Syst. Biol. 56(3):445–452, 2007

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150701416682

Majority-Rule Supertrees

JAMES A. COTTON[1,2] AND MARK WILKINSON[1]

[1]*Department of Zoology, The Natural History Museum, London SW7 5BD, UK*
[2]*Current Address: Bioinformatics Laboratory, Department of Biology, National University of Ireland Maynooth, County Kildare, Ireland; E-mail: james.cotton@nuim.ie*

Abstract.— Most supertree methods proposed to date are essentially ad hoc, rather than designed with particular properties in mind. Although the supertree problem remains difficult, one promising avenue is to develop from better understood consensus methods to the more general supertree setting. Here, we generalize the widely used majority-rule consensus method to the supertree setting. The majority-rule consensus tree is the strict consensus of the median trees under the symmetric-difference metric, so we can generalize the consensus method by generalizing this metric to trees with differing leaf sets. There are two different natural generalizations, based on pruning or grafting leaves to produce comparable trees, and these two generalizations produce two different, but related, majority-rule supertree methods. [Consensus; phylogeny; symmetric-difference metric; Tree of Life.]

Supertree methods (SMs) take as input a set of phylogenetic trees and return one or more trees (supertrees) that provide a synthesis of the input trees. Supertrees are thus phylogenetic trees inferred from other phylogenetic trees, and alternative supertree methods differ in the way this inference is made (e.g., Eulenstein et al., 2004; Wilkinson et al., 2005a). The problem of providing a synthesis of a set of input trees was first addressed by Adams (1972) for the consensus problem, which is the special case where the input trees have identical leaf sets. Alternative approaches led to the subsequent development of many different consensus methods (CMs) designed for this special case (see Bryant, 2003, for a review).

The supertree problem is a generalization of the consensus problem and the first attempt to develop supertree methods (Gordon, 1986) focused on generalizing from the well-known strict CM (see also Semple and Steel, 2000; Constantinescu and Sankoff, 1995). To be precise, we will say that an SM is a generalization of a particular CM if both methods handle consensus problems identically. For example, Goloboff and Pol (2000) developed a semistrict SM that is a generalization of the semistrict CM (Bremer, 1990) and several other supertree methods are generalizations of less well-known CMs (see Wilkinson et al., 2005a).

Of the various CMs, the majority-rule consensus has proven particularly important because of its use in summarizing bootstrap or jackknife replicates (Felsenstein, 1985), quartet puzzling steps (Strimmer and von Haeseler, 1996), and Bayesian posterior probability distributions on trees. The majority-rule also seems quite natural when the input trees are inferred from independent data as is often the case in supertree construction. However, as yet no SM has yet been proposed as a generalization of the majority-rule CM, and Goloboff and Pol (2002) and Goloboff (2005) doubted that such a generalization is possible. Here, we define two SMs that are alternative generalizations of the majority-rule CM.

PRELIMINARIES

We are concerned with leaf-labeled trees, which are acyclic graphs displaying exclusively branching phylogenetic relationships (see Semple and Steel, 2003, for a more formal definition of a phylogenetic X-tree). Polytomies in these trees are interpreted as soft, representing uncertainty rather than the simultaneous divergence of a set of taxa. Let L be a set of leaves (a leaf set) and L_T denote the leaf set of tree T (i.e., all and only the leaves of T). We call a tuple of trees $P = (t_1,\ldots,t_k)$ a *profile* and denote their leaf sets $L_1,\ldots,L_k$. The leaf set L_P is the union of the leaf sets of all trees in P. We write $T|_L$ for the restriction of tree T to leafset L—i.e., the subtree of T induced by the leaves in L (see Semple and Steel, 2003:110–111). A split is a bipartition of the leaf set, and a (nontrivial) split has at least two taxa in each set (equivalent to a tree with one internal branch). A tree displays a set of compatible splits, and a set of splits is compatible if all the splits could be displayed by a single tree (see Meacham 1983; Semple and Steel, 2003; Wilkinson et al., 2007, for formal definitions) A split is *full* with respect to a tree T if its leaf set is precisely L_T and is *partial* with respect to any more inclusive leaf set. A split is *plenary* with respect to a tuple of trees P if its leaf set is precisely L_P. Note that in the special case of consensus, where all trees have the same leaf sets, there is no difference between full and plenary splits. We will find it useful to further distinguish *majority* splits as those that are displayed by a majority of the input trees. Our examples are all of rooted trees in which the root (though not always shown as such) can be considered an additional leaf. In our application, the initial order of trees within the tuple P does not matter.

