Cladistics 17, S26–S34 (2001)
doi:10.1006/clad.2000.0156, available online at http://www.idealibrary.com on IDEAL®

Methods for Quick Consensus Estimation

Pablo A. Goloboff* and James S. Farris†

**Instituto Superior de Entomología e Instituto Miguel Lillo, Consejo Nacional de Investigaciones Científicas y Técnicas, Miguel Lillo 205, 4000 S. M. de Tucumán, Argentina; and †Molekylärsystematiska laboratoriet, Naturhistoriska riksmuseet, P.O. Box 50007, S-104 05, Stockholm, Sweden*

Accepted December 12, 2000

A method that allows estimating consensus trees without exhaustive searches is described. The method consists of comparing the results of different independent superficial searches. The results of the searches are then summarized through a majority rule, consensed with the strict consensus tree of the best trees found overall. This assumes that to the extent that a group is recovered by most searches, it is more likely to be actually supported by the data. The effect of different parameters on the accuracy and reliability of the results is discussed. Increasing the cutoff frequency decreases the number of spurious groups, although it also decreases the number of correct nodes recovered. Collapsing trees during swapping reduces the number of spurious groups without significantly decreasing the number of correct nodes recovered. A way to collapse branches considering suboptimal trees is described, which can be extended as a measure of relative support for groups; the relative support is based on the Bremer support, but takes into account relative amounts of favorable and contradictory evidence. More exhaustive searches increase the number of correct nodes recovered, but leave unaffected (or increase) the number of spurious groups. Within some limits, the number of replications does not strongly affect the accuracy of the results, so that using relatively small numbers of replications normally suffices to produce a reliable estimation.

INTRODUCTION

Several authors have criticized parsimony on the grounds that it is an intractable problem. The degree to which parsimony analysis of large data sets is impossible, however, has been much exaggerated. This exaggeration comes, no doubt, from the use of both poor search strategies and programs. For example, Rice *et al.* (1997) were unable to find shortest trees for Chase *et al.*'s (1993) 500-taxon data set in 11.5 months of CPU time using PAUP (Swofford, 1993). In contrast, Nixon (1999), using NONA (Goloboff, 1994) and the ratchet, was able to find trees of minimum length for Zilla in an average time of 4 h, and Goloboff *et al.* (1999), using TNT and combined search strategies, were able to find them in an average time of 10 min. However, even with good strategies and good implementation, finding shortest trees may become very difficult for much larger, or much more complex, data sets. This does not mean that parsimony cannot be used as a criterion in phylogenetic inference. As proposed by Farris *et al.* (1996), finding all most parsimonious trees for large data sets is not only impractical, but also unnecessary. Methods that simply attempt to produce an estimation of the consensus tree, without doing exhaustive searches, are possible, and that problem is explored in this paper.

