

Points of View

Syst. Biol. 49(2):363-368, 2000

Simple but Fundamental Limitations on Supertree and Consensus Tree Methods

MIKE STEEL,¹ ANDREAS W. M. DRESS,^{2,3} and SEBASTIAN BÖCKER²

¹Biomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch, New Zealand; E-mail: m.steel@math.canterbury.ac.nz

²GK Strukturbildungsprozesse, FSP Mathematisierung, Universität Bielefeld, PF 100 131, 33501 Bielefeld, Germany

A fundamental problem in biological classification is the question of how best to combine into one phylogenetic tree a collection of phylogenetic trees that classify the same or overlapping sets of taxa. When the sets of taxa are the same, this problem is known as the *consensus tree problem*; the more general situation, where the trees classify possibly different, though overlapping sets of taxa, has become known as the *supertree problem* (Sanderson et al., 1998). For the supertree problem, even if the input trees are compatible (defined below), there may be many, essentially different ways to amalgamate them. A further complication is that, for either problem, the given trees may not be compatible. Furthermore, even the task of merely determining compatibility is difficult for the supertree problem on unrooted trees (Steel, 1992).

The search for "good" new algorithms to amalgamate trees is topical, particularly for the supertree problem because existing methods appear to have certain limitations and seem somewhat ad hoc: A common supertree method is to recode trees by characters and apply maximum parsimony reconstruction; see for example, Baum (1992),

Ragan (1992), Purvis (1995a), and Sanderson et al. (1998).

The purpose of this note is to point out that inherent limitations exist as to what any supertree or consensus tree method can possibly achieve. More precisely, we demonstrate here by simple examples that, although one can easily list various desirable properties that one would like such a method to possess, no method can possess all of these simultaneously. Although this may be viewed as a somewhat negative result, its positive spin-off is that it focuses attention on those types of supertree constructions one should aim for and indicates promising lines for future work. Although we insist that any supertree method must return a single tree, we will briefly discuss extensions that allow several trees to be in the output.

One inherent limitation on the solution to the consensus tree problem for unrooted trees has been described by McMorris (1985), who discussed three conditions that cannot simultaneously be satisfied by any consensus method. The conditions we consider here are quite different and somewhat simpler. We will also see that there is a fundamental distinction between rooted and unrooted trees.

LIMITATIONS ON SUPERTREE METHODS

We begin with some terminology. Unless indicated otherwise, we will regard a phy-

³Current address: The City College, The City University of New York, Department of Chemical Engineering, Convent Avenue at 140th Street, New York, NY 10031, USA.

logenetetic tree as unrooted because phylogeny reconstruction algorithms generally do not locate the root of a tree (despite the tendency of some programs to represent trees as rooted). We say that a phylogenetic tree T resolves a phylogenetic tree T' if T' can be obtained from T by collapsing edges. A phylogenetic tree is fully resolved if it cannot be resolved any further; this is the same as requiring that the tree has no (non-hard) polytomies (i.e., for an unrooted tree, each interior vertex has degree 3; for a rooted tree, each interior vertex has exactly two descendant vertices). We say that a phylogenetic tree T displays a phylogenetic tree t if the phylogenetic tree that one obtains from T by deleting all taxa (and connecting edges) that do not appear in t is equal to, or resolves t . (Allowing for resolution is in line with the "soft polytomy" interpretation, whereby a multifurcation in the tree is regarded as uncertainty as to the exact order of speciation, rather than certainty as to a multiple and simultaneous speciation event.)

A collection C of phylogenetic trees is said to be compatible if there exists a phylogenetic tree T that displays each tree in C . In this case, T is said to be a parent tree for the collection C .

For the supertree (or the consensus tree) problem, we will say that a supertree (or a consensus tree) method A takes a collection of input trees and returns exactly one output tree.

Consensus methods are sometimes described as applying to unordered inputs (sets) of trees, or to ordered inputs of trees, or (more generally), to a partial ordering of input trees, where the partial order might reflect one's relative confidence in each tree. Whatever viewpoint is adopted, any consensus method should be able to deal with the case in which the trees have equal support. This motivates the first entry in our list of three desirable properties:

- P1. The method can be applied to any unordered set of input trees.
- P2. If we rename all the species, and then apply the method to the new input trees, the output tree is simply the old output tree, but with the species renamed accordingly.