THE MAJORITY-RULE CONSENSUS AND BEYOND

The majority-rule CM was introduced by Margush and McMorris (1981). It returns a unique tree that displays exactly (all and only) the majority full splits of a set of input trees. Thus the content of a majority-rule consensus tree

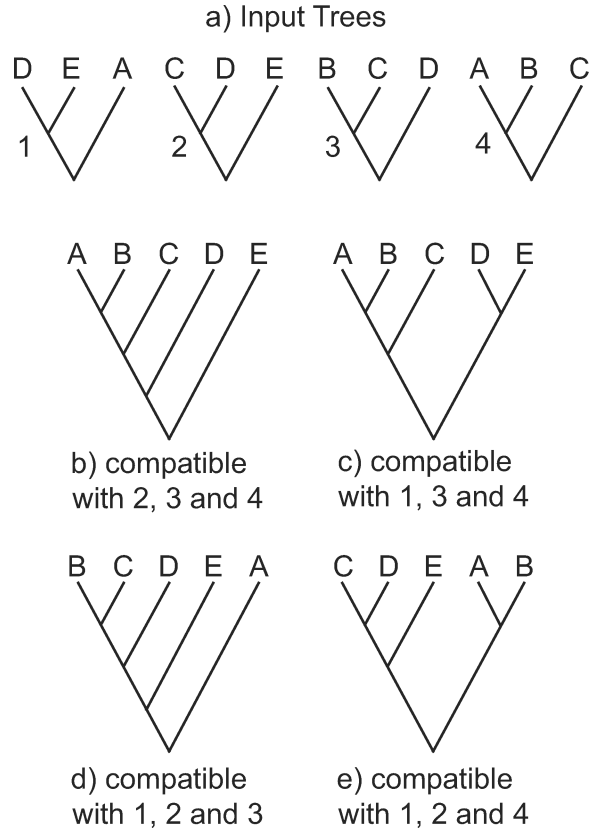

FIGURE 1. Four input trees (a) and four trees (b–e), each of which is entailed by a combination of three input trees. After (relabeled from) Goloboff and Pol (2002).

can be determined simply by counting the frequency of occurrence of the full splits displayed by the input trees. Goloboff and Pol (2002:522) doubted that an equivalent method is possible when input trees have nonidentical leaf sets because "In many cases it is not possible to count how many trees support (or contradict) a group." They illustrated their concerns with a simple example of four input trees (Fig. 1a), any three of which are mutually compatible and jointly entail a different fully resolved tree (Fig. 1b–e) that is incompatible with the fourth input tree. This "shows that a tree may be required to both support a group, or to contradict it, depending on the trees with which it is to be combined." Given that support and conflict, thus construed, are not exclusive, they concluded that "Therefore it is not possible to produce a conceptual equivalent of the majority-rule consensus tree when the trees have different sets of taxa. A method can check on how many input trees a given partition appears only as long as the taxa involved in the partition . . . are present in each and every one of the input trees." Note that it is not that we cannot count how many times a supertree split is displayed by less inclusive input trees (it is always zero) but that this count cannot determine which relationships should be in a majority-rule supertree. More importantly, as we shall show, the counting problem is not an insurmountable difficulty in defining a majority-rule SM.

Many CMs can be defined, as above, in terms of conditions required for relationships to be included in the consensus tree (Bryant, 2003), but alternative characterizations are sometimes possible and useful. Bathélemy and McMorris (1986) showed that the majority-rule consensus tree minimizes the sum of the symmetric-difference metric (the number of full splits present in one but not both of two trees) between it and each of the input trees. The majority-rule consensus tree is thus shown to be a median of the input trees with respect to the (full-split) symmetric-difference metric. When the number of input trees is odd, the majority-rule consensus tree is the unique median tree. If the number of input trees is even, there may be multiple median trees, including the majority-rule tree and resolutions of it that include full splits occurring in exactly half the input trees. Our key observation is that, in this case, the majority-rule consensus is the strict consensus of the median trees.

Although it may be simpler to construct majority-rule consensus trees by counting splits, in principle they can also be found by searching treespace using the sum of the symmetric differences between any candidate consensus tree and each of the input trees as an objective function to distinguish median trees from suboptimal trees and then constructing the strict consensus of the median trees. In seeking an SM that corresponds to the majority-rule CM, we can avoid conceptual difficulties associated with counting splits by generalizing the CM objective function. The full-split symmetric difference is defined only for trees with identical leaf sets, and two different generalizations of the symmetric-difference metric to the case of trees with differing leaf sets lead to two different majority-rule supertree methods.

MAJORITY-RULE(-) SUPERTREES

One means of generalizing to the case where one tree is a supertree and the other an input tree is by comparing the input tree to the subtree of the supertree induced by the leaf set of the input tree (i.e., the supertree pruned of any leaves not in the input tree). These two trees have the same leaf set, so the symmetric difference is defined, and we take this as the symmetric difference of the supertree and input tree. Given as input a profile $P = (t_1, \ldots, t_k)$, the objective function minimized in the majority-rule(-) supertree is

$$\sum_{i=1}^{k} d(T|_{L_i}, t_i)$$

where d is the standard symmetric-difference metric and T ranges over all trees with leaf set L_P. This seems reasonable when we consider that other relationships in the supertree that pertain to leaves not in the input tree seem irrelevant to the comparison. Note also that any input tree full split displayed by a pruned supertree is a partial split displayed and entailed by the supertree. A median supertree with respect to the symmetric-difference

metric is a supertree that minimizes the sum of the symmetric differences between each input tree and the (appropriately pruned) supertree (Bryant, 1997:204). We can now define the *majority-rule(-) supertree* as the strict consensus of the median supertrees. The minus signals that the method depends on pruning leaves to determine the symmetric difference.

