$.

3.3 Relative entropy

Entropy represents the amount of chaos in the system. Here, our system consists of a collection of trees that can be distinguished by two features: parsimony score and topological distance. High entropy describes the presence of many unique trees that are evenly distributed in our population (or collection) of trees. Low entropy values describe a population which contains fewer unique trees. In other words, many of the trees in the collection share identical features (i.e, parsimony score or RF rates). Entropy quantitatively captures the distribution of parsimony scores and RF rates among the collection of trees of interest. In our plots, we show *relative entropy*, which is a normalization of entropy, to allow the comparison of entropy values across different population sizes.

Let λ represent the total number of objects (parsimony scores or RF rates) in the population of trees. For example, suppose we want to partition a population of 100 trees based on their parsimony scores. Then, $\lambda = 100$. However, if we are interested in partitioning the 100 trees based on the upper triangle of the corresponding 100×100 RF rates matrix, then $\lambda = \frac{100(99)}{2}$ or 4,950 since the RF matrix is symmetric. Next, we group the λ objects into P total partitions. Each partition, i , contains n_i individuals with identical values. That is, if the objects of interest are parsimony scores, then each of the n_i individuals in partition i will have the same parsimony score. For RF rates, each individual in partition i will have the same RF rate.

The proportion, p_i , of the population of trees occupied by population partition i is $p_i = \frac{n_i}{\lambda}$. We can compute the entropy (E_T) of the collection of parsimony scores as:

$$E_T = - \sum_i^P p_i \log p_i.$$

Higher entropy values indicated more diversity (heterogeneity) among the population of trees. Lower entropy values indicate less diversity (homogeneity) in the population. The lowest entropy value is 0.

Assuming P partitions, the highest entropy value (E_{max}) is $\log P$. To compare collections with different numbers of objects λ , we introduce relative entropy (E_{rel}) defined as the quotient between the entropy E_T and the maximum entropy E_{max} and multiplying by 100 to obtain a percentage. Thus,

$$E = \frac{E_T}{E_{max}} \times 100.$$

3.4 Resolution rate

For n taxa, a complete, unrooted binary tree will have $n - 3$ bipartitions (or internal edges), and all of the evolutionary relationships in the tree are completely resolved. Trees with less than $n - 3$ bipartitions are considered to have unresolved relationships among the n taxa. In general, binary (or 100% resolved) trees are preferred by life scientists. The *resolution rate* of a tree is the percentage of bipartitions that are resolved. One common use of this measure is related to evaluating consensus trees, which are used to summarize the information from a collection of trees. The strict consensus method returns a tree such that the bipartitions of the tree are only those bipartitions that occur in all the trees. The majority consensus uses only those bipartitions that occur in at least 50% of the trees of interest. Thus, the desired result is a highly resolved majority or strict consensus tree, which represents that a high degree of similarity was found among the trees in the collection of interest.

4 Experimental Methodology

4.1 Datasets

We used the following biological datasets as input to study the behavior of the maximum parsimony heuristics.

1. A 60 taxa dataset (2,000 sites) of *ensign* wasps composed of three genes (28S ribosomal RNA (rRNA), 16S rRNA, and cytochrome oxidase I (COI)) [2]. The best-known parsimony score is 8,698, which was established by both Pauprat and Rec-I-DCM3.
2. A 174 taxa dataset (1,867 sites) of insects and their close relatives for the nuclear small subunit ribosomal RNA (SSU rRNA) gene (18S). The sequences were manually aligned according to the secondary structure of the molecule [3]. The best-known parsimony score is 7,440, which was established by both Pauprat and Rec-I-DCM3.
3. A set of 500 aligned *rbcl* DNA sequences (759 parsimony-informative sites) [10] of seed plants. The best-known parsimony is 16,218, which both Pauprat and Rec-I-DCM3 found.