- P3. If there exists at least one parent tree for the given collection of input trees, then the output tree is one of those parent trees (that is, the output tree displays the input trees whenever the latter are compatible).

Property P1 requires that in case we have equal confidence in all the input trees, then the arbitrary order in which we input these trees does not matter (formally, method A is symmetric for such inputs in this case). Property P2 requires that the manner in which we name or label our species does not affect the outcome (formally, method A operates equivariantly with respect to permutations of the taxa set). Property P3 requires that if the input trees do fit together, then the method selects one of the trees that displays all the input trees.

Properties P1 and P2 seem absolutely compelling, whereas P3 also appears to be desirable. However, our first result says that all three properties cannot be simultaneously achieved for unrooted trees.

Proposition 1. There is no supertree method that satisfies P1–P3 for unrooted phylogenetic trees, even when the input trees are restricted to being fully resolved.

We offer the following simple argument to establish this claim. For taxa I, J, K, L , let us write $(IJ)(KL)$ to denote the fully resolved unrooted phylogenetic tree that groups together taxa I and J versus taxa K and L . Consider six taxa, labeled 1, 2, . . . , 6, and the input trees (12)(45), (34)(16), (56)(23). Note that these three input trees are compatible. Suppose a supertree method A satisfies P1 and P2; we will show that P3 fails. The key observation (from Böcker et al., 1999) is that there are precisely two trees that are parent trees for this collection of trees. Therefore, precisely two potential candidates exist for the output tree of any supertree method that satisfies P3—the two trees shown in Figure 1. Now, notice that if we had interchanged the names of taxa 2 and 6, and also interchanged the names of taxa 3 and 5, then our collection of input trees would become (16)(43), (54)(12), (32)(65) which are the same three input trees, just listed in different order. Thus, by P1 and P2, the output tree resulting from the application of method A to the un-

ordered set consisting of the three input trees must remain unchanged under these two simultaneous taxa interchanges. Yet, neither of the two possible parent trees (in Fig. 1) has this property (in fact, performing simultaneously the taxa interchanges of 2 with 6, and 3 with 5 simply interchanges the two parent trees). Consequently, regardless of which of the two possible parent trees method A chose as its single output tree, we see that P3 cannot hold, as we have claimed.

This result suggests that the supertree problem simply does not have a globally satisfactory solution, at least if one allows unrooted trees. The imperative nature of P1 and P2 suggests either (1) abandoning property P3, or (2) dealing with more restrictive inputs such as rooted trees, or (3)

outputting a collection of supertrees that lists all of the competing phylogenetic inferences that can be drawn from the collection of input trees.

With regard to option (1), it is easy to satisfy P1 and P2 if we abandon P3—for example, by always outputting the completely uninformative star tree! An interesting problem is how to devise a more sensitive approach that retains some phylogenetic information, yet can be carried out when the input trees are both large and numerous. For example, a first simple approach might be to take the strict or majority rule consensus of the set S of all parent trees for the input trees (outputting a star tree if S is empty). However, it is not at all clear how to calculate this tree in realistic time for large-scale inputs; and as is apparent from the proof of Proposition 1, the resulting tree is, in general, not a parent tree, even if parent trees exist.

Option (2)—working with rooted trees—also offers some promise. For example, it can be shown (see Semple and Steel, 1999) that a suitable modification of the algorithm developed by Aho et al. (1981) (see also Constantinescu and Sankoff, 1995; Ng and Wormald, 1996) provides a supertree method for rooted trees that satisfies P1–P3 as well as the following two desirable properties:

- P4. Each leaf (taxon) that occurs in at least one input tree occurs in the output tree.
- P5. The time required to compute the output tree grows polynomially with the total number of species.