Applied to Goloboff and Pol's example (Fig. 1) we see that each of the four trees entailed by some combination of three input trees has a symmetric difference to the fourth input tree (and to the input trees as a whole) of two, which is the minimum. All are construed as median supertrees given our generalization of a symmetric difference, and their strict consensus, which is completely unresolved, is the majority-rule(-) supertree. This lack of resolution is to be expected in this artificial example in which there are no full or partial splits shared by any two input trees—and other supertree methods, such as matrix representation with parsimony (MRP; Baum, 1992; Ragan, 1992), give the same result. A second example, a modification of the first, illustrates a case where the majority-rule(-) supertree is completely resolved (Fig. 2). The four input trees conflict over the relationships of D and E. Tree 1 asserts that D is more closely related to E than to either A or B, whereas trees 2 and 3 assert that D is closer to A and B, respectively, than it is to E. The fourth does not include both D and E and is compatible with all the others. There is a unique median tree in this case with a minimally pruned symmetric difference of two (Fig. 2b, c). This is the majority-rule(-) supertree. Note that the relationships of D and E are resolved, as we might expect, in favor of the two input trees that place D closer to A or B than to E.

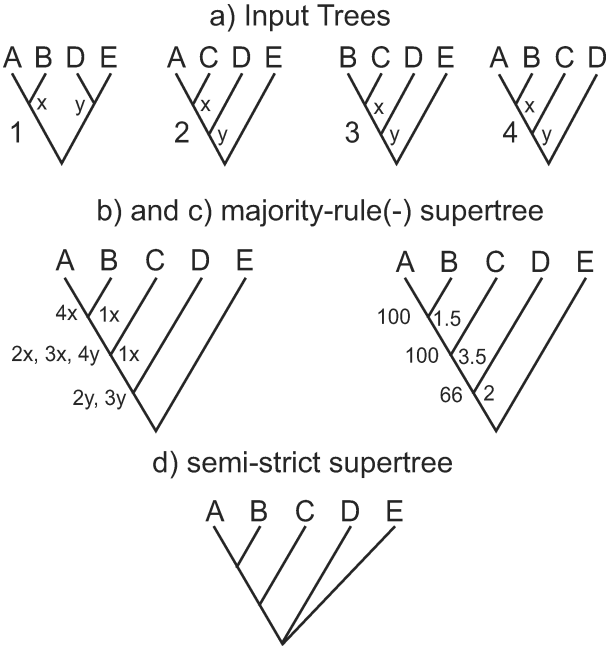

FIGURE 2. Four input trees (a) modified from those in Figure 1 through the addition of a single leaf to each, and their majority-rule(-) supertree showing alternatively (b) the mapping of supportive input tree splits and (c) the number of supportive input trees as a percentage of the input trees that by virtue of their leaf set could have supported the supertree split, and a weighted sum of the support provided by the input trees. (d) The semistrict supertree for these input trees.

Majority-rule consensus trees are usually decorated with the frequencies of their full splits, which greatly enhances their usefulness. There has been some debate about how to measure support for supertree relationships (Bininda-Emonds, 2003; Wilkinson et al., 2005b; Cotton et al., 2006). Here we use the measures of supertree support defined by Wilkinson et al. (2005b), who argued that a plenary split in a supertree is supported by any input tree full split that it entails. We can thus map the input tree splits to the supertree splits they support and for any supertree split we can count s, the number of input trees supporting any supertree split (Fig. 2b). Note that some input tree splits can map to, and support, more than one split in the majority-rule supertree. In the example, tree 1 lacks C, and AB is taken as supporting both AB and ABC in the majority-rule supertree. Wilkinson et al. (2005b) suggest down-weighting such support by the number of supertree splits it is shared among, leading to a weighted sum of the total support (ws) for any supertree split.

In contrast to the consensus case, we expect that in many real supertree analyses some input trees will have leaf sets that make them irrelevant to the support for a supertree split. Thus we suggest expressing s as a percentage of the number of input trees that, by virtue of their leaf sets, could support the supertree split and reporting some measure of absolute support such as ws in addition. Applied to our example, the decorated majority-rule tree (Fig. 2c) shows that AB and ABC have 100% support, but that more input tree splits support ABC (3.5) than AB (1.5). In contrast, ABCD has 66% support, reflecting the two trees that support it and the one tree that contradicts (the other being irrelevant). That the results seem reasonable and intuitive in this case is underlined by the semistrict supertree (Fig. 2d), which includes only the two groups with 100% support and excludes the group ABCD. Note that these measures can be used to decorate any supertree.