0748-3007/01 $35.00

Farris *et al.*'s (1996) method combines two different aspects. One is doing a quick estimation based on multiple Wagner trees; the other is using resampling (jackknifing or bootstrapping) as a way to assess support. The quick estimation part is based on using the information provided by the degree to which the results produced by independent searches are similar. It can be assumed that groups found for all (or most) of the independent searches are more likely to be actually supported by the data (but see below, under "Problems"). This is the case regardless of the specific search algorithm used in each individual search and regardless of resampling. Thus, none of the methods discussed here uses character resampling, since they are not primarily intended to measure support; just doing several independent Wagner trees followed by TBR swapping, and keeping only those groups that occured in a high percentage of the replications, normally produces a good estimation of the actual consensus, without the need to ever find multiple trees through branch-swapping. Although this is ideally the case, two problems occur in practice. First, some groups actually supported by the data are not recovered. Second, some of the groups recovered are not actually supported by the data (i.e., the analysis often produces "spurious" groups). Ideally, one would want a method that recovered all the supported groups and no spurious groups, but no quick method proposed so far attains this ideal (and perhaps none will). The extent to which these methods can be trusted, however, can be judged only by comparing their results to the "correct" results (i.e., those obtained by exhaustive searches); the comparisons reported here were automated, by means of several programs and batch files. The "error rate" is the number of spurious groups found, divided by the total number of nodes in the estimated consensus. This gives an idea of the probability of a group picked from inspecting the results being actually unsupported by the data. Note that just recovering many true groups does not produce reliable results; for example, recovering all true groups is not of much help if an equal number of unsupported groups are found: this is a reliability of 50%. The data sets used in the comparisons reported here are Chase *et al.*'s (1993) "Search II" (500 rbcL plant sequences, 398 nodes in the consensus, called "Zilla" by some workers), a data set from Szumik (1997) (85 embiids, 40 nodes in the consensus), and Ballard *et al.*'s (1992) data set (40 invertebrates, 21 nodes in the consensus). Our general conclusions regarding the influence of different factors have also been observed for other data sets. All the timings reported are in a double Pentium II (266 MHz), running under Windows NT. A multithread program, Est, was used (Est is available at ftp.unt.edu.ar/pub/parsimony; a sister program, West, estimates consensus trees under implied weights). All the tables report averages based on repeating 30 runs with different random seeds, except Table 3a (which reports averages for only 15 runs).

The method that has been explored in more detail is the "double consensus" method, implemented in Est/West, which consists of five steps:

(1) do several independent searches (e.g., Wagner tree plus TBR, saving a single tree per replication);

(2) collapse the trees to eliminate unsupported groups;

(3) calculate the majority rule consensus tree for the collapsed trees, using a relatively high cutoff frequency (85 to 95%);

(4) calculate the strict consensus tree of a percentage of the best trees found (here, we used 25%, or all the shortest trees, if more than 25%).

(5) calculate the consensus of the trees produced in steps 3 and 4.

Step (4) indicates that groups not appearing in the best trees will not appear in the final result, which is a form of making more influential those replications that produced better results. To maximize the number of correct groups and minimize the number of spurious groups, we have experimented with changing several parameters.

PROBLEMS

None of the methods studied so far guarantees perfect results in every case. Whether the degree of accuracy is acceptable may well depend on the problem being analyzed and on whether obtaining more reliable results is feasible. The reason for consistently recovering unsupported groups is that some groups may be more frequent in optimal or nearly optimal trees, even if unsupported. It is then much more likely that a given individual search will end up in one tree having the frequent group, even if the group is entirely unsupported. Thus, unsupported but frequent groups are

TABLE 1

Effect of Increasing the Cutoff Frequency, on Zilla (for 16 Replications of Random Addition Sequence Wagner Tree Followed by TBR with No Mulpars, Collapsing with TBR)

Cutoff	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
75	322.03 ± 5.49	81.32 ± 1.39	2.333 ± 1.348	0.718 ± 0.409
85	312.63 ± 5.48	78.95 ± 1.38	1.500 ± 1.137	0.477 ± 0.354
90	299.20 ± 6.39	75.56 ± 1.61	1.067 ± 1.048	0.353 ± 0.342
95	274.40 ± 6.41	69.29 ± 1.62	0.433 ± 0.568	0.158 ± 0.208

consistently recovered. Since this effect is systematic, not random, repeating the estimation will often retrieve the same unsupported groups. This problem affects more than just the estimation methods described here; unless very large numbers of trees are saved (or minimum length is hit many times independently), it also affects more traditional search strategies, such as multiple random addition sequences.

EFFECT OF DIFFERENT PARAMETERS

Cutoff Frequency

The error rate is decreased at higher cutoff frequencies, but then, fewer correct groups are recovered (Table 1). Since some unsupported groups may be found more frequently than some supported groups, it is obvious that increasing the cutoff will lose some supported (infrequent) groups before losing the unsupported (frequent) groups. Increasing the cutoff frequency therefore does not in itself solve the problem, although it makes for more reliable results (i.e., lower error rates).