Property P4 states that no taxa are discarded, even if the input trees are not compatible. Note that this property is not implied by property P3, because P3 applies only when the input trees are compatible. Property P5 is desirable if one hopes to compute an output tree in realistic time for many input trees involving large numbers of species (for a recent example of a large-scale study, see Purvis, 1995b). Methods that do not have an exact polynomial-time implementation (such as those based on maximum parsimony reconstructions) may nevertheless be approximated by using fast

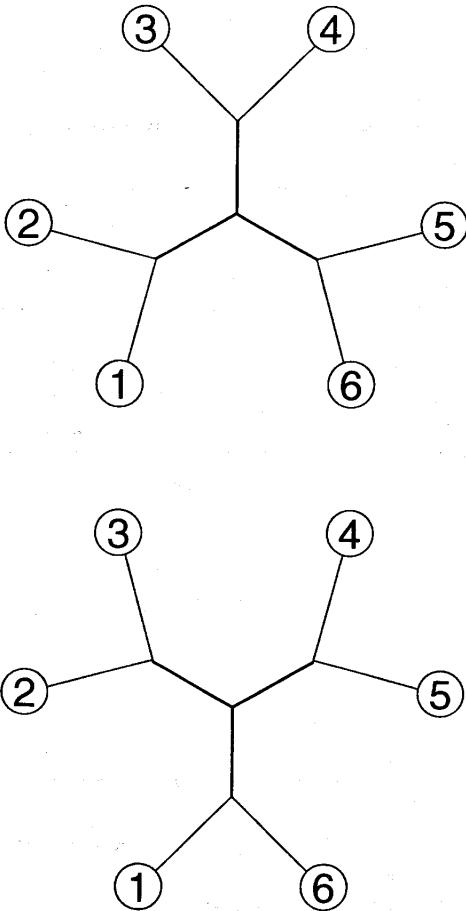


FIGURE 1. The two phylogenetic trees that display the trees (12)(45), (34)(16), (56)(23).

heuristic techniques; however, requirements (for example, P3) known to be met by the exact method may no longer be satisfied under a heuristic implementation.

Thus, option (2) may offer a promising line of approach for the development of a good supertree method. Note that the "root" of the input trees need not correspond to the temporal (ancestral) root: If, for example, a collection of unrooted input trees has the property that some taxon J appears in the taxa set of each tree, then we may regard J in a purely formal way as a "pseudo-root" and proceed as follows: We delete taxon J and its incident edge from each tree to obtain a collection of rooted trees, the root of each tree being the vertex that was adjacent to the leaf labeled by taxon J . We can then apply an algorithm that amalgamates rooted trees to this collection of rooted trees, and then reattach taxon J to the resulting (rooted) output tree by making it adjacent to the root of this tree.

Unfortunately, selecting as our "pseudo-root" a different taxon (J') appearing in the taxa set of each tree may, by the process described, lead to a different output tree. Thus P2 does not necessarily hold across all taxa, under this form of rooting.

Option (3) is also appealing. For example, supertree methods based on maximum parsimony reconstruction may return more than one tree. This holds also for the consensus tree method described by Wilkinson (1994). If the collection of input trees is compatible, one would, ideally, like to describe the set of all of its parent trees. However, this set may be very large. For example, if there is a parent tree that is highly unresolved, then any of the many possible resolutions of this tree would provide a legitimate parent tree, too. Thus, a more realistic aim would be to describe the set of all minimally resolved parent trees. Unfortunately, this set can also be very large, because one can construct a set of compatible input trees for which every parent tree is fully resolved (binary), and yet the number of these parent trees grows exponentially with the number of taxa (Böcker, 1999).

Any method that outputs a set of trees, and satisfies P4 for each output tree, can be extended to return a single tree by feeding the output trees into a consensus tree method. For example, following the ap-

proach of the previous paragraph, given any collection of compatible input trees, we could compute the set of all trees (or all minimally resolved trees) that display each input tree, and then apply the strict consensus method to all the resulting trees. This is Gordon's (1986) consensus supertree for compatible sets of trees. Restricted to compatible sets of input trees, this method fails P3 (it would return a star tree for the two trees shown in Fig. 1). However, it satisfies P1, P2, and P4; and, at least for rooted trees, it also satisfies P5 (Steel, 1992).