MAJORITY-RULE(+) SUPERTREES

The majority-rule(-) SM defined above relies on pruning leaves from the supertree to construct an analogue of the symmetric-difference metric applicable to trees with different, but overlapping, leaf sets. An alternative generalization stems from the reverse operation of grafting missing leaves onto each of the input trees in a set P so as to convert them all into plenary trees (trees with leaf set L_P). We define the binary supertree span $<t>$ of an input tree t to be the set of binary (fully resolved) trees on L_P that display t. In other words, $<t>$ includes all the supertrees that can be produced by grafting any missing leaves (those in P but not in t) onto t and resolving any polytomies in t. Thus, for a k-tuple of trees $P = (t_1, \ldots, t_k)$ we have a k-tuple of their binary supertree spans $Z = (<t_1>, \ldots, <t_k>)$ and we use this to define a representative (because it displays all the input trees) selection,

$R = (T_1, \ldots, T_k)$, where $T_i \in <t_i>$ for $i = 1,\ldots,k$ as any k-tuple comprising precisely one supertree selected from every span in Z. Because all the members of any R display plenary splits we can use the majority-rule consensus method to summarize any R, producing a well-defined and unique *candidate supertree* T_R for each. We are interested in the best representative selections, as judged by the median objective function. Thus we can rank the majority-rule consensus trees (the T_Rs) for the different Rs by their median scores, calling those for which this is minimized the optimal candidate supertrees.

More precisely, given a representative selection $R = (T_1, \ldots, T_k)$ for P, let $s(R)$ denote the median score of R, defined by

$$s(R) = \min_T \sum_{i=1}^{k} d(T, T_i)$$

where T ranges over all trees with leaf set L_P. An *optimal candidate supertree* is the candidate supertree T_R of any representative selection $R = (T_1, \ldots, T_k)$ for P that has the smallest possible median score $s(R)$. We can now define the *majority-rule(+) supertree* as the strict consensus of all the optimal candidate supertrees. The plus sign in the name signals that the method depends on grafting leaves to convert the supertree problem into a consensus problem and distinguishes it from the former method based on pruning.

Figure 3 illustrates the method applied to Goloboff and Pol's example (Fig. 1). We consider four likely candidate trees, each of which displays three of the four input trees (and is thus in each of their supertree spans) and is incompatible with the fourth. In this case, each is the majority-rule consensus tree of a representative selection from the supertree spans of the input trees that comprises the candidate tree, selected (three times) from the spans of the three input trees it displays, and a tree chosen from the span of the incompatible input tree such as to minimize the symmetric difference between it and the candidate tree. The minimal frequency of the full splits in the majority-rule consensus trees is therefore 75% and any differences between them result from differences in how badly the incompatible input tree conflicts with the consensus.

With two of the four candidate trees (Fig. 3b and c), it is possible to graft the missing leaves onto the incompatible input tree so as to produce two of the three full splits in the candidate tree. For one candidate tree (Fig. 3e), only a single common full split can be created by grafting, whereas no common splits can be produced by any grafting for the fourth (Fig. 3d). Thus the first two are the optimal candidate supertrees and the majority-rule(+) supertree is the strict consensus of these, decorated with the lowest of the frequencies of the included splits.

In this case, the majority-rule(+) tree differs substantially from the majority-rule(-) (and other supertrees), which is unresolved. The additional resolution results from considering relationships that are possible (through