Collapsing the Trees

To eliminate (or reduce) spurious groups, large numbers of multiple trees per replication should be saved. However, this would require a lot of time in swapping, while the very purpose of these methods is to provide a quick estimation. A solution is to collapse the trees by doing branch-swapping and, when a rearrangement produces a tree of the same length as the one being swapped, collapsing all of the nodes between source and destination (and new root, in the case of TBR). This is equivalent to saving all the trees of equal length produced by swapping the tree and then producing their strict consensus, but takes less time and needs no extra memory space. Collapsing during swapping is the best means of eliminating unsupported groups without greatly reducing the number of correct groups. The swapping algorithm for collapsing can be either SPR or TBR; TBR eliminates more spurious groups than SPR, losing few correct groups (Table 2). The net effect of using TBR for collapsing is then an increase in the reliability of the results.

Exhaustiveness of Searches

Instead of doing a Wagner tree (as in the original method of Farris *et al.*, 1996), every independent replication may search by means of a Wagner tree plus branch-swapping. For more exhaustive searches, the results of the independent replications will be more similar among them, and thus the number of correct groups is increased. However (unexpectedly), the

TABLE 2

Effect of Different Collapsing Algorithms on Zilla (16 Replications of Wagner Tree Followed by TBR with No Mulpars, All with Cut = 85)

Collapse algorithm	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
None	321.60 ± 4.91	81.21 ± 1.24	7.767 ± 2.029	2.356 ± 0.608
SPR	315.23 ± 5.15	79.60 ± 1.30	2.000 ± 1.145	0.630 ± 0.355
TBR	312.63 ± 5.48	78.95 ± 1.38	1.500 ± 1.137	0.477 ± 0.354

TABLE 3a
Effect of Different Search Algorithms, on the Embiid Data Set of Szumik (16 Replications, Cut = 85, Collapsing with SPR)

Search algorithm	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
Wagner	7.467	18.7	0.000	0
NNI	11.133	27.8	0.000	0
SPR	16.733	41.8	1.267	7.04
TBR	18.733	46.8	2.133	11.39

TABLE 3b
Effect of Different Search Algorithms, on Zilla (16 Replications, Cut = 85, Collapsing with SPR)

Search algorithm	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
Wagner	150.23 ± 7.38	37.94 ± 1.86	0.800 ± 0.484	0.528 ± 0.319
NNI	152.73 ± 9.04	38.57 ± 2.28	1.033 ± 0.414	0.674 ± 0.279
SPR	299.57 ± 5.02	75.65 ± 1.27	2.167 ± 1.053	0.716 ± 0.340

number of spurious groups was not significantly reduced when using more exhaustive swapping algorithms (SPR or TBR). When more exhaustive algorithms are used, it is possible to increase the cutoff frequency, losing fewer correct groups, but the absolute number of spurious groups at a given cutoff frequency is the same—or higher. Wagner trees or nearest neighbor interchange has a number of wrong nodes lower than SPR or TBR, because applying the SPR-collapsing algorithm to a Wagner tree produces a very strict collapsing (shorter trees are treated just like trees of equal length). The effect is very pronounced in the embiid data set (Tables 3a and 3b). This allows doing very quick, but still very conservative, searches. For example, doing 16 Wagner trees (collapsed with SPR) for Zilla takes 1 min and produces an error rate below 1% (recovering 38% of the supported nodes).

The best results were obtained when using the parsimony ratchet.[1] The ratchet (Nixon, 1999) is much more exhaustive than using just TBR and also takes longer. As observed with SPR/TBR, increasing the exhaustiveness by using the ratchet often increases the error rate at similar cutoff frequencies. However, the ratchet with a cutoff of 95% recovers more nodes than TBR with a cutoff of 85% (Table 4). It is then possible to have more reliable estimations by combining more exhaustive search algorithms with higher cutoff frequencies. Using the ratchet at a cutoff of 95%, over 80% of the nodes supported were recovered, and the average error rate was 0.174%.