4.2 Starting trees

All methods used PAUP*'s random sequence addition module to generate the starting trees. First, the ordering of the sequences in the dataset is randomized. Afterwards, the first three taxa are used to create an unrooted binary tree, T . The fourth taxon is added to the internal edge of T that results in the best MP score. This process continues until all taxa have been added to the tree. The resulting tree is then used as the starting tree for a phylogenetic analysis.

4.3 Parameter settings

We set the parameters of the Pauprat and Rec-I-DCM3 algorithms according to the recommended settings in the literature. We use PAUP* [14] to analyze our four datasets using the parsimony ratchet heuristic. The implementation of the parsimony ratchet was implemented using PAUP* [14]. For our analysis, we randomly selected 25% of the sites and doubled their weight; initially, all sites are equally weighted. On each dataset, we ran 5 independent runs of the parsimony ratchet, each time running the heuristic for 1,000 iterations.

For Rec-I-DCM3, it is recommended that the maximum subproblem size is 50% of the number of sequences for datasets with 1,000 or less sequences and 25% of then number of sequences for larger datasets not containing over 10,000 sequences. We used the recommended settings established by Roshan et. al [12] for using TNT as a base method within the Rec-I-DCM3 algorithm.

4.4 Implementation and platform

We used the HashRF algorithm to compute the RF distances between trees [13]. Each heuristic was run five times on each of the biological datasets. All experiments were run on a Linux Beowulf cluster, which consists of four, 64-bit, dual dual-core processor nodes (16 total CPUs with gigabit-switched interconnects). Each node contains four, 2 GHz AMD Operton processors and they share 4GB of memory. We note that both Rec-I-DCM3 and parsimony ratchet are sequential algorithms. The parallel computing environment was used as a way to execute multiple, independent batch runs concurrently.

5 Results

5.1 Frequency of the top-scoring trees

Table 1 shows the number of trees found by the Pauprat and Rec-I-DCM3 heuristics in terms of the number of steps they are from the best score, b , we found. Let x represent the parsimony score of a tree T . Then, tree T is $x - b$ steps

away from the best score. In Table 1, $step_0$, $step_1$, and $step_2$ represents trees that are 0, 1 and 2 steps away from the best score, b , respectively. Hence, $step_0$ trees are the trees with the best-known scores. It is clear that the top-scoring trees from Pauprat comprise a large proportion of the total collection of 5,000 trees for the smaller datasets (60 and 174 taxa). On the other hand, the top trees for Rec-I-DCM3 comprise the majority of its collection of trees for the larger dataset. So, if one is simply interested in frequency counts, Pauprat finds best-scoring trees more often than Rec-I-DCM3 on the smaller datasets and Rec-I-DCM3 prevails on the 500 taxa dataset.

5.2 Topological comparisons of top trees

Figures 1 and 2 show the topological differences between the top-scoring trees found by the different search heuristics. We use a heatmap representation, where each value (cell) in the two-dimensional 6×6 matrix is represented as a color. Darker (lighter) colors represent smaller (higher) values such as consensus tree resolution and all-pairs RF rates, which are described in Section 3. Our heatmaps are symmetric two-dimensional matrices. For each heatmap, the bottom values are x coordinates and the values on the left are y coordinates. Each heatmap show 3 types of comparisons among the top-scoring trees (i.e., $step_0$, $step_1$, and $step_2$ trees) found by the search heuristics.

1. Pauprat trees compared to Pauprat trees, which corresponds to cells (x, y) , where $x \leq 3$ and $y \leq 3$.
2. Rec-I-DCM3 trees compared to Rec-I-DCM3 trees, which relates to cells (x, y) , where $x \geq 4$ and $y \geq 4$.
3. Pauprat trees compared to Rec-I-DCM3 trees, which occurs in cells (x, y) , where $x \geq 4$ and $y \leq 3$ or $x \leq 3$ and $y \geq 4$.