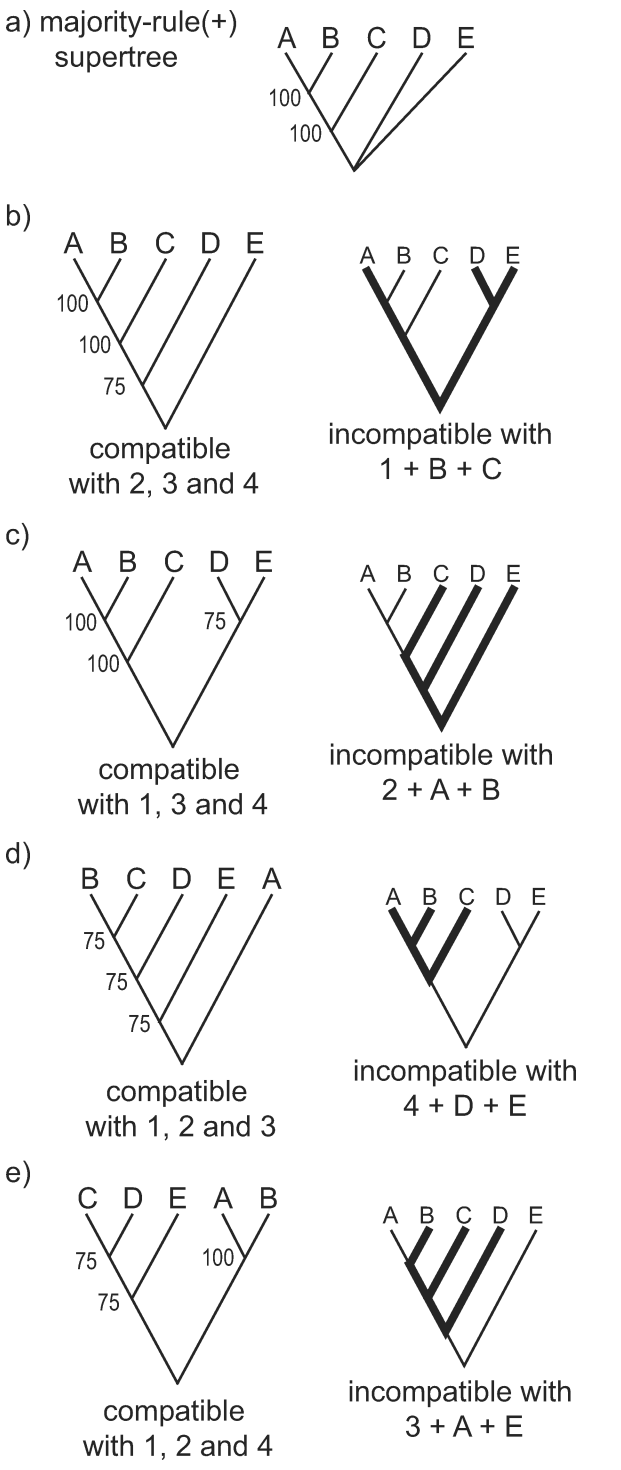

FIGURE 3. Constructing the majority-rule(+) supertree (a) for the four input trees in Figure 1a. Each pair of trees (b–e) corresponds to a representative sample from the spans of the input trees, comprising three copies of the tree entailed by a combination of three input trees (which is also the majority-rule consensus of the selection), and one tree that displays the conflicting input tree (thickened branches) and has the missing leaves grafted on so as to minimize its symmetric difference to the other tree. Numbers indicate the frequency of occurrence of the full splits in the representative selection. Two of the majority-rule consensus trees (b and c) are optimal candidates and the majority-rule(+) supertree is their strict consensus.

alternative graftings) in addition to relationships present in each input tree. In this case, the additional resolution reflects the facts that (1) there is some grafting, and so some representative selection, under which all the input trees display the two groups in the supertree; and (2) there is no alternative grafting, under which all the input trees display different groups, and so does not seem unreasonable. Applied to the second example, both methods return the same supertree but they differ slightly in their decoration (Fig. 2c). The frequency of ABCD is 75%, indicating that three of the four input trees are compatible with this full split. The corresponding decoration on the majority-rule(-) supertree (66%) indicates that two input trees support this full split and that only one other tree could have supported the split (by virtue of its leaf set) but that it does not. The effect of grafting rather than pruning is to allow the otherwise irrelevant input tree to support this split.

Properties of Majority-Rule Supertrees

Based on their definitions, we conjecture that the majority-rule supertrees have the following properties. In the consensus setting, these properties follow from the definition of the majority-rule consensus as the tree containing all and only the majority plenary splits. As we have generalized from the median property of majority-rule consensus, these properties await formal proof (or disproof) for our more general methods:

1. All majority plenary splits are in the majority-rule tree. This property is familiar in the consensus case where all input trees display only plenary splits. More generally, input trees may display no plenary splits (as in our examples), in which case the property is uninteresting, but whenever such majority plenary splits exist they will be in the majority-rule supertree. Such cases may arise in small-scale supertree studies of genomic data (e.g., Creevey et al., 2004).
2. The majority-rule tree is compatible with any majority partial splits. The majority-rule CM does not display all majority partial splits (only those entailed by majority full splits), but it is compatible with any majority partial splits. In the supertree case there may be no majority partial splits but if there are, the majority-rule tree is expected to be compatible with them, as in the consensus case.
3. Although a split in the majority-rule trees may not be displayed by any input tree, all splits in the majority-rule tree are compatible with a majority of the input trees. Note that we cannot expect every split that is compatible with a majority of input trees to be in the majority-rule tree even in the consensus case, as the set of such splits may be incompatible.
4. Every plenary split in the majority-rule tree entails at least one input tree full split. Without support, in the sense of Wilkinson et al. (2005b), from one or more input trees, a split cannot occur in the majority-rule tree. The converse does not hold, so not every plenary split that entails at least one input tree full split will be in the majority-rule tree. The set of such splits may be incompatible.

Additionally, for the majority-rule(-) supertree, support (sensu Wilkinson et al., 2005b) must be greater than conflict, so that every plenary split in the majority-rule tree must entail splits in more input trees than it conflicts with; i.e., in a majority of the relevant trees. It is not clear whether this also holds for majority-rule(+) supertrees, where otherwise irrelevant trees can be interpreted as either supporting or conflicting a given supertree split, depending on the other input trees.

The two methods differ in their treatment of polytomies. The objective function of the minus method penalizes the supertree if it resolves a group permitted by an input tree polytomy, whereas in the plus method this carries no cost, which seems more reasonable in the supertree setting (Page, 2002). This difference does not explain their very different behaviors with the Golobof and Pol (2002) example (Figs. 1 and 3), which instead reflects the different treatment of missing leaves.

In common with many other supertree methods (Wilkinson et al., 2005a), we expect the majority-rule(-) method to be quite sensitive to input tree size, given that larger trees have more splits and a greater potential contribution to the sum of the full-split symmetric differences. One way to reduce this effect might be to use a normalized symmetric-difference score (Robinson and Foulds, 1981) between the pruned supertree and input trees but this would result in splits in different trees having different weights. Both majority-rule methods can be readily extended to employ differential weights of trees and of their splits by defining a weighted sum of the symmetric differences. Using normalized scores or other differential weighting, properties 1, 2, and 3 would not generally hold. The representation of input trees by supertrees ensures that in the majority-rule(+) method, all input trees potentially have equal weight. However, the best representative selections will be ones in which missing leaves are grafted to input trees so as to best represent their relationships as evidenced by the other input trees, and we would expect larger input trees to have a greater impact on this. In the consensus case, the majority-rule is also a median tree, but this does not hold more generally.