Number of Replications

On first thought, it would appear that doing more replications would improve the results. However, within some limits, the number of replications does not strongly affect the number of correct groups recovered (Table 5). That is determined more by the frequency with which spurious groups appear, and they do appear in a large part of the replications, no matter how many replications are done. The error rate is, anyway, decreased when more replications are used, because then the strict consensus is done for more trees (4 trees for 16 replications, 11 for 45). If fewer replications are done, the results are more subject to sampling error: the variances then are smaller for more replications, even when the means are about the same. The only source for variation in the methods described here is the variability of the results of searches, while in jackknifing it is the same plus the possible combinations of characters sampled. This suggests that, if the only

---

[1] Possibly, better results could be obtained by using the methods of Goloboff *et al.* (1999). Those methods had not yet been developed when the present paper was first submitted.

TABLE 4
Results of TBR and 15 Iterations of TBR-ratchet in Each Replication (Collapsing with TBR)

	Cutoff	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
TBR	85	312.63 ± 5.48	78.95 ± 1.38	1.500 ± 1.137	0.477 ± 0.354
	90	299.20 ± 6.39	75.56 ± 1.61	1.067 ± 1.048	0.353 ± 0.342
	95	274.40 ± 6.41	69.29 ± 1.62	0.433 ± 0.568	0.158 ± 0.208
RAT	85	358.90 ± 6.71	90.63 ± 1.70	2.633 ± 1.245	0.726 ± 0.340
	90	346.47 ± 5.54	87.49 ± 1.40	1.567 ± 1.406	0.446 ± 0.395
	95	326.77 ± 7.23	82.52 ± 1.83	0.567 ± 0.971	0.174 ± 0.297

purpose is estimating the consensus, doing large numbers of replications (as required in jacknifing or bootstrapping) is actually unnecessary. In general, 16 replications provides a reasonable estimation. Because the strict consensus of a fourth of the replications is done, the net effect of using more replications is more reliable and conservative estimations.

Using Only the Best Trees

If larger numbers of replications are used, one could discard some of the trees. In some sense this is equivalent to increasing the exhaustiveness of the search, and it is therefore unsurprising that it increases the number of correct nodes recovered, but also increases slightly the number of spurious nodes (Table 6). The gain in correct nodes recovered by doing more TBR replications and discarding the worse trees, however, is lower than the gain by doing a smaller number of replications with a more exhaustive search algorithm (like the ratchet), especially at higher cutoffs, and therefore this option is not of much help, unless very large numbers of replications are done (i.e., hundreds).

RELATIVE SUPPORTS

Because the best way found to decrease the error rate without decreasing the number of correct nodes is to use better collapsing algorithms, collapsing on rearrangements that produce slightly suboptimal trees might improve the results. The degree to which a tree is suboptimal compared to another could be determined by the absolute step (or fit) difference between the two trees, as in the Bremer support. A defect of that method, however, is that it does not always take into account the relative amounts of evidence contradictory and favorable to the group. For example, according to the Bremer support, a group supported by 2 uncontradicted characters is better supported than a group supported by 100 and contradicted by 97. The first case, however, is a relatively well-supported group, while in the second case there is about as much evidence in favor of the group as against. This problem is diminished if the support for the group is calculated as the ratio between the amounts of favorable and contradictory evidence. The amounts of contradictory and favorable evidence can be estimated by comparing

TABLE 5
Effect of the Number of Replications, on Zilla (Each Replication with Wagner Tree Plus TBR, Collapsing with TBR)