LIMITATIONS ON CONSENSUS TREE METHODS

Given that the input trees in the consensus tree problem all have the same taxa set, we might hope that the nastiness evident in Proposition 1 would disappear for such input trees. Indeed, it does; for example, the "combinable component" consensus method (also called the "loose" consensus method, see Bremer, 1990; Barthélemy et al., 1992) satisfies all properties P1–P5. More popular methods, such as the majority-rule and strict consensus methods, also satisfy P1, P2, P4, and P5, and can be easily modified to satisfy P3 as well.

One problem with all such methods is that when the input trees are not compatible, these methods often lead to highly unresolved (starlike) output trees, and even phylogenetic relationships that are shared by all input trees may disappear in the output tree. We now show that this is inevitable for unrooted phylogenetic trees (in contrast to the consensus problem on rooted trees, where the Adams consensus method preserves shared phylogenetic relationships; see Adams, 1986).

In other words, we might hope for a consensus method that satisfies the following property:

- P6. If all input trees display $(IJ)(KL)$, then the output tree displays $(IJ)(KL)$.

However, if we retain the requirement that a consensus method always outputs a single tree then we have:

Proposition 2. There is no consensus method for unrooted phylogenetic trees that satisfies properties P1, P2, and P6.

The justification here is similar to that for the previous proposition. Consider as input the two trees shown in Figure 1. Both trees display the trees (12)(45), (34)(16), (56)(23) and are, furthermore, the only two phylogenetic trees displaying these three quartet trees simultaneously. Consequently, if P6 were satisfied, the output tree would have to be one of the two trees in Figure 1.

Now, suppose there were a consensus method that also satisfied P1 and P2 for these two input trees. Recall that by the interchange of taxa 2 with 6 and 3 with 5, the two input trees (in Fig. 1) are interchanged, and so the unordered set of input trees is fixed. Consequently, applying method A to the unordered set consisting of the two input trees must produce an output tree that remains fixed under the simultaneous interchange of taxa 2 with 6 and 3 with 5. But as we have seen above, the trees in Figure 1 are in fact interchanged by this property; given that these were the only candidates for the output tree, property P6 cannot be satisfied.

As before, the situation is more appealing for rooted trees. Let us write $(IJ)K$ to denote the fully resolved rooted phylogenetic tree with taxa I and J on one side and taxon K on the other side. Then, consider the following two properties—the first being the analog of property P6 for rooted trees, the second being obviously much stronger:

P6'. If all the input trees display $(IJ)K$, then the output tree displays $(IJ)K$.

P7. If at least one input tree displays $(IJ)K$ and no input tree displays $(IK)J$ or $(JK)I$, then the output tree displays $(IJ)K$.

This last condition requires that if at least one input tree displays a particular tree $(IJ)K$ for three species, I , J , and K , and if each of the other input trees displays either this tree or an unresolved tree for I , J , and K , then the tree $(IJ)K$ is displayed by the output tree.

Proposition 3. There exists a consensus method for rooted phylogenetic trees satisfying properties P1–P5 and P6'; but there is no consensus method for rooted phylogenetic trees that satisfies property P7.

The Adams consensus method satisfies P1–P5 and P6' (see Adams, 1986). To justify

the second claim, consider four rooted trees (Fig. 2), each involving species 1, 2, . . . , 5, in which each tree has one nontrivial cluster, thereby grouping together the following pairs of taxa: {1,2}, {2,3}, {3,4}, {4,5}. These trees display, respectively, the trees (12)5, (23)5, (34)1, (45)1; and none of the input trees displays $(IK)J$ or $(JK)I$ for any tree $(IJ)K$ in this list. If there were a consensus method satisfying P7, then the output tree would have to display (12)5, (23)5, (34)1,