An Empirical Example

As an empirical example, we have constructed the majority-rule supertrees for five molecular phylogenies of *Drosophila*, extracted from larger studies of *Drosophila* phylogeny. More complete versions of these trees were analyzed with a range of supertree methods in Cotton and Page (2005). The input trees are shown in Figure 4. The majority-rule(+) and majority-rule(-) trees for this example are identical. They include only the clade containing *D. melanogaster, D. sechellia,* and *D. simulans* and its resolution and have the same support values (Figure 5). The majority-rule(-) supertree is the strict consensus of 79 median trees, whereas there are 43 median trees under the majority-rule(+) objective

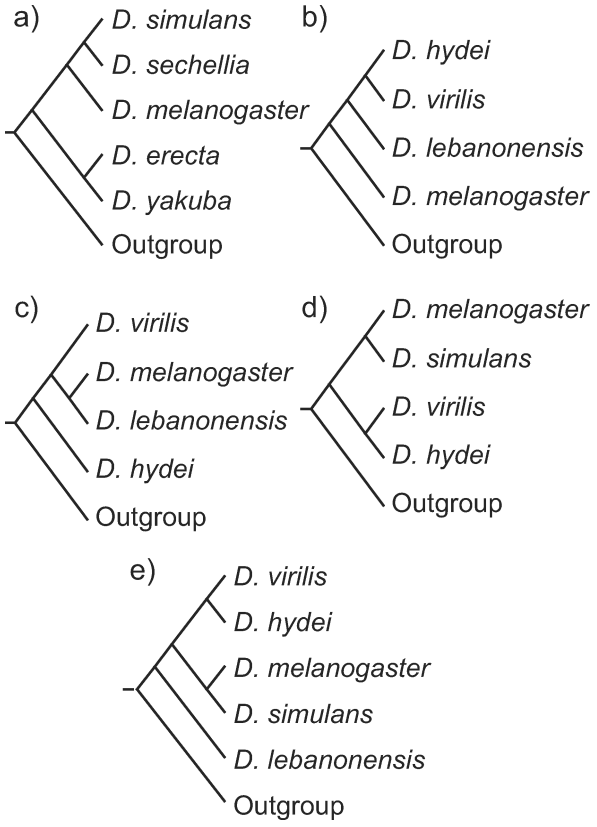

FIGURE 4. Input trees for the *Drosophila* example. Each tree is extracted from a wider study of *Drosophila* phylogeny from the following genes and sources. (a) *Roughex* (Avedisov et al., 2001). (b and c) *Alcohol dehydrogenase* and *alcohol dehydrogenase-related*, respectively (Bertrán and Ashburner, 2000). (d) *Dopa decarboxylase* (Tatarenkov et al., 1999). (e) *Cu-Zn superoxide dismutase* (Kwiatowski et al., 1994).

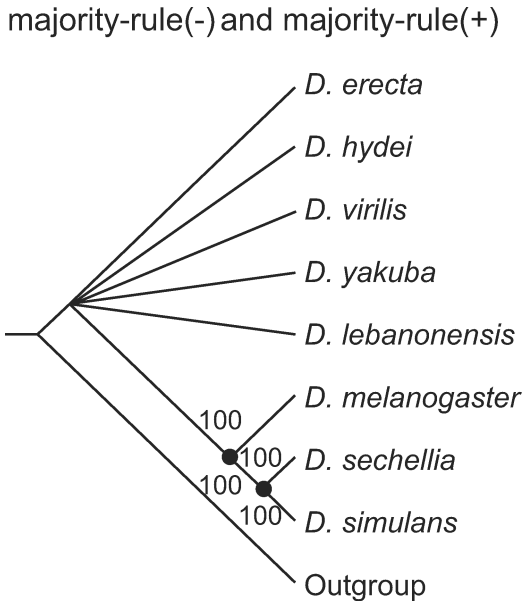

FIGURE 5. Majority-rule supertrees for *Drosophila* species. The majority-rule(-) and majority-rule(+)supertrees are both identical and show identical decoration (above and below nodes, respectively) for this data set.

function, all of which are also median trees under the majority-rule(-) criterion. The strict consensus of the 77 standard MRP supertrees is also identical to this tree, whereas the strict consensus of 10,200 Purvis MRP supertrees is unresolved. In this case, the majority-rule methods perform as well as standard MRP, and the general lack of resolution in the supertrees results both from incongruence and from a lack of effective overlap in the input trees (see Wilkinson and Cotton, 2006).

FINDING MAJORITY-RULE SUPERTREES

Although not our prime concern, some comments on the implementation of the majority-rule SMs is warranted. Majority-rule(-) supertrees can be approximated with heuristic searches of tree space as are used for many other supertree methods and taking the strict consensus of all equally optimal candidate supertrees. Clann (Creevey and McInerney, 2004) implements a number of methods that use objective functions based on comparisons of input treees with pruned supertrees. Bryant (1997) has shown that finding a single median tree under the pruned symmetric-difference metric is NP-complete, and this implies that finding our majority-rule(-) supertree is also.

There can be many representative selections of a set of input trees, too many for majority-rule(+) supertrees to be found through their exhaustive enumeration. The objective function is minimized when there is a lot of agreement in a representative selection; i.e., when the addition of leaves is maximally consistent with information on their relationships in the other input trees. Thus we might be able to use this information to efficiently construct optimal or near optimal representative selections. Doubtless other, better methods that do this await invention, but we could, for example, use a greedy polynomial time method such as quartet joining (Wilkinson and Cotton, 2006) to construct representative selections from the input trees and repeat this with as many different starting trees and orders of adding missing leaves as we desire in the search for the optimal candidates. In the worst case, there are exponentially many representative selections, so finding the majority-rule(+) supertree is likely to have at least exponential complexity.

DISCUSSION

The majority-rule CM of Margush and McMorris (1981) is the most widely used CM in systematics and a reasonable choice when input trees are independent. Thus, in searching for a useful supertree method it is natural to look for generalizations of the majority-rule CM. Goloboff and Pol (2002) suggested that a meaningful generalization is not possible, but we have defined two nontrivial generalizations that overcome the difficulties they identified. As Goloboff and Pol (2002) and Goloboff (2005) acknowledge, a majority-rule SM seems ideal for a range of important applications. In particular, Goloboff (2005) takes the widely used MRP method to be an attempt at something like a majority-rule supertree,

and he highlights that it does not behave how we might expect from a majority-rule supertree method. That the definition of majority-rule SMs is not as problematic as Goloboff and Pol (2002) supposed only strengthens Goloboff's (2005) critique of other supertree methods. The majority-rule methods we have defined appear more similar to (though not equivalent to) the split-fit or matrix representation with compatibility (Rodrigo, 1996) method than to parsimony-based methods.

Many CMs can be defined in a number of alternative, equivalent ways: in terms of conditions for splits to be included, in terms of algorithms used to construct them, and, in some cases, in terms of the objective function a consensus tree optimizes. Our treatment highlights that alternative generalizations from consensus to supertree methods based on these different views of consensus may be possible and that some may be more fruitful than others. In particular, we emphasize the potential use of pruning or grafting leaves to achieve consensus comparisons and offer some insight into how and why they might differ. We note, in passing, that although the objective functions defined here are well-defined for comparing a supertree with an input tree (a tree on a subset of the whole leafset), they are not well defined metrics on the set of all trees and may not be useful generalizations of the symmetric-difference metric in other settings.

Majority-rule supertrees are not expected to include the unsupported groups for which MRP, MinFlip, and other SMs have been criticized (e.g., Pisani and Wilkinson, 2002) and because the two objective functions are symmetric we do not expect the method to be biased with respect to input tree shape (Wilkinson et al., 2005a). We therefore consider the majority-rule supertree methods we have defined to be promising approaches to supertree construction that merit further study of their behavior with real examples and thorough simulation. It might be anticipated that majority-rule supertrees will sometimes be too conservative and poorly resolved and that there will sometimes be partial splits that are well supported in the absence of any well-supported plenary splits. Thus, extending the approach to include compatible minority splits and to find such well-supported partial splits (Wilkinson, 1996) would be useful areas of further work. We would expect more liberal methods to extend rather than conflict with majority-rule trees. It might also be instructive to investigate supertree methods more widely in the better understood context of consensus methods (Day and McMorris, 2003; Wilkinson et al., 2007).

ACKNOWLEDGEMENTS

The concept of what we have termed a representative selection from the supertree span of a set of input trees is due to David Bryant, who used this notion in a presentation on strict consensus supertrees at the DIMACS 2003 meeting on bioconsensus. We acknowledge the financial support of BBSRC 40/G18385, the stimulating atmosphere fostered at the DIMACS meeting, and most especially David Bryant for providing both inspiration and conceptual tools needed for a solution of our problem. STsupport, a program to calculate measures of support for supertrees, is available at http://taxonomy.zoology.gla.ac.uk/~jcotton/software.html. We thank David Bryant, Barbara Holland, Buck McMorris, Rod Page, Claire Pickthall, Davide Pisani, and particularly Mike Steel for comments on the manuscript.

REFERENCES

Adams, E. N. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. 21:390–397.

Avedisov, S. N., I. B. Rogozin, E. V. Koonin, and B. J. Thomas. 2001. Rapid evolution of a cyclin A inhibitor gene, *roughex*, in Drosophila. Mol. Biol. Evol. 18:2110–2118.

Barthélemy, J. P., and F. R. McMorris. 1986. The median procedure for n-trees. J. Classif. 3:329–334.

Baum, B. R. 2002. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

Betrán, E., and M. Ashburner. 2000. Duplication, dicistronic transcription, and subsequent evolution of the *alcohol dehydrogenase* and *alcohol dehydrogenase-related* genes in Drosophila. Mol. Biol. Evol. 17:1344–1352.