Replications	Cutoff	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
16	85	312.63 ± 5.48	78.95 ± 1.38	1.500 ± 1.137	0.477 ± 0.354
	90	299.20 ± 6.39	75.56 ± 1.61	1.067 ± 1.048	0.353 ± 0.342
	95	274.40 ± 6.41	69.29 ± 1.62	0.433 ± 0.568	0.158 ± 0.208
45	85	299.30 ± 4.02	75.58 ± 1.01	0.467 ± 0.571	0.155 ± 0.190
	90	293.37 ± 3.85	74.08 ± 0.97	0.367 ± 0.556	0.124 ± 0.188
	95	283.30 ± 5.31	71.54 ± 1.34	0.233 ± 0.430	0.082 ± 0.151

TABLE 6

Effect of Considering Only the Best Trees for the Estimation (45 Replications of Wagner Tree Plus TBR, Collapsing with TBR)

Cutoff	Use best	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
85	100%	299.30 ± 4.02	75.58 ± 1.01	0.467 ± 0.571	0.155 ± 0.190
	50%	317.20 ± 4.54	80.10 ± 1.15	1.600 ± 1.545	0.498 ± 0.475
	33%	323.87 ± 5.85	81.78 ± 1.48	2.400 ± 1.610	0.733 ± 0.487
90	100%	293.37 ± 3.85	74.08 ± 0.97	0.367 ± 0.556	0.124 ± 0.188
	50%	310.27 ± 3.42	78.35 ± 0.86	1.133 ± 1.196	0.361 ± 0.377
	33%	312.77 ± 4.29	78.98 ± 1.08	1.533 ± 1.432	0.485 ± 0.447
95	100%	283.30 ± 5.31	71.54 ± 1.34	0.233 ± 0.430	0.082 ± 0.151
	50%	300.10 ± 4.77	75.78 ± 1.21	0.567 ± 0.679	0.186 ± 0.222
	33%	293.73 ± 5.94	74.18 ± 1.50	0.667 ± 0.994	0.225 ± 0.336

the steps, character by character, among different trees. Given two trees, when character i fits the most parsimonious tree better, the (weighted) fit difference for that character in the two trees is favorable to the most parsimonious tree (f_i). When character i fits the least parsimonious tree better, the fit difference is contradictory (c_i). Define $F = \Sigma f_i$, and $C = \Sigma c_i$. The Bremer support measures support of groups using simply the difference $F - C$, but this is merely one of the aspects of the support of a group. The other aspect of the support of a group is given by the ratio C/F. C/F varies between 0 and 1, inversely with the support of a group, so it is better to use its complement as a measure of support, $1 - C/F = (F - C)/F$. The ratio $(F - C)/F$ is the relative fit difference, RFD, between two trees. If RFD $= 0$, the group is entirely unsupported, and if RFD $=$ 1, the group is entirely uncontradicted. A measure of relative support based on the RFD could be calculated like the Bremer supports, comparing optimal and suboptimal trees. What is more important for present purposes is that the relative supports can be easily estimated during branch-swapping for tree collapsing. In this case, all the nodes between source and destination for a rearrangement that produces a tree with a relative fit difference below a specified value Q can be collapsed. When a clade is clipped off the main tree (to be moved to a different location), the decrease in length places an upper bound in the maximum increase that still can lead to a relative fit difference below Q. Denote the length difference (summed over all characters) between the entire tree and the divided tree as D, and the difference between the divided tree and the new (suboptimal) tree as X. Then,

$$\frac{X - D}{X} \leq \text{RFD}$$

(because $X - D = F - C$ and $X \geq F$). Therefore, only when

$$\frac{X - D}{X} < Q$$

will it be possible for the RFD between the original tree and the new tree to be less than Q. But further, that inequality can be rewritten as

$$X < \frac{D}{(1 - Q)}$$

and therefore, when calculating length for a rearrangement, as soon as the length increase X is greater than D divided by $(1 - Q)$, the rearrangement can be abandoned without further calculations (i.e., without calculating length increases for any additional characters). When this is done, a round of swapping on an optimal tree takes almost as little time as needed to collapse when the trees are equally optimal; for $Q =$ 0.10, only an additional 5% of time is necessary for tree collapsing.