Consider the heatmap representation in Figure 1(a). In cell (1,1), the $step_0$ trees found by Pauprat are compared to each other. In particular, the strict consensus is computed for the 1,508 $step_0$ trees (see Tables 1 to get the number of $step_0$ trees). The heatmap plots the resolution rate of the resulting strict consensus tree. High resolution rates (e.g., above 85%) reflect high similarity among the trees of interest. For Pauprat, the $step_0$ (best-scoring) trees for the 60 taxa dataset are all identical resulting in a strict consensus resolution rate of 100%. The heatmap also shows comparisons of trees with different number of steps from the best. For example, cell (3,2) compares $step_2$ ($x = 3$), and $step_1$ ($y = 2$) trees from Pauprat. The resulting strict consensus tree has a resolution rate of 68%, which is based on 1,872 trees (see Table 1). For both Pauprat and Rec-I-DCM3, the majority resolution of comparing the top trees always resulted in a resolution of at least 90% (not shown).

No. of taxa	Pauprat				Rec-I-DCM3			
	step ₀	step ₁	step ₂	% of total	step ₀	step ₁	step ₂	% of total
60	1,508	1,509	363	67.6%	59	343	134	10.7%
174	2,626	1,042	635	86.1%	170	491	1,301	39.2%
500	184	562	955	34.0%	1,231	1,279	983	69.9%

Table 1. Count of the top 3 scoring trees from Pauprat and Rec-I-DCM3. This total is based on the number of top-scoring trees across all five runs of each algorithm. For Pauprat (Rec-I-DCM3), the step₀, step₁, and step₂ trees make up 67.6% (10.7%) of the 5,000 total trees in the collection for the 60 taxa dataset.

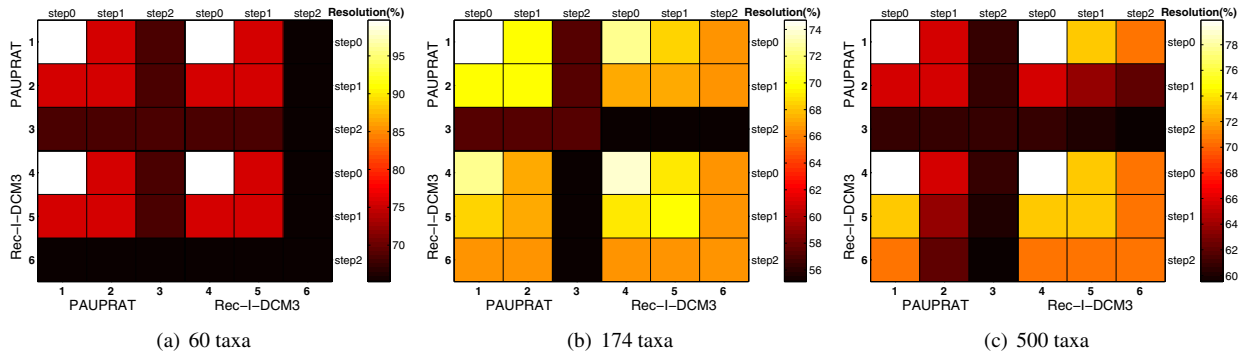


Figure 1. Comparing the strict consensus tree resolution of the top-scoring trees found by the Pauprat and Rec-I-DCM3 heuristics. For best viewing results, please view electronically.

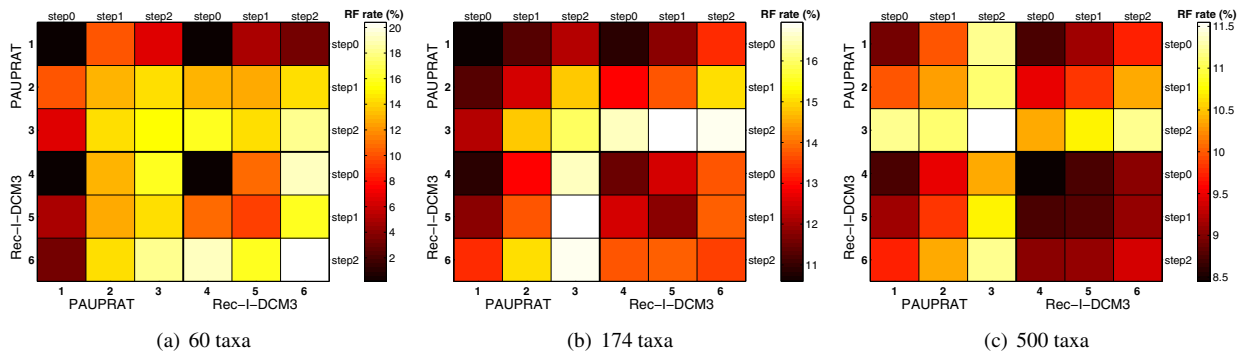


Figure 2. Comparing the all-to-all RF rates of the top-scoring trees found by the Pauprat and Rec-I-DCM3 heuristics. For best viewing results, please view electronically.

Figure 2 is read similarly to Figure 1 except that instead of computing the strict consensus resolution rate of the relevant trees, the average all-pairs RF distance is computed. Let t represent the total number of trees of interest for entry (i, j) in the heatmap. Next, the $t \times t$ RF distance matrix is computed for cell. Since the RF matrix is symmetric, the upper triangle (without the diagonal elements) is used to compute the average. That is, all of the values in the upper triangle are summed and then divided by $\frac{t(t-1)}{2}$ to get the

average RF distance. We normalize this distance by dividing it by $n - 3$ to report the RF rate between 0% and 100% in the heatmap as it makes it easier to compare topological distances across different number of taxa.

Overall, the heatmaps show that the Pauprat and Rec-I-DCM3 algorithms find topologically similar best (step₀) trees, regardless of whether one uses the strict consensus resolution rate or the average RF rate measures. In fact, this is quite interesting that the heuristics land on the same

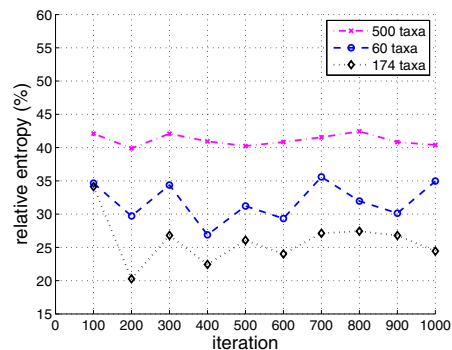
best trees even though they start with different starting tree topologies. There is more variety in the topological structure of the step_1 and step_2 trees of the algorithms. Hence, if trees that have a slightly higher score than the best score are of interest, then both algorithms have value since they explore different areas of tree space.

5.3 Comparisons over time

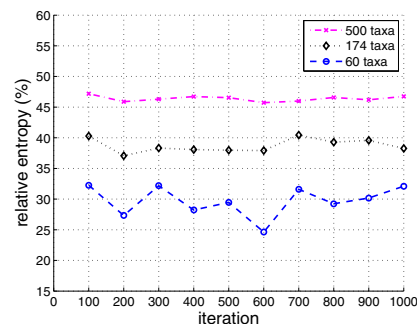
Next, we focus on performance in terms of time using all of the trees returned by each search heuristic. Here, time is measured by number of iterations (which is CPU time independent) and not on wall-clock time (e.g., number of hours required). Although number of iterations is an architecture-independent measure, it may not be completely adequate as each algorithm may do more work than the other per iteration. But, given that we are trying to compare heuristics based on solely their input/output behavior, that is the collection of trees returned after 1,000 iterations, we believe that using iterations as a basis of time is adequate for our purposes in this paper.

Figures 3 and 4 use relative entropy as a measure for uniformly quantifying the information content of parsimony scores and RF rates. Relative entropy is shown as a percentage of the maximum possible entropy. Higher relative entropy means that there is more diversity (heterogeneity) among the values of interest, and hence higher information content. Lower relative entropy values denote homogeneous values and lower information content. One implication of low entropy values is that the search has reached a local optimum. Higher entropy values signify that more diverse trees are found by a phylogenetic heuristic, which lessen its probability of being trapped in local optima.

For 174 and 500 taxa datasets, Pauprat has a higher relative entropy than Rec-I-DCM3 when comparing parsimony scores and RF distances. That is, Pauprat trees are more diverse than Rec-I-DCM3 trees. For the 60 taxa curves, Rec-I-DCM3 has a much higher relative entropy than Pauprat. Moreover, for Rec-I-DCM3, parsimony score entropy values are much higher than RF rate values for 60 taxa. Such a result implies that the parsimony scores of trees are more diverse than their topologies. In other words, trees with different scores when compared topologically are similar. For Pauprat, the relative entropy values vary quite a bit more than for Rec-I-DCM3, which has relative entropy values that are fairly constant across iterations. Essentially such behavior denotes that the Rec-I-DCM3 search has converged as there is not much change in the parsimony or RF rates among the trees found.



(a) Parsimony score



(b) RF rate

Figure 3. Relative entropy values obtained from Pauprat trees every 100 iterations.

6 Conclusions

In this paper, we use *novel* approaches for comparing the phylogenetic trees obtained from two well-known maximum parsimony (MP) heuristics—Pauprat and Rec-I-DCM3. Our results show that although Pauprat is a slower algorithm than Rec-I-DCM3, it is a useful approach to use when reconstructing MP trees. The value of slower algorithms come from finding good-scoring trees that are sufficiently different from their faster counterparts. In essence, the heuristics are finding different classes of trees, which may have different implications about evolution. Of course, in the real-world, speed does matter. As dataset sizes continue to increase, the performance of phylogenetic heuristics must increase as well to keep up with the ever-increasing size of molecular datasets. If the trees found by the slower heuristic finds are good ones, then it may be worth the effort to spend time improving the implementation to make it more competitive in terms of speed.

It is not unusual for phylogenetic heuristics to find hundreds to thousands of best-scoring trees. In some sense this could imply that parsimony scores alone are not fine-grained enough to distinguish between the different topolo-

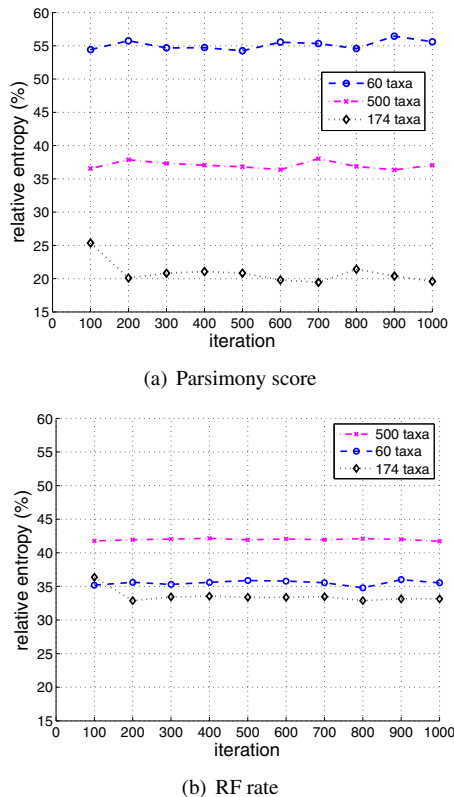


Figure 4. Relative entropy values obtained from Rec-I-DCM3 trees every 100 iterations.

gies of the trees. By using relative entropy, our results show that for trees obtained from our larger datasets, there is more information content in topological distance measures (such as the Robinson-Foulds distance) than in parsimony scores. Hence, heuristics could benefit from actively using topological distance (in addition to parsimony scores) to guide their search through tree space. Our entropy plots also show that Pauprat trees are more diverse than their Rec-I-DCM3 counterparts. Thus, Pauprat appears less likely than Rec-I-DCM3 to get stuck in local optima.

In the future, we plan to develop more measures of comparing heuristics by the collection of trees they find. Of particular interest is developing new heuristics that incorporate topological distance measures as part of the search. Furthermore, we plan to apply our approach for comparing tree collections to other types of heuristics such as maximum likelihood techniques.

7 Acknowledgments

Funding for this project was supported by the National Science Foundation under grants DEB-0629849 and IIS-

0713618. We also thank Matt Yoder for providing us with some of the biological datasets used in this study.

References

- [1] D. Bader, B. M. Moret, and L. Vawter. Industrial applications of high-performance computing for phylogeny reconstruction. In H. Siegel, editor, *Proceedings of SPIE Commercial Applications for High-Performance Computing*, volume 4528, pages 159–168, Denver, CO, Aug. 2001.
- [2] A. R. Deans, J. J. Gillespie, and M. J. Yoder. An evaluation of ensign wasp classification (Hymenoptera: Evanilidae) based on molecular data and insights from ribosomal rna secondary structure. *Syst. Ento.*, 31:517–528, 2006.
- [3] J. Gillespie, C. McKenna, M. Yoder, R. Gutell, J. Johnston, J. Kathirithamby, and A. Cognato. Assessing the odd secondary structural properties of nuclear small subunit ribosomal rna sequences (18s) of the twisted-wing parasites (Insecta: Strepsiptera). *Insect Mol. Biol.*, 15:625–643, 2005.
- [4] P. Goloboff. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics*, 15:415–428, 1999.
- [5] D. Huson, S. Nettles, and T. Warnow. Disk-covering, a fast-converging method for phylogenetic tree reconstruction. *Journal of Computational Biology*, 6:369–386, 1999.
- [6] D. Huson, L. Vawter, and T. Warnow. Solving large scale phylogenetic problems using DCM2. In *Proc. 7th Int'l Conf. on Intelligent Systems for Molecular Biology (ISMB'99)*, pages 118–129. AAAI Press, 1999.
- [7] M. L. Metzker, D. P. Mindell, X.-M. Liu, R. G. Ptak, R. A. Gibbs, and D. M. Hillis. Molecular evidence of HIV-1 transmission in a criminal case. *PNAS*, 99(2):14292–14297, 2002.
- [8] L. Nakhleh, U. Roshan, K. St. John, J. Sun, and T. Warnow. Designing fast converging phylogenetic methods. In *Proc. 9th Int'l Conf. on Intelligent Systems for Molecular Biology (ISMB'01)*, volume 17 of *Bioinformatics*, pages S190–S198. Oxford University Press, 2001.
- [9] K. C. Nixon. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, 15:407–414, 1999.
- [10] K. Rice, M. Donoghue, and R. Olmstead. Analyzing large datasets: *rbcL* 500 revisited. *Systematic Biology*, 46(3):554–563, 1997.
- [11] D. Robinson and L. Foulds. Comparison of weighted labelled trees. In *Proc. Sixth Austral. Conf.*, volume 748 of *Lecture Notes in Mathematics*, pages 119–126. Springer-Verlag, 1979.
- [12] U. Roshan, B. M. E. Moret, T. L. Williams, and T. Warnow. Rec-I-DCM3: a fast algorithmic techniques for reconstructing large phylogenetic trees. In *Proc. IEEE Computer Society Bioinformatics Conference (CSB 2004)*, pages 98–109. IEEE Press, 2004.
- [13] S.-J. Sul and T. L. Williams. A randomized algorithm for comparing sets of phylogenetic trees. In *Proc. Fifth Asia Pacific Bioinformatics Conference (APBC'07)*, pages 121–130, 2007.
- [14] D. L. Swofford. PAUP*: Phylogenetic analysis using parsimony (and other methods), 2002. Sinauer Associates, Sunderland, Massachusetts, Version 4.0.