Bininda-Emonds, O. R. P. 2003. Novel versus unsupported clades: Assessing the qualitative support for MRP supertrees. Syst. Biol. 52:839–848.

Bremer, K. 1990. Combinable component consensus. Cladistics 6:369–372.

Bryant, D. 1997. Building trees, hunting for trees and comparing trees. PhD thesis. Department of Mathematics, University of Canterbury, New Zealand.

Bryant, D. 2003. A classification of consensus methods for phylogenetics. Pages 163–183 *in* Bioconsensus (M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts, eds.), DIMACS series in discrete mathematics and theoretical computer science. American Mathematical Society, Providence, Rhode Island.

Consantinescu, M., and D. Sankoff. 1995. An efficient algorithm for supertrees. J. Classif. 12:101–112.

Cotton, J. A., and R. D. M. Page. 2004. Tangled trees from molecular markers: reconciling conflict between phylogenies to build molecular supertrees. Pages 107–125 *in* Phylogenetic supertrees: Combining information to reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Cotton, J. A., C. S. C. Slater, and M. Wilkinson. 2006. Discriminating supported and unsupported relationships in supertrees using triplets. Syst. Biol. 55:345–350.

Creevey, C. J., D. A. Fitzpatrick, G. K. Philip, R. J. Kinsella, M. J. O'Connell, M. M. Pentony, S. A. Travers, M. Wilkinson, and J. O. McInerney. 2004. Does a tree-like phylogeny exist only at the tips in the Prokaryotes? Proc. R. Soc. B 271:2552–2558.

Creevey, C. J., and J. O. McInerney. 2004. Clann: Investigating phylogenetic information using supertree analyzes. Bioinformatics 21:390–392.

Day, W. H. E., and F. R. McMorris. 2003. Axiomatic consensus theory in group choice and biomathematics. Frontiers in applied mathematics, volume 39. Society for Industrial and Applied Mathematics, Philadelphia.

Eulenstein, O., D. Chen, J. G. Burleigh, D. Fernández-Baca, and M.J. Sanderson. 2004. Performance of flip supetree construction with a heuristic algorithm. Syst. Biol. 53:299–308.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Goloboff, P. A. 2005. Minority-rule supertrees? MRP, Compatibility, and MinFlip may display the least frequent groups. Cladistics 21:282–294.

Goloboff, P. A., and D. Pol. 2002. Semistrict supertrees. Cladistics 18:514–525.

Gordon, A. 1986. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. J. Classif. 3:335–348.

Kwiatowski, J., D. Skarecky, K. Bailey, and F. J. Ayala. 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide-sequence of the Cu,Zn Sod gene. J. Mol. Evol. 38:443–454.

Margush, T., and F. R. McMorris. 1981. Consensus *n*-trees. Bull. Math. Biol. 43:239–244.

Meacham, C. A. 1983. Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. Pages 304–314 *in* Numerical taxonomy (J. Felsenstein, ed.), NATO ASI series, volume G1. Springer-Verlag, Berlin.

Page, R. D. M. 2002. Modified mincut supertrees. Lect. Notes Comput. Sci. 2452:537–551.

Pisani, D., and M. Wilkinson. 2002. MRP, total evidence and taxonomic congruence. Syst. Biol. 51:151–155.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representations of trees. Mol. Phylogenet. Evol. 1:53–58.

Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

Rodrigo, A. G. 1996. On combining cladograms. Taxon 45:267–274.

Semple, C., and M. Steel. 2000. A supertree method for rooted trees. Discr. Appl. Math. 105:147–158.

Semple, C., and M. Steel. 2003. Phylogenetics. Oxford University Press, Oxford, UK.

Strimmer, K., and A. von Haesseler. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

Tatarenkov, A., J. Kwiatowski, D. Skarecky, E. Barrio, and F. J. Ayala. 1999. On the evolution of Dopa decarboxylase (Ddc) and Drosophila systematics. J. Mol. Evol. 48:445–462.

Wilkinson, M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. Mol. Biol. Evol. 13:437–444.

Wilkinson, M., and J. A. Cotton. 2006. Supertree methods for building the tree of life: Divide-and-conquer approaches to large phylogenetic problems. Pages 61–76 *in* Towards the Tree of Life: Taxonomy and Systematics of large and species rich taxa (T. Hodkinson, J. Parnell, and S. Waldren, eds.). CRC Press, Boca Raton, Florida.

Wilkinson, M., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F.-J. Lapointe, C. Levasseur, J. O. McInerney, D. Pisani, and J. L. Thorley. 2005a. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. Syst. Biol. 54:419–431.

Wilkinson, M., J. A. Cotton, F.-J. Lapointe, and D. Pisani. 2007. Properties of supertree methods in the consensus setting. Syst. Biol. 56:330–337.

Wilkinson, M., D. Pisani, J. A. Cotton, and I. Corfe. 2005b. Measuring support and finding unsupported relationships in supertrees. Syst. Biol. 54:823–831.

First submitted 25 August 2006; reviews returned 15 October 2006; final acceptance 31 January 2007

Associate Editor: Mike Steel