Potential advantages of the relative supports (i.e., measured using RFD) over normal Bremer support are that they vary between 0 and 1 and they provide an approximate measure of the amount of favorable/contradictory evidence (for example, if the RFD is 0.25, the amount of contradictory evidence is 75% the amount of favorable evidence, so it is equivalent to the conflict of 4 characters versus 3). Under weighting methods such as those of Goloboff (1993, 1997), the Bremer supports may be hard to interpret, but the relative supports—for different weighting strengths—are directly comparable.

A disadvantage of the relative supports is that the values of RFD in different pairs of trees must be calculated carefully. In particular, if **a** and **b** are two equally parsimonious trees, and **c** is less parsimonious, it is possible that $\mathrm{RFD}_{\mathrm{ac}} \neq \mathrm{RFD}_{\mathrm{bc}}$. This is because $F_{\mathrm{ab}} = C_{\mathrm{ab}} = 0$, and $F_{\mathrm{ac}} - C_{\mathrm{ac}} = F_{\mathrm{bc}} - C_{\mathrm{bc}}$, but possibly $F_{\mathrm{ac}} \neq F_{\mathrm{bc}}$ and $C_{\mathrm{ac}} \neq C_{\mathrm{bc}}$. Consider the data set of Fig. 1 as an example. That data set produces two trees of 14 steps. The "correct" value of relative support for group (ab) is 0.5, since there are two characters in favor and one against. This value of RFD is obtained for nonmonophyly of group (ab) if comparing the first tree in Fig. 1 (suboptimal) with the second (optimal). The only difference between the two trees is the monophyly of (a b). However, if the comparison is done between the first and the third (optimal) trees, RFD = 0.25; $F - C$ is the same when comparing the suboptimal tree to any of the optimal trees, but in the second case both F and C are larger because of conflict within (d e f). Conflict within (d e f) is actually irrelevant at the time of determining conflict within (a b c), and to factor out cases like this, the RFD for a suboptimal tree should be considered as the maximum RFD between the suboptimal tree and each of the optimal trees. Other cases where the RFD value may be misleading are possible, as in the example shown in Fig. 2. In that case, there is a single optimal tree, with group (a b) contradicted by no characters, but there are some suboptimal trees with RFD < 1 and having the group as not monophyletic. An example is the second tree shown, suboptimal by virtue of (1) making the uncontradicted (a b) nonmonophyletic and (2) making (d e) monophyletic. Because there are some characters in favor of (d e), C is greater than 0, and the RFD between the two trees is 0.50.

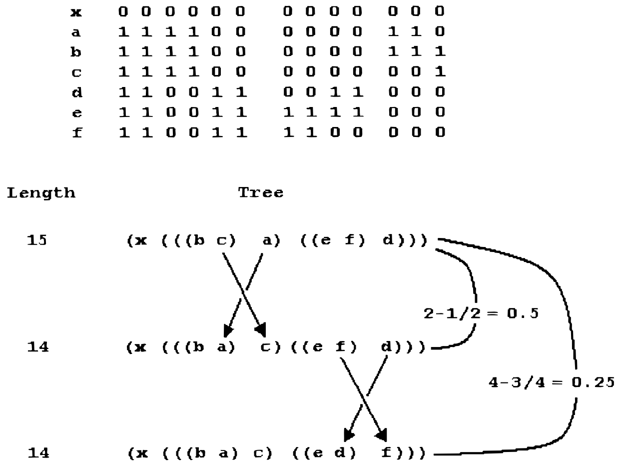

FIG. 1. See text for explanation.