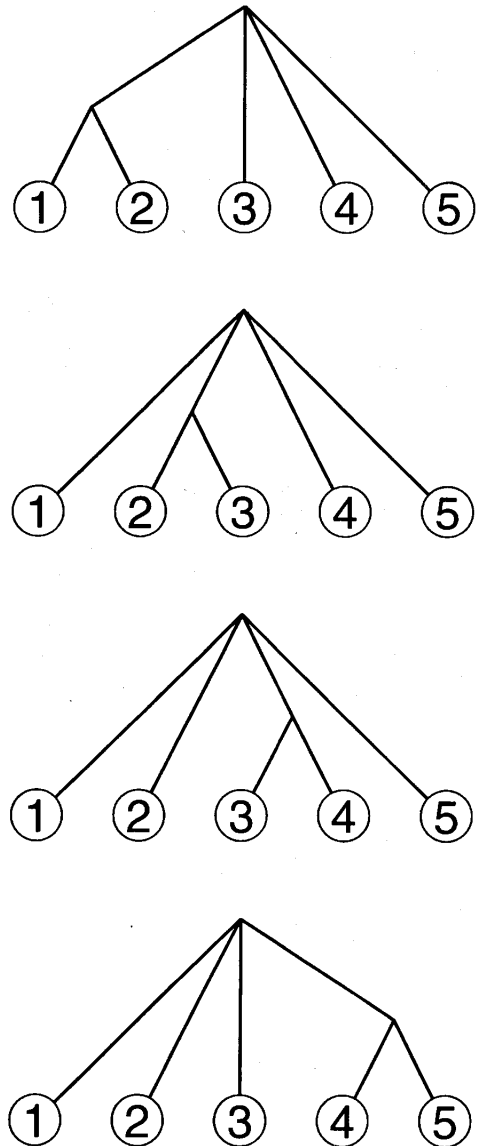


FIGURE 2. Four trees with nontrivial clusters {1,2}, {2,3}, {3,4}, {4,5}.

(45)1. However, it is easily verified that no rooted phylogenetic tree can display these last four trees simultaneously.

Comparing Proposition 2 with the first part of Proposition 3 explains why a natural analog of the Adams consensus method for unrooted trees has never been found: It simply does not exist!

Finally, we remark that although our arguments relied on particular small-scale examples, they can be extended easily to apply to larger numbers of taxa.

ACKNOWLEDGMENTS

We thank the two anonymous referees for some helpful suggestions.

REFERENCES

- ADAMS, E. III. 1986. N-trees as nestings: Complexity, similarity and consensus. *J. Classif.* 3:299–317.
- AHO, A. V., S. YEHOOSHUA, T. G. SZYMANSKI, AND J. D. ULLMAN. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. *SIAM J. Comput.* 10:405–421.
- BATHÉLEMY, J. P., F. R. MCMORRIS, AND R. C. POWERS. 1992. Dictatorial consensus functions on n -trees. *Math. Biosci.* 25:59–64.
- BAUM, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference. *Taxon* 41:3–10.
- BÖCKER, S. 1999. From subtrees to supertrees. Ph.D. thesis, Univ. Bielefeld, Germany.
- BÖCKER, S., A. W. M. DRESS, M. AND M. STEEL. 1999. Patching up X-trees. *Ann. Combin.* 3:1–12.
- BREMER, K. 1990. Combinable component consensus. *Cladistics* 6:369–372.
- CONSTANTINESCU, M., AND D. SANKOFF. 1995. An efficient algorithm for supertrees. *J. Classif.* 12:101–112.
- GORDON, A. G. 1986. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labelled leaves. *J. Classif.* 3:335–348.
- MCMORRIS, F. R. 1985. Axioms for consensus functions on undirected phylogenetic trees. *Math. Biosci.* 74:17–21.
- NG, M. P., AND N. C. WORMALD. 1996. Reconstruction of rooted trees from subtrees. *Discrete Appl. Math.* 69:19–31.
- PURVIS, A. 1995a. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251–255.
- PURVIS, A. 1995b. A composite estimate of primate phylogeny. *Phil. Trans. R. Soc. Lond. B* 348:405–421.
- RAGAN, M. A. 1992. Phylogenetic inference based on matrix representations of trees. *Mol. Phylogenet. Evol.* 1:53–58.
- SANDERSON, M. J., A. PURVIS, AND C. HENZE. 1998. Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol. Evol.* 13:105–109.
- SEMPLE, C., AND M. STEEL. 1999. A supertree method for rooted trees. *Discr. Appl. Math.* (in press).
- STEEL, M. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. *J. Classif.* 9:91–116.
- WILKINSON, M. 1994. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* 43:343–368.

Received 17 March 1999; accepted 16 April 1999

Associate Editor: R. Olmstead