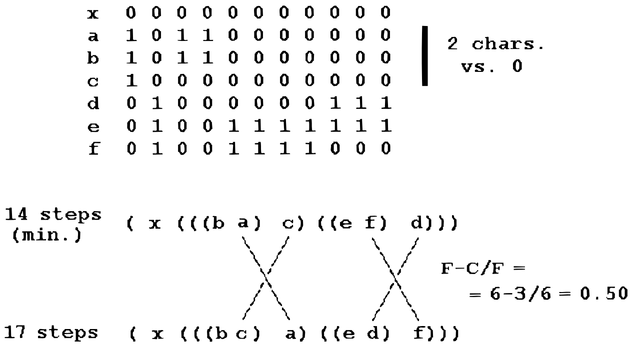

FIG. 2. See text for explanation.

The problems mentioned could perhaps be avoided if F and C are measured onto trees chosen more carefully. A reasonable possibility (suggested by M. Ramírez, personal communication) is to use, in order to evaluate support for each group, only those trees that differ from the optimal trees in a number of steps no greater than the Bremer support of the group. This modification would produce a correct evaluation of the support for the example in Fig. 2, but we have not tested it on a wide variety of cases. The modification, however, seems unnecessary for our purpose, which is just using the RFD as a quick and approximate way to consider suboptimal trees, not using it as a measure of support. The RFD is best used in combination with the absolute fit difference, that is, to collapse nodes with RFD below Q, and a step (or fit) difference below S steps.

The number of wrong groups and the error rate decrease when collapsing the nodes in which the estimated support is low (Table 7), but so does the number of correct nodes (obviously, many nodes in the consensus are poorly supported). Using this leads to estimations with fewer errors, but with less resolution. Using low values of Q, the resolution is not decreased so much, and then this may be of help in combination with higher cutoff frequencies.

TABLE 7

Effect of Increasing the *Q* Value, for Zilla (16 Replications of TBR, Collapsing with TBR, Cutoff 90%)

Q	Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
0	299.20 ± 6.39	75.56 ± 1.61	1.067 ± 1.048	0.353 ± 0.342
10	292.17 ± 8.25	73.78 ± 2.08	0.967 ± 0.809	0.328 ± 0.271
15	272.40 ± 9.23	68.79 ± 2.33	0.833 ± 0.747	0.303 ± 0.269
25	212.93 ± 8.00	53.77 ± 2.02	0.433 ± 0.504	0.203 ± 0.237
33	164.07 ± 10.71	41.43 ± 2.70	0.400 ± 0.498	0.247 ± 0.310

Additionally, when tree collapsing is done with absolute and relative fit differences combined, the least supported groups tend to be eliminated first. For Ballard *et al.*'s (1992) data set, the groups were first arranged according to their jacknife supports (as measured with the GC statistic, calculated with more or less exhaustive searches), subsequently checking which groups were and were not recovered with the estimation procedure (Fig. 3, using 5 to 10 runs in each case; a dark square indicates presence of the group, and a light square indicates absence). Using either RFD or absolute fit difference alone to collapse groups during swapping (first three columns in Fig. 3), it was found that many recovered groups have a support lower than that of some unrecovered groups. This effect is decreased if the absolute fit difference is used in combination with RFD, where in 6 of 10 cases a single group of low support is recovered, and in the remaining 4, no group of low support is recovered (rightmost column in Fig. 3).

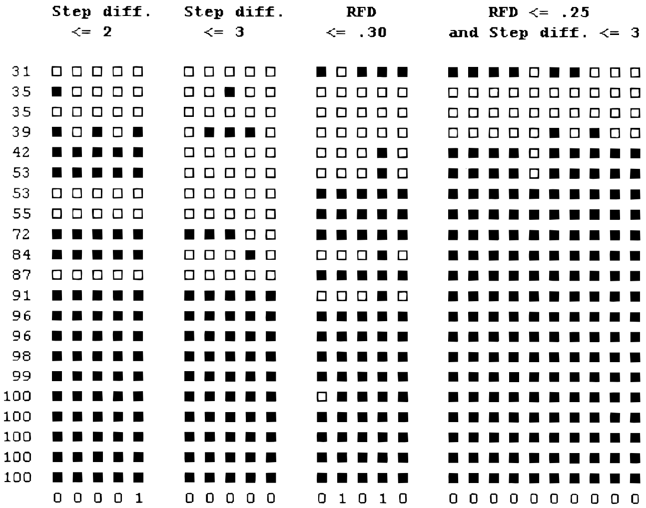

FIG. 3. Groups for Ballard *et al.*'s (1992) data set, arranged according to their jacknife GC values (indicated in the leftmost column). The numbers in the bottom row indicate the number of spurious groups found in each case. See text for additional explanation.

DISCUSSION

Rice *et al.* (1997) ran Zilla on PAUP for 3.5 months on three machines; the consensus they found recovers more correct nodes than the estimation methods described here, but also has a large number of spurious nodes (Table 8). Thus, the chances of a group picked by inspecting the results of Rice *et al.* (1997) being wrong (i.e., not actually supported by the data) are over 10%. Rice *et al.* (1997) made an effort in trying to find the minimum length possible, but they found that length only once, so that their result is based on a single tree island. The high error rate of Rice *et al.*'s (1997) tree shows that just finding one or a few reasonably short trees does not solve the problem of discovering which groups are actually supported by the data.[2] In this regard, the estimation methods described here are not only much faster than Rice *et al.*'s (1997) strategy, but also significantly more precise and reliable. On a double Pentium II (266 MHz), both the TBR

TABLE 8

Statistics for Rice *et al.*'s (1997) Tree

Correct nodes	True nodes recovered (%)	Wrong nodes	Error rate (%)
388	97.49	46	10.60

---

[2] This would apply as well if Rice *et al.* (1997) had succeeded in finding a tree of minimum length, but only once or twice.

and ratchet estimations (at cutoff 95%) take 10 min to 1 h to complete (2500 to 15,000 times faster). The probability of a group based on such methods being wrong is only 0.17 to 0.16%—60 times more reliable than Rice *et al.*'s (1997) consensus tree.

ACKNOWLEDGMENTS

We thank G. Giribet and W. Wheeler for comments. We also acknowledge support from CONICET (PEI 0324/97) and ANPCyT (PICT 01-04347) to P.A.G. and NFR (10204-303) to J.S.F.

REFERENCES

Ballard, J. W. O., Olsen, G. J., Faith, D. P., Odgers, W. A., Rowell, D. M., and Atkinson, P. W. (1992). Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science* **258,** 1345–1348.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K.-J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q.-Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Willimas, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H., Jr., Graham, S. W., Barret, S. C. H., Dayanandan, S., and Albert, V. A. (1993). Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Mo. Bot. Gard.* **80,** 528–580.

Farris, J., Albert, V., Kallersjo, M., Lipscomb, D., and Kluge, A. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12,** 99–124.

Goloboff, P. (1993). Estimating character weights during tree search. *Cladistics* **9,** 83–91.

Goloboff, P. (1994). NONA: A tree searching program. Program and documentation, available at ftp.unt.edu.ar/pub/parsimony.

Goloboff, P. (1997). Self-weighted optimization: Character state reconstructions and tree searches under implied transformation costs. *Cladistics* **13,** 225–245.

Goloboff, P., Farris, J., and Nixon, K. (1999). T.N.T.: Tree Analysis Using New Technology. Program available at www.cladistics.com.

Nixon, K. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* **15,** 407–414.

Rice, K., Donoghue, M., and Olmstead, R. (1997). Analyzing large data sets: rbcL 500 revisited. *Syst. Biol.* **46,** 554–563.

Swofford, D. (1993). PAUP: Phylogenetic Analysis Using Parsimony, version 3.1. Program and documentation, Laboratory of Molecular Systematics, Smithsonian Institution, Washington, DC.

Szumik, C. (1997). Contribución al conocimiento de la sistemática de Embioptera (Insecta) Neotropicales. Tesis doctoral, Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán.