The Shape of Supertrees to Come:
Tree Shape Related Properties of Fourteen Supertree Methods

Mark Wilkinson,[1] James A. Cotton,[1] Chris Creevey,[2] Oliver Eulenstein,[3] Simon R. Harris,[1,4]
Francois-Joseph Lapointe,[5] Claudine Levasseur,[5] James O. McInerney,[2] Davide Pisani,[1]
and Joseph L. Thorley[6]

[1]*Department of Zoology, The Natural History Museum, London SW7 5BD, United Kingdom;*
E-mail: marw@nhm.ac.uk (M.W.)
[2]*Department of Biology, National University of Ireland, Maynooth, County Kildare, Ireland*
[3]*Department of Computer Science, Iowa State University, Ames, Iowa 50011-1040, USA*
[4]*Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, United Kingdom*
[5]*Département de sciences biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville,*
Montréal (Québec), H3C 3J7, Canada
[6]*Fisheries Research Services, Freshwater Laboratory, Faskally,*
Pitlochry, Perthshire PH16 5LB, United Kingdom

Abstract.—Using a simple example and simulations, we explore the impact of input tree shape upon a broad range of supertree methods. We find that input tree shape can affect how conflict is resolved by several supertree methods and that input tree shape effects may be substantial. Standard and irreversible matrix representation with parsimony (MRP), MinFlip, duplication-only Gene Tree Parsimony (GTP), and an implementation of the average consensus method have a tendency to resolve conflict in favor of relationships in unbalanced trees. Purvis MRP and the average dendrogram method appear to have an opposite tendency. Biases with respect to tree shape are correlated with objective functions that are based upon unusual asymmetric tree-to-tree distance or fit measures. Split, quartet, and triplet fit, most similar supertree, and MinCut methods (provided the latter are interpreted as Adams consensus-like supertrees) are not revealed to have any bias with respect to tree shape by our example, but whether this holds more generally is an open problem. Future development and evaluation of supertree methods should consider explicitly the undesirable biases and other properties that we highlight. In the meantime, use of a single, arbitrarily chosen supertree method is discouraged. Use of multiple methods and/or weighting schemes may allow practical assessment of the extent to which inferences from real data depend upon methodological biases with respect to input tree shape or size. [Consensus; parsimony; phylogeny; tree similarity; Tree of Life.]

There is considerable interest in phylogenetic supertrees. To date, most practical works have used matrix representation with parsimony (MRP) to combine the information in a set of input trees with nonidentical leaf sets into a larger-scale phylogeny including all the leaves (Baum, 1992; Ragan, 1992; Sanderson et al., 1998; Bininda-Emonds et al., 2002). More generally, supertree methods take as input a set of phylogenetic trees and return one or more phylogenetic supertrees that represent the input trees or an inference based upon them. Characterized this way, supertree methods include the consensus methods developed for use in the special case where all input trees have the same leaf set (Steel et al., 2000), and the supertree problem is a generalization of the consensus tree problem (Semple and Steel, 2000).

Strict and semistrict supertree methods (e.g., Gordon, 1986; Steel, 1992; Lanyon, 1993; Constantinescu and Sankoff, 1995; Goloboff and Pol, 2002), like their consensus namesakes, output trees that do not conflict with any input trees. These methods are conservative in that they do not resolve conflicts. In contrast, most supertree methods are, like MRP, more liberal in that they are capable of resolving or reconciling conflicts among the input trees.

MRP has been advocated because of its potential to produce more comprehensive and well-resolved phylogenies more efficiently than by assembling and analyzing ever larger supermatrices (Sanderson et al., 1998;

Bininda-Emonds et al., 2002). Unfortunately, the properties of most liberal supertree methods, and thus the suitability of the supertrees they yield as stepping-stones to the Tree of Life, are only poorly understood. It has been shown that some methods perform well in simulations (e.g., Bininda-Emonds and Sanderson, 2001; Chen et al., 2003; Lapointe and Levasseur, 2004), but these studies have only just begun to provide useful comparisons of alternative methods (Eulenstein et al., 2004). The nascent methodological literature includes discussion of desirable properties of supertree methods (e.g., Bininda-Emonds and Bryant, 1998; Semple and Steel, 2000; Pisani and Wilkinson, 2002; Wilkinson et al., 2004), of possible biases with respect to tree size (e.g., Purvis, 1995a; Ronquist, 1996) and shape (Wilkinson et al., 2001), and fundamental limitations on all supertree methods (Steel et al., 2000). With a steady increase in supertree methods, there is increasing need for comparative study of the methods.

Here we use a simple example to investigate the impact of input tree shape (balance, symmetry) on how conflict is resolved by 14 liberal supertree methods. We contend that resolution of conflict should be independent of tree shape and ask whether methods ever resolve conflict in favor of relationships in more or less balanced trees in the absence of any evidential basis for resolving conflict in that way. We argue that input tree shape effects (ITSEs) revealed by our example are related to asymmetric tree-to-tree distances in the objective functions of

some supertree methods. In considering the issues raised by our limited investigations, we make a number of suggestions for the future of supertree construction.

RELATIONSHIPS

Relationships in trees are given by internal branches, each of which gives a full split of the leaves into two nontrivial disjoint subsets. For simplicity (and because not all methods we examine are applicable to unrooted trees) we consider only rooted trees. In rooted trees, cladistic relationships are expressed in terms of some leaves (terminal taxa, OTUs) being more closely related to (sharing a more recent common ancestry with) each other than to (with) some other leaves (see, e.g., Wilkinson, 1994a). An internal branch splits the members of a clade from the nonmembers (including the root), and we refer to such relationships as components. Components may entail less inclusive relationships. The irreducible cladistic relationship is the resolved triplet (three-taxon statement), which places two leaves closer to each other than a third (in practice a resolved triplet is a resolved quartet because a fourth leaf, representing the root, is implied or included).

Rooted trees can be thought of as sets of resolved triplets, or as sets of more inclusive relationships, most commonly the components or the sister-group relationships that they display. Components, like trees, can be thought of as composite hypotheses of relationships that can be built up from the irreducible relationships among triplets of leaves. Trees can also be conceived of as comprising sets of additive path-length distances between pairs of leaves that jointly entail the relationships in the tree.

A different notion of relationship, due to Adams (1986), is subset nesting (nesting for short). For a set of taxa S, a nesting is a subset of taxa that have a more recent last common ancestor than the last common ancestor of S. The cladistic information conveyed if we know that A and B nest within A, B, C, and D is that A and B are more closely related to each other than they are either to C or to D or to both, i.e., an ambiguous combination of otherwise unambiguous cladistic relationships. Adams consensus trees represent common nestings and

the polytomies they contain do not have an unambiguous cladistic interpretation (see Wilkinson, 1994a).

SUPERTREE METHODS

We explored 14 distinct supertree methods or important methodological variants. The methods are summarized in Table 1 and described here. In the MRP approach to supertree construction of Baum (1992) and Ragan (1992), referred to here as *standard MRP*, binary coding of the components of each input tree is used to generate a "pseudocharacter" matrix representation (MR) of the tree (Farris, 1973). The pseudocharacters or matrix elements for all source trees are combined, with leaves that are not present in a given tree scored as missing entries in the matrix elements for that tree, and the combined matrix is analyzed with reversible (Fitch or Wagner) parsimony to produce one or more most parsimonious MRP supertrees. *Irreversible MRP* (Bininda-Emonds and Bryant, 1998) differs only in its use of irreversible (Camin-Sokal) parsimony. *Purvis MRP* (Purvis, 1995a) uses reversible parsimony but matrix elements represent sister-group relationships rather than components. There is one matrix element for each clade and it distinguishes the members of the clade from the members of its sister group (or of all possible sister groups in the case of polytomies) and the root (MRP outgroup), with any other leaves scored as missing.

The *MinFlip* supertree method as used thus far also uses the component MR (Chen et al., 2003; Eulenstein et al., 2004). Conflict can be removed from the combined MR by flipping the scores of individual matrix cells (i.e., from 0 to 1 or vice versa). MinFlip supertrees are those corresponding to matrices in which conflict has been removed with a minimum number of flips. Several authors have suggested analyzing component MRs with compatibility (specifically, clique analysis) in place of parsimony (e.g., Purvis, 1995b; Rodrigo, 1996; Pisani, 2002). We call this approach *split fit*. The objective function maximized by split fit supertrees is the number of matrix elements (components) entailed or displayed by the supertree (i.e., that fit the supertree with no extra steps or "homoplasy").

TABLE 1.   Supertree methods examined in this work. $+$, $-$ indicate objective functions that are maximized or minimized, respectively.

Method	Representation	Objective function
1. Standard MRP	Components	Fitch parsimony fit of composite matrix to supertree ($-$)
2. Irreversible MRP	Components	Camin-Sokal parsimony fit of composite matrix to supertree ($-$)
3. Purvis MRP	Sister groups	Fitch parsimony fit of composite matrix to supertree ($-$)
5. MinFlip	Components	Number of flips needed to render all input trees compatible with the supertree ($-$)
4. Split fit	Components	Number of matrix elements entailed by the supertree ($+$)
6. Triplet fit	Resolved triplets	Number of matrix elements entailed by the supertree ($+$)
7. Quartet fit	Resolved quartets	Number of matrix elements entailed by the supertree ($+$)
8. Average consensus	EBL distances (EBL)	Least squares fit of average matrix to supertree ($-$)
9. Average dendrogram	Ultrametric distances	Least squares fit of average matrix to ultrametric supertree ($-$)
10. MSS	EBL distances	Weighted sum of absolute differences between each tree and corresponding pruned supertree
11. MinCut	—	—
12. Modified MinCut	—	—
13. GTP (DL)	—	Number of duplications and losses needed to reconcile all input trees with the supertree
14. GTP (D)	—	Number of gene duplications to needed reconcile all input trees with the supertree

MRs in which each resolved triplet in the tree contributes a matrix element, with other leaves scored as missing, can be analyzed with parsimony (reversible or irreversible), compatibility (Wilkinson, 1994b), and MinFlip methods. All these approaches seek supertrees that display the maximum number of resolved triplets in the composite MR. This approach has been called three-item consensus (Nelson and Ladiges, 1994) and triplet MRP (Wilkinson et al., 2001), and is referred to here as *triplet fit*. *Quartet fit* (= quartet MRP; Wilkinson et al., 2001) differs only in the MR, which includes all quartets rather than only those that include the root.

A variety of supertree methods employ distance MRs of trees. The *average consensus* uses least-squares analysis of the matrix of average pairwise path-length distances (an average distance MR of the trees), with missing values that may arise in the supertree context estimated from the available distances (Lapointe and Cucumel, 1997). The average consensus procedure is the only method considered here that uses branch length information when it is available, whereas the examples considered here are cladograms. To implement the average consensus, we calculated path lengths assuming all branches have length of unity (Lapointe and Levasseur, 2004) and refer to these as equal branch length (EBL) distances. A variant, which we call the *average dendrogram* method, was implemented by treating the rooted input trees as ultrametric matrices (see Lapointe and Legendre, 1995) and using a least-squares algorithm that imposes a molecular clock. The *most similar supertree* method (MSS; Creevey et al., 2004) also uses matrix representations of EBL path-length distances, but the optimal tree is that which minimizes a weighted sum of the absolute differences in path lengths between the supertree and each of the input trees. This is determined by pruning the supertree of irrelevant leaves and collapsing redundant branches for each comparison, with the scores for each tree divided against the number of pairwise distances to account for input tree size.

Given a set of compatible input trees, one or more supertrees exist that display all the relationships in all the input trees. The Aho et al. (1981) algorithm returns the Adams consensus of the set of such supertrees (Bryant, 1997), which it finds by building a graph representing all the nestings in the input trees in which each vertex is a leaf, and an edge connects two vertices if the leaves are nested in any input tree. If these nestings are compatible, this graph is disconnected, and the components of the graph are grouped in the output tree. An iterative procedure of rebuilding the graph after restricting the input trees to the appropriate leaves then resolves relationships within each component. The *MinCut* (Semple and Steel, 2000) supertree method is an extension of the Aho et al. algorithm to deal with input tree conflict. Semple and Steel (2000) simply suggest that if the nestings graph is fully connected, it be disconnected by making all the cuts in any minimal cut-set of the graph. Page's (2002) *modified MinCut* method differs in attempting to ensure that uncontradicted relationships in the input trees are present in the output trees.

These MinCut methods have a number of desirable properties, not least that they can be computed in polynomial time. They yield Adams consensus-like supertrees, which include any nestings that are common to the input trees, but unlike Adams consensus trees they are not limited to this information (Semple and Steel, 2000). The Adams-like properties call into question whether the clusters in MinCut supertrees should be interpreted as components, as in other supertrees, or as nestings, as in Adams consensus trees, or as some ambiguous mixture of the two. We prefer to interpret MinCut supertrees conservatively, like Adams consensus trees, but we expect some practitioners will interpret them as any other supertree and thus we explore both interpretations.

Gene tree parsimony (GTP; Slowinski and Page, 1999) methods depend on the idea that incongruence between two gene trees could be due to a limited number of molecular events, such as gene duplication and subsequent gene loss or lateral gene transfer. We can use the number of these events needed to explain the difference between input trees as an optimality criterion for choosing between supertrees. GTP defines the best supertrees as those that imply the minimum number of cophylogenetic events—for ease of computation, restricted to either gene duplications alone (D) or both gene duplications and gene losses (DL)—on the input trees. Parsimony-based reconciled tree methods are used to infer the events (Page and Charleston, 1997; Slowinski and Page, 1999). Unlike other supertree methods, GTP invokes specific biological explanations for incongruence between input trees (Cotton and Page, 2004).

Methods

The (dis)similarities of pairs of trees were quantified with the symmetric difference metric (SD; Robinson and Foulds, 1981) and explicitly agree (EA$_T$; Thorley and Wilkinson, 2000). The former is the sum of the components present in one but not both trees and is negatively correlated with similarity. The latter is the proportion of triplets that are resolved identically in the two trees, and was originally defined as a distance as for quartets by Estabrook et al. (1985). As used here, it is positively correlated with similarity and, because there are more resolved triplets than components in most trees, it is a potentially more discerning measure than SD, and one that is not as dramatically affected by instability in a single leaf.

EA$_T$ values were determined and summarized using a program written by SRH. Pseudocharacter MRs were prepared and some EA$_T$ values were determined with RadCon (Thorley and Page, 2000). PAUP* (Swofford, 1998) was used for exact (Branch and Bound) Fitch and irreversible parsimony (MRP, triplet and quartet fit) analyses, to construct consensus trees, and to determine SD. All matrix elements were weighted equally and zero-length branches were not collapsed, so that all supertrees were binary. MinFlip supertrees were constructed with D. Chen's heuristic supertree software (http://genome.cs.iastate.edu/CBL/download/).

MSS analysis was implemented with CC's CLANN software (Creevey and McInerney, 2004). Split fit was implemented with CLANN and with the MIX program of PHYLIP (Felsenstein, 1993). Quartet fit was implemented with CLANN and PAUP, and triplet fit with PAUP. All CLANN analyses used heuristic searches with 100 Random addition sequences and SPR branch swapping. Average consensus and dendrograms were constructed with the FITCH and KITCH programs of PHYLIP (Felsenstein, 1993), respectively, after computation of average distance matrices. MinCut and modified MinCut supertrees were constructed with R. Page's supertree program (http://darwin.zoology.gla.ac.uk/cgi-bin/supertree.pl). Heuristic GTP analyses were performed with GeneTree (Page, 1998) using 100 random starting points and 100 searches with alternating NNI and SPR branch swapping. Where methods return more than one supertree, the strict component consensus of the set of supertrees is referred to specifically as the consensus supertree. All other reference to supertrees is to optimal supertrees found using heuristic or exact searches.

For methods where our example revealed an ITSE we further tested for a bias with respect to tree shape by building supertrees for the unbalanced input tree of Figure 1 and for random, balanced input trees con-

structed by randomly permuting the labels of the original balanced example. For each test, 500 random permutations were used, and the difference in mean EA_T between the output trees and each of the two input trees calculated, giving an indication of any preference for relationships in balanced or unbalanced input trees. For methods with an optimality criterion we also calculated the asymmetry in the optimality criterion, when used to fit the unbalanced input tree onto the balanced tree or vice versa. Thus, for example, for MRP analyses this was measured as the difference in minimum number of parsimony steps for the MR of the unbalanced input tree fitted onto the balanced tree and the matrix for the balanced tree fitted onto the unbalanced tree.

THE EFFECT OF INPUT TREE SHAPE

Based on two empirical examples, Wilkinson et al. (2001) suggested that standard MRP may be biased with respect to tree shape, so that, in cases of conflicting input trees, the method favors relationships in more unbalanced trees. We contend that if supertrees are used to resolve conflict in input trees, then such resolution should be based on some reasonable assessment of the relative weight of evidence for the conflicting relationships. We further contend that, in the absence of any explicit justification for considering it otherwise, tree shape should be considered an irrelevant variable in this context and should exert no influence in the resolution of conflict. In the terminology of Wilkinson et al. (2004), we would like to have supertree methods that are "shapeless," for which input tree shape plays no part in the resolution of conflict.

In order to investigate whether supertree methods are shapeless we confronted them with a simple contrived example (Fig. 1). The two input trees have the same leaf sets, and thus represent the special case of a consensus problem. Given that properties, and our expectations, are mostly much better understood for consensus than for supertree methods, it can be helpful to investigate how the latter handle consensus problems. It allows direct comparison with consensus methods, and, importantly here, if methods do have biases with respect to input tree size, as has been much discussed (Purvis, 1995a; Ronquist, 1996; Bininda-Emonds and Bryant, 1998; Page, 2002), then analysis of trees with the same leaf set provides the opportunity to investigate other properties free from the confounding influence of tree size. Because our input trees are also fully resolved, there is no potentially confounding difference in their cladistic information content (Thorley et al., 1998). We contend that because there are just two equal sized and equally resolved input trees, there is no evidential basis for resolving any conflicts between them. Conflicts are substantial. There are no components in common (SD = 28), and about one third of the resolved triplets are shared ($EA_T = 0.33$). Of the most commonly used consensus methods (Adams and strict, semistrict, and majority-rule component), only the Adams (Fig. 2) is not completely unresolved. With a pair of such highly incongruent input
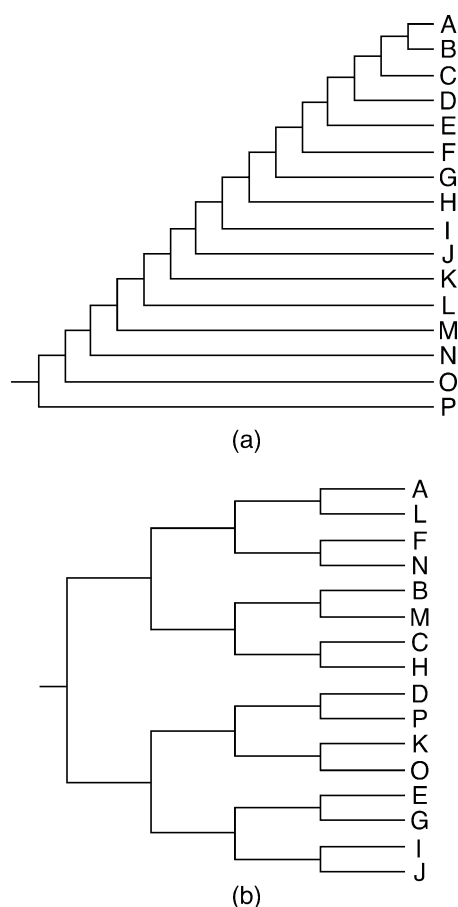

(a)

(b)

FIGURE 1. Two highly incongruent binary input trees of equal size and information content that are maximally unbalanced (a) or maximally balanced (b).
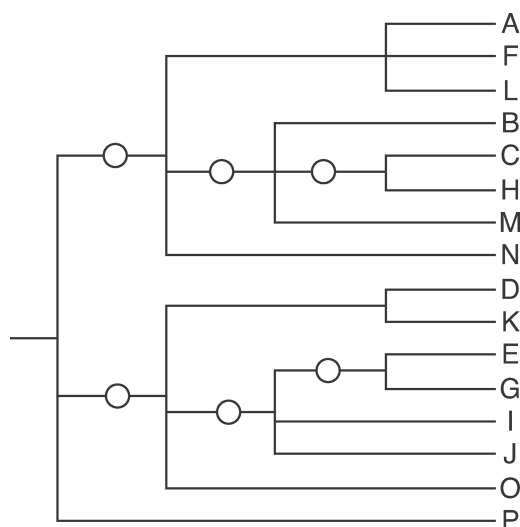
FIGURE 2.   The Adams consensus of the two trees in Figure 1. This tree is also the MinCut and the modified MinCut supertree. Circles indicate components present in the balanced input tree.

trees, there is little basis for resolving conflict and we would not expect any supertree or consensus method to perform well, in the sense of producing an intuitively acceptable and well-resolved synthesis.

If shape does not affect conflict resolution, i.e., if a supertree method is shapeless, then we would not expect supertrees to resolve conflict substantially in favor of the relationships in either the balanced or the unbalanced input tree. Where supertrees are substantially more similar to one of these input trees, we would like to take this as indicative of an ITSE in this case, and that the supertree method is not shapeless in general. However, although shape is the most obvious difference between the input trees, some other feature of the trees may be responsible for the preference, or the preference could be random. The specific effect in either case would still be worrying (because we see no basis for resolving the conflict) but attributing the effect to tree shape would be incorrect. The only feature of these input trees, other than shape, that we can envisage affecting the resolution of conflict in our example is the leaf labeling, which was contrived to produce substantial conflict. Thus, to test both these alternatives, we extended the example to examine many random permutations of the leaf labelings of the input trees for the subset of the supertree methods that appeared not to be shapeless on the basis of the example.

We computed supertrees for the two input trees using the 14 methods described above (Table 1). Most published supertrees have been constructed using standard MRP. This method returns two relatively unbalanced MRP supertrees, the consensus supertree of which is shown in Figure 3a. Each of the standard MRP supertrees is much more similar to the unbalanced ($EA_T = 0.90$, SD = 12–14) than to the balanced ($EA_T = 0.39$, SD = 24) input tree, and this is reflected also in the consensus supertree: of the 13 clades, 7 are also in the unbalanced

tree compared to only 2 that are in the balanced input tree. In this instance, standard MRP produces a result in which conflicts are resolved in favor of relationships in the unbalanced tree, suggesting an ITSE and that the method is not shapeless. To check that the preference was related to tree shape, the analysis was repeated with 500 random permutations of the leaf labeling of the balanced tree. Standard MRP trees are consistently more similar to the unbalanced input tree, indicating a bias with respect to input tree shape in how this method resolves conflict between trees with these topologies (Fig. 4a).

We further assessed the strength of the bias experimentally, using differential weighting of the matrix elements derived from the unbalanced and balanced trees (Fig. 1). Weighting the matrix elements representing the balanced tree 2.7 times as heavily as those of the unbalanced tree produces MRP supertrees that are more similar to the unbalanced ($EA_T = 0.68$) than to the balanced input tree ($EA_T = 0.46$). Weighting by a factor of 2.8 yields MRP supertrees that are more similar to the balanced than to the unbalanced input tree ($EA_T = 0.94$ and 0.38, respectively). With weighting by a factor of 2.75 there are two MRP supertrees, one of which is more similar to the unbalanced ($EA_T = 0.68$ versus 0.38) and one more similar to the balanced input tree ($EA_T = 0.94$ versus 0.48).

Table 2 summarizes results for the various methods. As with the standard approach, irreversible MRP prefers relationships in the unbalanced tree in this case, with all 72 irreversible MRP supertrees more similar to it. The corresponding consensus supertree is fairly well resolved and includes six components that are present in the unbalanced input tree and two from the balanced input tree (Fig. 3b). Random permutation of the leaf labelling confirms that this method is biased with respect to input tree shape (Fig. 4a).

Purvis MRP appears to have a less strong and opposite preference for relationships in the balanced tree. All 80 of the Purvis MRP supertrees are more similar to the balanced input tree. The corresponding consensus supertree (Fig. 3c) is poorly resolved and includes only four components, all of which are present in the balanced tree. Random permutation of the leaf labelling demonstrates that Purvis MRP favors relationships in balanced trees on average, but not exclusively, at least as measured by EA_T (Fig. 4a).

MinFlip also appears to show an ITSE in this case, yielding 31 supertrees that are all much more similar to the unbalanced than to the balanced tree (Table 1). The effect is less apparent in the corresponding consensus supertree (Fig. 3d), which includes only four components, one of which is present in the balanced input tree and two of which are in the unbalanced input tree. Random permutation of the leaf labeling confirms that this method is biased with respect to input tree shape (Fig. 4b).

Both split and quartet fit yield the same three supertrees, the two input trees and one hybrid that is identical to the unbalanced input tree except for the inclusion of one clade (I + J) from the balanced input tree (not shown). The MSS method returns the two input trees only. The consensus supertree is the completely uninformative
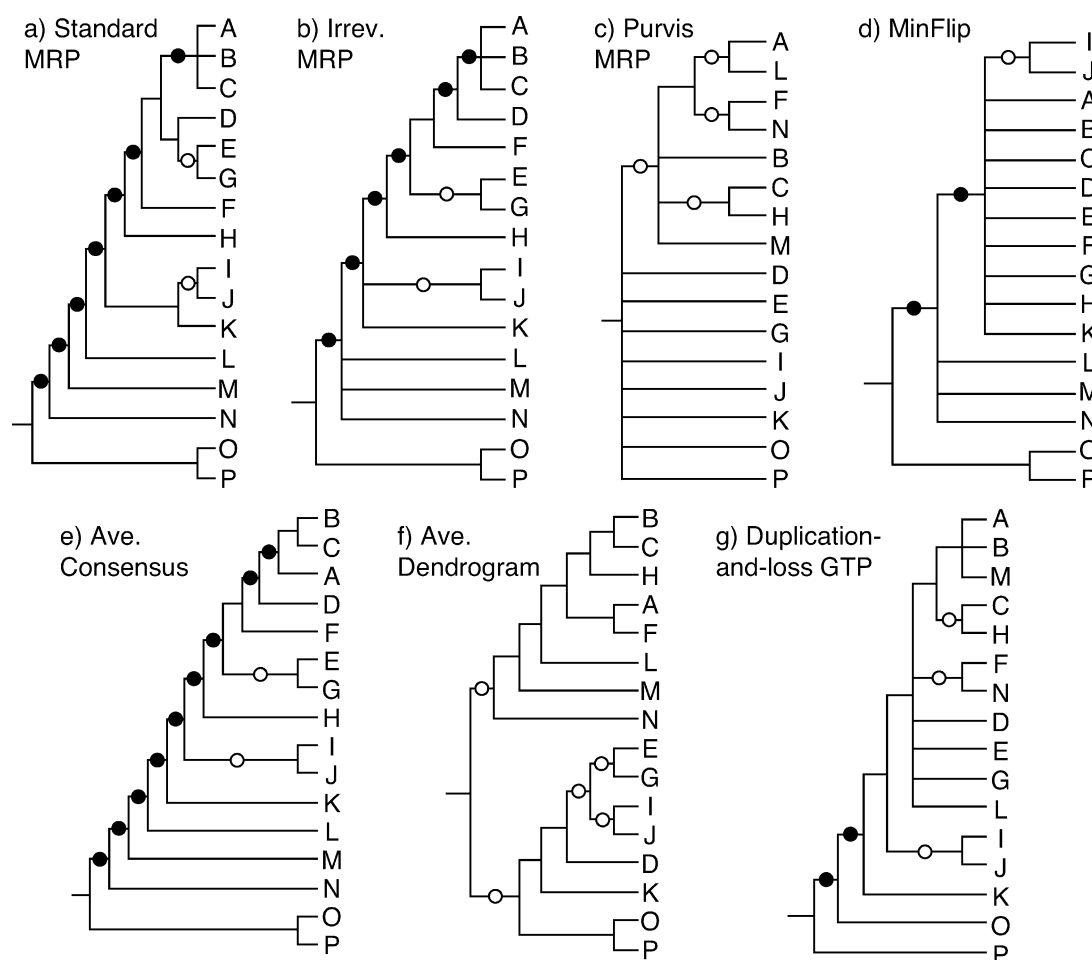
FIGURE 3. Unique optimal supertrees and strict component consensus of optimal supertrees for the two input trees in Figure 1, produced using standard MRP (a), irreversible MRP (b), Purvis MRP (c), MinFlip (d), average consensus (e), average dendrogram (f), and duplication-and-loss GTP (g). The unique optimal supertree found with duplication-only GTP is identical to the unbalanced input tree in Figure 1a. The strict component consensus supertree for the MSS and the component, quartet and triplet fit methods is the unresolved bush. The MinCut supertrees are identical to the Adams consensus of the input trees in Figure 2. White and black circles indicate components present in the balanced and unbalanced input trees respectively.

bush in each case. By returning the input trees, these methods offer no resolution of the extensive conflicts in the input trees and our example reveals nothing of any potential impact of tree shape upon these methods.

Triplet fit yielded 1677 supertrees, the most of any method. Average (dis)similarity scores (Table 2) might suggest a tendency toward favoring relationships in the balanced input tree. Further comparison reveals that

TABLE 2. Numbers (N) of supertrees for the unbalanced (U) and balanced (B) input trees in Figure 1, and measures of distances and similarities between supertrees and input trees, given as ranges with means in parentheses. See text for explanation of other acronyms.

Method	N	EA$_T$ (U)	EA$_T$ (B)	SD (U)	SD (B)
Standard MRP	2	0.90	0.39	12–14 (13)	24
Irreversible MRP	72	0.85–0.94 (0.90)	0.38–0.41 (0.40)	8–16 (12.67)	24
Purvis MRP	80	0.29–0.47 (0.36)	0.53–0.66 (0.56)	28	6–14 (10.10)
MinFlip	31	0.82–0.96 (0.89)	0.36–0.46 (0.40)	4–18 (11.61)	22–26 (25.10)
Split fit/quartet fit	3	0.33–1.00 (0.77)	0.33–1.00 (0.56)	0–28 (10)	0–28 (18)
Triplet fit	1677	0.33–1.00 (0.54)	0.33–1.00 (0.79)	0–28 (24.55)	0–28 (17.35)
Average consensus	1	0.95	0.36	8	26
Average dendrogram	1	0.43	0.89	28	18
MSS	2	0.33–1.00	0.33–1.00	0–28	0–28
MinCut/modified MinCut	1	0.36	0.83	22	12
GTP (DL)	10	0.65–0.68 (0.67)	0.43–0.46 (0.45)	22–24 (22.8)	18–22 (19.8)
GTP (D)	2	0.99–1.00	0.32–0.34	0–2	26–28

large majorities (1327 and 1317) of the triplet fit supertrees are more similar to the balanced tree than they are to the unbalanced tree (measured with EA_T and SD, respectively). However, as with split fit, quartet fit, and MSS, the triplet fit supertrees include both of the input trees and their strict component consensus is the completely unresolved bush.
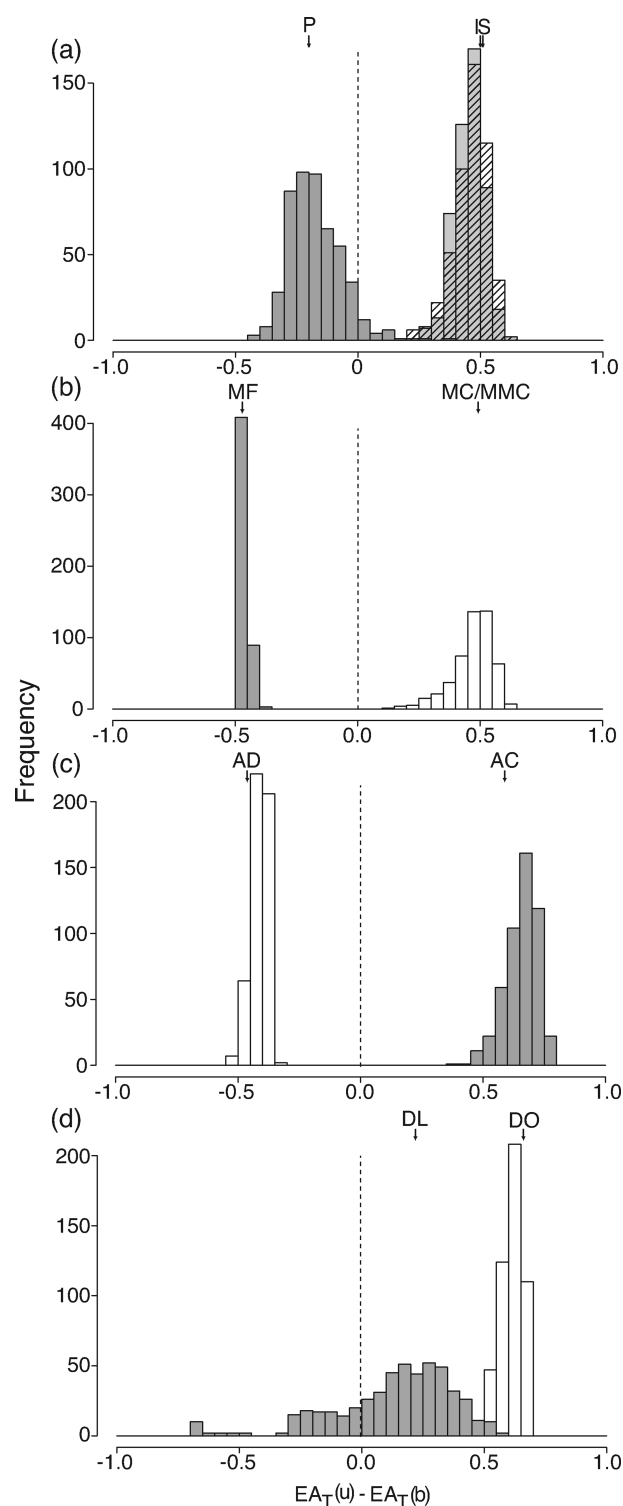
The average consensus method yielded a single tree (Fig. 3e) that is much more similar to the unbalanced ($EA_T = 0.95$) than to the balanced ($EA_T = 0.36$) input tree, and which includes nine components that are present in the unbalanced tree, two that are present in the balanced tree, and two that are not present in either input tree. Conversely, the average dendrogram (Fig. 3f) appears more similar to the balanced ($EA_T = 0.89$) than to the unbalanced ($EA_T = 0.43$) input tree, including five components present in the former and none from the latter. Random permutation of the leaf labeling confirms that these methods are biased with respect to input tree shape in the directions suggested by the example (Fig. 4c).

MinCut supertree methods are Adams consensus-like in that they include all nestings that are present in all the input trees (Semple and Steel, 2000) and both MinCut methods return the Adams consensus in this case (Fig. 2). Interpreted as a cladogram (i.e., as a set of components), this tree is much more similar to the balanced than to the unbalanced tree ($EA_T = 0.83$ versus 0.36 and SD = 12 versus 22, respectively). Random permutation of the leaf labelling confirms that this method is biased with respect to input tree shape (Fig. 4b) when interpreted this way. Interpreted as a set of common nestings (Adams, 1986; Wilkinson, 1994a), the MinCut supertrees are necessarily compatible with both input trees and thus show no ITSE, and we expect this generally.

Duplication-only GTP returned two trees, the unbalanced input tree and the similar hybrid tree found also by the split fit and quartet fit methods, indicating a strong preference for relationships in the unbalanced tree in this case. Random permutation of the leaf labeling confirms that duplication-only GTP is biased with respect to input tree shape as suggested by the example (Fig. 4d). In contrast, duplication-and-loss GTP returned 10 supertrees that were all a little more similar to the unbalanced than to the balanced input tree, as measured by EA_T, but equally similar or more similar to the balanced tree, as measured by SD (Table 2). The consensus supertree (Fig. 3g) includes three components that are present in the balanced input tree, two components that are in the unbalanced input tree, and three that are in neither input tree. This result is inconclusive, and random

FIGURE 4.   Input tree shape biases demonstrated from sets of 500 random permutations of the leaf labeling of the balanced input tree in Figure 1. The mean similarity (EA_T) between the optimal supertrees and the corresponding unbalanced and balanced input trees was calculated ($EA_T(u)$ and $EA_T(b)$, respectively). The figure shows the distribution of the statistic $EA_T(u) - EA_T(b)$ across permutations, so positive and negative values thus indicate greater similarity to the unbalanced and balanced trees, respectively. (a) Distributions for standard (light grey, S), Purvis (dark grey, P), and irreversible (transparent, hatched bars, I) MRP; (b) distributions for MinFlip (light grey, MF) and MinCut and modified MinCut (unshaded, MC/MMC) methods; (c) distributions for average consensus (light grey, AC) and average dendrogram (unshaded, AD) methods; (d) distributions for duplication and loss (light grey, DL) and duplication-only (unshaded, DO) GTP. Arrows indicate the values for the example labeling of Figure 1 for each method. An unbiased method would lead to a distribution centred on 0 for this statistic.

permutation reveals a broad range and left-skewed distribution of similarities to balanced and unbalanced input trees and no clear bias as measured by EA_T, although the mean similarity is significantly non-zero at -0.0089 (Fig. 4d).

MRP has been widely used to combine and synthesize the information in available phylogenies into larger, and therefore potentially more useful, supertrees (e.g., Purvis, 1995b; Liu et al., 2001; Pisani et al., 2002). Our concern is with how reasonably the information in the input trees is combined. There would be little merit in methods that arbitrarily resolve conflict as opposed to resolving it only when warranted by the evidence. Our simple example indicates that standard and irreversible MRP, MinFlip, average consensus, and duplication-only GTP can sometimes favor relationships in more unbalanced trees, and that Purvis MRP, average dendrogram, and MinCut and its modification (if interpreted as producing cladograms) can sometimes favor relationships in more balanced trees. Random permutation confirms that these methods suffer input tree shape biases. Although we do not know the extent to which these biases are important in practice, we consider them arbitrary and worrying.

An argument could be made that some tree shapes and thus some trees are more probable, as is the case if trees are generated under a Markovian model (because there are more ways to grow balanced cladograms), and that relative probability might reasonably be taken into account in the relative weight attached to relationships in input trees. Without wishing to dismiss such notions entirely, we do not expect the seemingly accidental tree shape effect demonstrated for these supertree methods to emulate a well-designed "tree balance weighting scheme" sufficiently well to dispel our concerns over their failure to be shapeless in general.

Causes of Input Tree Shape Effects

Thorley and Wilkinson (2003) suggested that supertree methods could be usefully characterized in terms of their objective functions, the measures of distance, similarity or fit between a supertree and input trees that are minimized or maximized by the methods (Table 1). For example, median (component) consensus trees are those that minimize the sum of the symmetric differences between the consensus and each of the input trees (Barthélemy and McMorris, 1986). By simple extension, the same objective function defines a median supertree method if we allow that, for trees with nonidentical leaf sets, SD is determined by comparing the subtrees induced by leaves in common (i.e., the input tree is compared to a supertree pruned of irrelevant taxa as in the MSS method).

Thorley and Wilkinson (2003) also demonstrated that the objective function of standard MRP is an asymmetric tree-to-tree distance. We argue here that the tree shape effects revealed by our example are related to the use of unusual asymmetric measures of supertree–input tree distances as the bases for the objective functions of those supertree methods that are not shapeless. We must first digress to consider conditions when asymmetric distances have been shown to be useful in order to show that these conditions do not pertain with the current example and that supertree methods that are not shapeless rely upon a distinct class of asymmetric distances.

Asymmetric Tree-to-Tree Distances

Phillips and Warnow (1996) noted that the median consensus is often poorly resolved as a consequence of the symmetry of the tree-to-tree distance measure that is the basis of its objective function. Comparing any input tree and a consensus tree (or pruned supertree), relationships (in this case components) can be classified as those that are in the input tree and not in the consensus (A), those that are in both (B), and those that are in the consensus and not in the input tree (C). $SD = A + C$, and is therefore symmetric, but a candidate consensus is penalized if it includes components that are in less than 50% of the input trees. This is the case even if the minority components are uncontradicted by (e.g., by soft polytomies in) the remaining input trees.

To solve this problem, Phillips and Warnow (1996) proposed the asymmetric median consensus method, which minimizes A rather than $A + C$, and they showed that their new method is more liberal in the sense of always being at least as well resolved as the median consensus. Relationships that are in the consensus but not in a particular input tree do not contribute to the asymmetric difference (A) between the input tree and the consensus.

Note that minimizing A is equivalent to maximizing B, and that B is a symmetric tree-to-tree similarity measure. B is also the objective function maximized across all input trees by split fit supertrees. Thus, extended to the supertree context, the corresponding asymmetric median supertree method is equivalent to split fit. Triplet and quartet fits also maximize B, where this is now understood to be the number of triplets or quartets present in both input and supertree. All three methods, triplet fit, quartet fit, and split fit are examples of methods that employ as the basis of their objective functions a special class of asymmetric tree-to-tree distance (A), the minimization of which is equivalent to maximizing a symmetric tree-to-tree similarity (B).

Phillips and Warnow (1996) clearly show the utility of this special class of asymmetric tree-to-tree distances in defining relatively liberal consensus or supertree methods that use MRs. The advantage is primarily in producing more resolved supertrees, but in our example all trees (input and supertrees) are binary so that there is no such advantage to be had, and these measures are necessarily symmetric in this case. In contrast, supertree methods that are not shapeless appear to have objective functions founded on asymmetric-tree-to-tree distances that are not known to correspond to any symmetric tree-to-tree similarity and which are asymmetric even when all trees are binary.

Asymmetric Distances and ITSEs

We used our example (Fig. 1) to explore MRP objective functions by measuring the parsimony fit of the

MRs of each input tree to the other input tree. Fitting the MR of the unbalanced tree to the balanced tree required 65, 73, and 27 steps using standard, irreversible, and Purvis MRP, respectively. Fitting the MR of the balanced tree to the unbalanced tree required 43, 46, and 39 steps, respectively. Asymmetric tree-to-tree distances are the foundation of the objective functions of Purvis and irreversible, as well as of standard MRP, and these distances are asymmetric even in the special case where all trees are binary. In contrast to triplet, quartet, and split fit methods, we have been unable to identify symmetric tree-to-tree similarity measures that correspond to the asymmetric tree-to-tree distance measures used in the MRP supertree methods. Note also that the differences in fit do not appear trivial, and that the direction of the difference is correlated with the direction of tree shape effect exhibited by these methods. Across the random permutations of leaf labelings, the strength of the ITSEs are also significantly correlated with the difference in the tree-to-tree distances (Fig. 5; N = 500 for all methods. Standard MRP, regression coefficient = 0.0053, $F = 109.7$, $P < 2 \times 10^{-16}$; Purvis MRP, regression coefficient = 0.0144, $F = 85.75$, $P < 2 \times 10^{-16}$; irreversible MRP, regression coefficient = 0.0064, $F = 290.8$, $P < 2 \times 10^{-16}$).

We know of no good reason for supertree or consensus methods to employ an asymmetric distance or fit measure that does not correspond to a symmetric similarity measure. Nor can we think of any reason to employ a distance measure that is asymmetric even in the special case where all trees are binary. Because of the asymmetry, with standard and irreversible MRP the fully balanced input tree is taken as not conflicting as strongly with an unbalanced supertree than vice versa. This is further reflected in the different maximum number of (reversible) parsimony steps of the component MRs of the balanced and unbalanced trees (43 versus 70 steps, respectively), which gives unbalanced trees a potentially bigger vote against candidate supertrees that conflict with them. Thus, random supertrees have on average worse parsimony fits to the MR of the unbalanced input tree than they do to the MR of the balanced input tree (Fig. 6). The opposite holds for maximum steps (48 versus 28, respectively) with the oppositely biased Purvis MRP. We believe that objective functions defined on asymmetric tree-to-tree distances provide a plausible mechanism to explain, at least in part, why some supertree methods are not shapeless.

MinFlip tree-to-tree distances, the number of flips needed to render the component MR of one tree compatible with (not identical to) another, are also asymmetric, as would be predicted from its favoring relationships in the unbalanced tree. The strength of the ITSE is also significantly correlated with the difference in the tree-to-tree MinFlip distances (Fig. 5; regression coefficient = 0.0066, $F = 59.13$, $P = 7.93 \times 10^{-14}$) as would be expected if ITSEs are caused by asymmetric distances. As with maximum number of parsimony steps, the maximum number of flips possible for a component MR is not independent of tree shape and gives unbalanced trees a potentially big-

ger vote against supertrees with which they conflict. Of the methods using component MR, only split fit shows no evidence of tree shape effects with our example, and this is the only one of these methods that has an objective function that can be stated as a symmetric tree-to-tree similarity measure.

The distance between two trees can be calculated as a function of the differences between their corresponding pairwise distance MRs. Where absolute or squared differences (i.e., least-squares fit) are used, the corresponding tree-to-supertree distances are necessarily symmetric. Thus, if tree shape effects are caused by asymmetric tree-to-tree distances, we might expect that distance matrix supertree methods would be free from them. However, of the methods examined here, only the MSS meets this expectation with our example.

The MSS method measures the distance between two trees as the sum of the absolute differences in the corresponding pairwise path-length distance matrices, with all branch lengths (i.e., those in the input tree and the pruned supertree) set at unity (i.e., EBL distances), and the overall distance standardized by dividing the sum by the number of comparisons. The average consensus also uses EBL distances in the representation of input trees and the construction of an average distance matrix, but branch lengths in the supertree are not also set at unity. Rather, they are chosen so as to optimize the least-squares fit of the supertree to the average distance matrix. Our example suggests this might be an important practical difference.

We measured the least-squares fit of the EBL distance matrices for each of the input trees to the other tree. The distance between the unbalanced EBL distance matrix and the balanced tree, when branch lengths of the latter are unconstrained, is 811.299, whereas that between the balanced EBL distance matrix and the unconstrained unbalanced tree is 3022.769. Least-squares fit is a symmetric tree-to-tree distance metric when we compare two trees with specified branch lengths. The example shows that when branch lengths are given for one tree and optimized for the other, the magnitude of the least-squares fit may depend upon which branch lengths are specified and which are optimized. In the special case of average consensus examined here, which relies upon EBL distance representations of input trees, the balanced input tree achieves a better least squares fit to an unbalanced (super)tree than vice versa, and the strength of the ITSE is correlated with the magnitude of the difference in least-squares fits of the input trees to each other (Fig. 5; regression coefficient = 2.75×10^{-4}, $F = 256.3$, $P < 2 \times 10^{-16}$). The distance between the unbalanced ultrametric distance matrix and the balanced tree, when the latter is constrained to be ultrametric but branch lengths are otherwise free to vary, is 6520.15, whereas the corresponding fit between the balanced ultrametric matrix distance matrix and the unbalanced tree is 3022.368. Thus, with the average dendrogram method, both the asymmetry in distance and the tree shape effect are reversed, and again the magnitude of the effect correlates with that of the
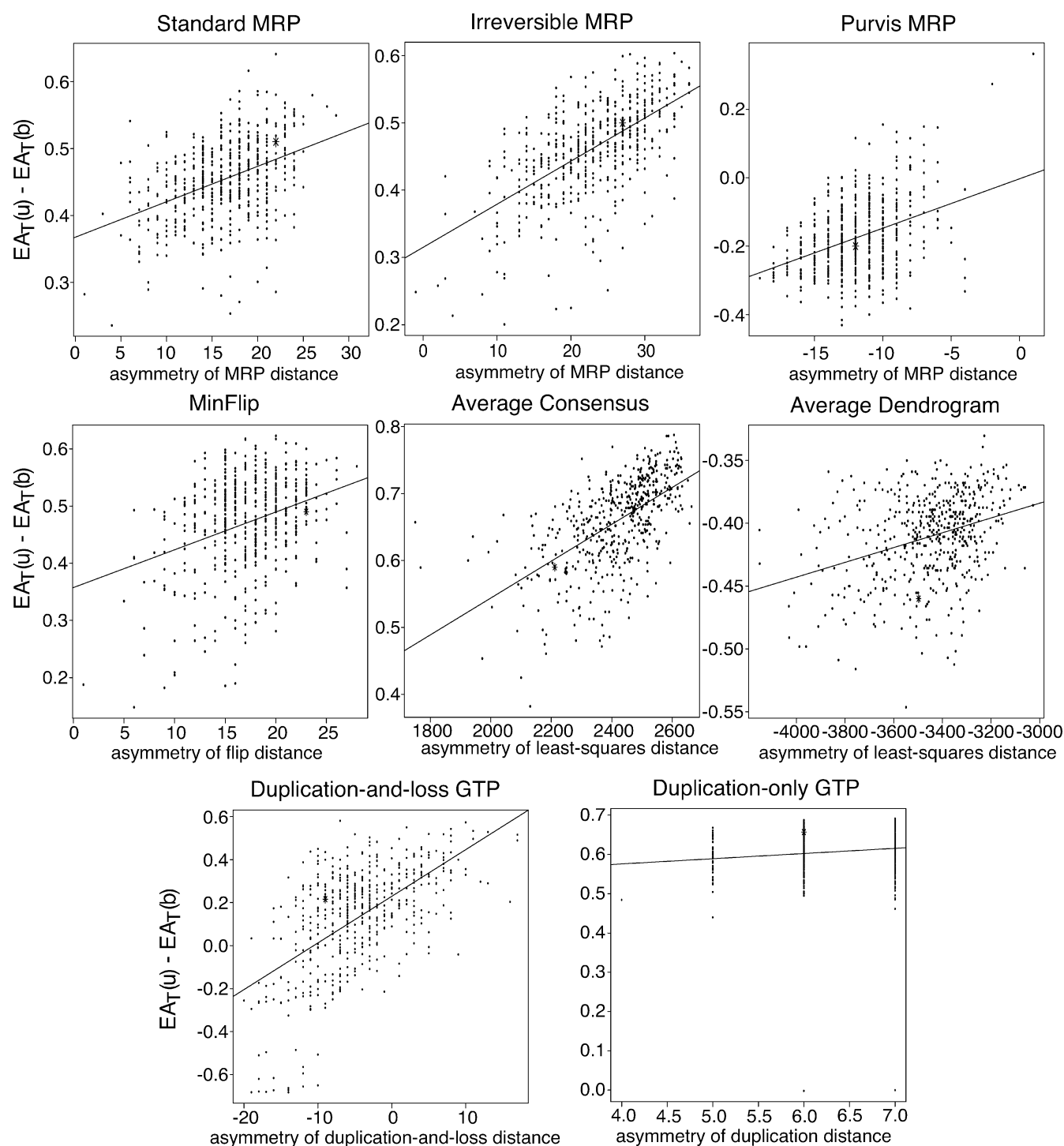
FIGURE 5. Correlation between asymmetry in tree-to-tree distances and input tree shape bias. We determined the cost of fitting the balanced input tree onto the unbalanced input tree and vice versa, with the difference between these two scores being used as a measure of the asymmetry of the particular tree-to-tree distance forming the basis of the methods objective function. This measure was plotted against the difference in similarity (EA_T) between the optimal supertrees and the corresponding unbalanced and balanced input trees ($EA_T(u)$ and $EA_T(b)$, respectively) for each of 500 random permutations of the labels on the balanced input tree topology of Figure 1. Fitted lines are a linear least-squares best-fit for each method, and an asterisk indicates values for the example labelling of Figure 1 in each case. Correlations between distance asymmetry and difference in similarity of the supertrees to the two input trees are highly significant for all methods ($N = 500$, $P < 0.01$).
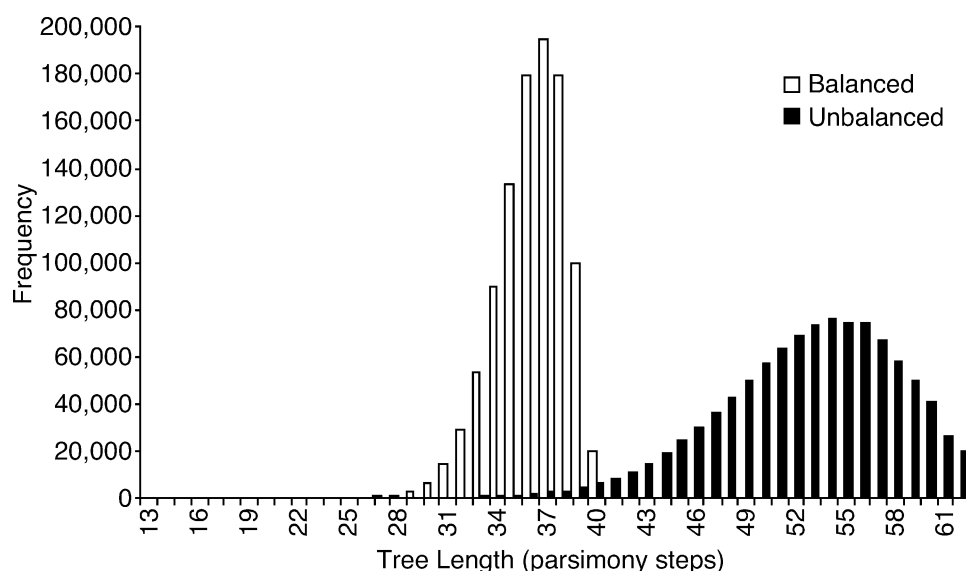
FIGURE 6. Histogram of parsimony lengths of 1,000,000 random trees for the MRs of the balanced (unshaded) and unbalanced (shaded) input trees. Candidate supertrees have on average a better fit to MRs of the balanced input tree.

asymmetry in distances (regression coefficient $= 5.88 \times 10^{-5}$, $F = 62.45$, $P = 1.77 \times 10^{-14}$).

With GTP, tree-to-tree distances are the number of events, duplications, alone, or duplications and losses, needed to reconcile an input tree with the supertree. The example shows that these distances are asymmetric. Using the duplication-only distance, the considerable asymmetry (13 versus 7 events) correlates with a clear bias towards relationships in the unbalanced input tree, with the strength of the ITSE correlated with the asymmetry of the tree-to-tree distances between the input trees (Fig. 5; regression coefficient $= 0.0133$, $F = 8.648$, $P = 0.00343$).

In contrast, with duplications and losses the asymmetry of the distances is reversed in our example. Reconciling the balanced tree with the unbalanced tree requires 71 events (13 duplications and 58 losses), whereas reconciling the unbalanced tree with the balanced tree requires fewer duplications (7) but more events in total (80). The asymmetry is less marked (71 versus 80) and no clear bias is demonstrated by the example or by the broad range of similarities to balanced and unbalanced input trees achieved by optimal supertrees. However, the random permutations reveal a correspondingly broad range of asymmetries in distances (Fig. 5). The duplication-and-loss distance of a balanced input tree to an unbalanced input tree can be smaller, equal to, or greater than the reverse distance, and there is a strong correlation between the direction and magnitude of this asymmetry and the strength and direction of the ITSE's (Fig. 5; regression coefficient $= 0.0217$, $F = 257.7$, $P < 2 \times 10^{-16}$), consistent with the hypothesis that shape effects and biases are caused, at least in part, by asymmetric distances.

MinCut and modified MinCut lack an explicit objective function and do not rely upon a tree-to-tree distance. Thus, we cannot explain the greater similarity of the MinCut supertrees to the balanced tree with our example, which pertains only when they are interpreted as a collection of components rather than as a collection of more ambiguous nestings, in terms of an asymmetric tree-to-tree distance in this case. With polytomies interpreted as nestings, MinCut supertrees show no evidence of an ITSE.

We cannot generalize from the lack of any ITSEs shown by the MSS, component, quartet fit, and triplet fit methods in the special case represented by our simple example. However, if tree shape biases result from objective functions founded on asymmetric tree-to-tree distance measures then we would expect these methods to be generally free of any such bias by virtue of their objective functions being founded on symmetric distances. The validity or otherwise of this conjecture is an important open problem that should be established analytically or by counterexample, or, failing that, should be investigated using simulations.

DISCUSSION

Supertree construction remains a very new field with an increasing number of methods, many of which are poorly known. Most published supertrees have been constructed using standard MRP, seemingly because this method was developed early in the history of supertree construction and is readily implemented. However, there has been relatively little consideration of the fundamental properties of standard MRP and we have little theoretical understanding of to what extent it and other methods are more or less well suited to the task of constructing useful (i.e., accurate) supertrees. The rush to build MRP supertrees in the absence of much understanding of the method demonstrates the strength of the perceived need for supertrees.

Previous workers have suggested that standard MRP may suffer from biases with respect to tree size (e.g., Purvis, 1995a) and tree shape (Wilkinson et al., 2001). Size biases are understood to be more complex than at first thought and to involve the relative sizes of substrees that span or cover conflicting relationships (Bininda-Emonds and Bryant, 1998). Given that tree shape is a function of relative subtree size, tree shape and size effects, as well as the poorly understood positional effects discussed by Wilkinson et al. (2004), may well be interrelated.

Our example, suggests that many supertree methods, including MRP, are not shapeless. They suffer from biases with respect to tree shape that can influence how input tree conflicts are resolved. With the exception of MinCut methods, the biased methods are characterized by objective functions based on asymmetric tree-to-tree distances (ones that are not equivalent to symmetric tree-to-tree similarity measures), with the direction of the asymmetry correlated with the direction of the effect, and the magnitude of the effect correlated with the absolute difference in tree-to-tree distances. If, as we conjecture, this is a causal relation, then supertree methods with objective functions based on symmetric tree-to-tree distances may be free of any input tree shape bias, and this merits further investigation. MinCut methods are biased only if their Adams consensus-like properties are not taken into account in the interpretation of MinCut supertrees.

It might be objected that our example is very unrealistic, and that input tree shape may be unimportant in the supertree context, or when there are more trees, or that it may not be strong enough to have any practical import. Simple thought experiments convince us that ITSEs are not restricted to the consensus setting and we would be concerned for other reasons if they were. Imagine adding some unique leaves to each of the input trees in our first example (so as to convert it into a supertree problem). We think it may be reasonable to consider the relationships of these unique leaves to be irrelevant to the resolution of the conflict between the trees over the relationships of the common leaves, and this expectation is related to independence axioms in bioconsensus (Wilkinson et al., 2004). If the addition of leaves somehow removed any ITSE it would do so only at the expense of violating at least one other "independence axiom."

We also think it unlikely that any shape bias will always be overwhelmed, rather than exacerbated, by larger numbers of input trees, although this remains an open question. We do expect conditions in which there is no good basis for resolving conflicts to be rarer when there is a good sample of input trees. This might ameliorate input tree shape biases, particularly if they are strongest, as well as the most easily revealed and investigated, when, as in our example, there is no basis for resolving conflict. Shape biases might be tolerated if they prove not to be much of a problem, but it seems to us that only unbiased methods will allow this problem to be avoided altogether. Biased methods might also be preferred if they have other important desirable properties that are not shared by unbiased methods. For example, the average consensus makes use of branch length infor-

mation that is treated as irrelevant by other methods but which may be helpful.

It is at present unclear to what extent properties revealed by the rather special case of our example will be, or have been, important in supertree construction, but potential users should be aware that several methods, including the currently most popular, do not satisfy some seemingly reasonable desiderata (see also Wilkinson et al., 2004). Tree shape can be used to investigate macroevolution (e.g., Mooers and Heard, 2002) and possible shape biases of supertree methods might be of particular concern if the shapes considered are of supertrees produced by biased methods. We hope that our findings encourage consideration of ITSEs and the potential role of asymmetric tree-to-tree distances in the future development or testing of supertree methods, and that they highlight the need for further study of the properties and comparative performance of supertree methods.

Given our current limited understanding of supertree methods, and hence our limited scope for justified choice among the alternative methods, we warn against an over-reliance upon any single arbitrarily chosen method. Uncritical use of a single method should be discouraged. We see no good reason for standard MRP to be the method of choice in supertree construction, particularly given its apparent bias toward relationships in unbalanced input trees, its potential to yield unsupported groups (Pisani and Wilkinson, 2002), and because recent simulation studies show that MinFlip performs at least as well as MRP (Eulenstein et al., 2004). Use of multiple methods allows practical assessment of the extent to which inferences depend upon method and of the impact of possible biases and should be encouraged. There is also a pressing need for validation methods for supertrees like those used to assess the reliability of phylogenetic trees (see Lapointe and Cucumel, 2003). There is still a long way to go before supertree methodology comes of age.

REFERENCES

Aho, A. V., Y. Sagiv, 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.

Baum, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41:3–10.

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

Bininda-Emonds, O. R. P., and H. N. Bryant. 1998. Properties of matrix representation with parsimony analyses. Syst. Biol. 47:497–508.

Bininda-Emonds, O. R. P., J. L. Gittleman, and M. A. Steel. 2002. The (super) tree of life: Procedures, problems and prospects. Ann. Rev. Ecol. Syst. 33:265–289.

Bininda-Emonds, O. R. P., and M. J. Sanderson. 2001. An assessment of the accuracy of MRP supertree construction. Syst. Biol. 50:565–579.

Bryant, D., and M. A. Steel. 1995. Extension operation on sets of leaf-labelled trees. Adv. Appl. Math. 16:425–453.

Chen, D., L. Diao, O. Eulenstein, D. Fenández-Baca, and M. J. Sanderson. 2003. Flipping: A supertree construction method. Pages 135–160 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.

Constantinescu, 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.

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 only exist 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 through supertree analyses. Bioinformatics 21:390–392.

Estabrook, G.F., F. R. McMorris, and C. A. Meacham. 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. Syst. Zool. 34:193–200.

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

Farris, J. S. 1973. On comparing the shapes of taxonomic trees. Syst. Zool. 22:50–54.

Felsenstein, J. 1993. PHYLIP: Phylogenetic inference package, version 3.5c. University of Washington, Seattle.

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

Gordon, A. D. 1986. Consensus supertrees: The synthesis of rooted trees containing overlapping sets of labeled leaves. J. Classif. 3:31–39.

Lanyon, S. M. 1993. Phylogenetic frameworks: Towards a firmer foundation for the comparative approach. Biol. J. Linn. Soc. 49:45–61.

Lapointe, F.-J., and G. Cucumel. 1997. The average consensus procedure: combination of weighted trees containing identical or overlapping sets of taxa. Syst. Biol. 46:306–312.

Lapointe, F.-J., and G. Cucumel. 2003. How good can a consensus get? Assessing the reliability of consensus trees in phylogenetic studies. Pages 205–219 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.

Lapointe, F.-J., and P. Legendre. 1995. Comparison tests for dendrograms: A comparative evaluation. J. Classif. 12:265–282.

Lapointe, F.-J., and C. Levasseur. 2004. Everything you always wanted to know about the average consensus, and more. Pages 87–105 in Phylogenetic supertrees: Combining information to reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Liu, F.-G. R., M. M. Miyamoto, N. P. Freire, P. Q. Ong, M. R. Tennant, T. S. Young, and K. F. Gugel. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. Science 291:1786–1789.

Mooers, A. O., and S. B. Heard. 2002. Using tree shape. Syst. Biol. 51:833–834.

Nelson, G., and P. Y. Ladiges. 1994. Three-item consensus: Empirical test of fractional weighting. Pages 193–209 in Models in phylogeny reconstruction (R. W. Scotland, D. J. Siebert, and D. M. Williams, eds.). The Systematics Association Special Volume No. 52. Oxford University Press, New York.

Page, R. D. M. 1998. GeneTree: Comparing gene and species phylogenies using reconciled trees. Bioinformatics 14:819–820.

Page, R. D. M. 2002. Modified mincut supertrees. Lecture Notes in Computer Science 2452:537–551.

Page, R. D. M., and M. A. Charleston. 1997. From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem. Mol. Phy. Evol. 7:231–240.

Phillips, C. A., and T. J. Warnow. 1996. The asymmetric median tree—A new model for building consensus trees. Discrete Appl. Math. 71:311–335.

Pisani, D. 2002. Comparing and combining data and trees in phylogenetic analysis. Ph.D. Thesis, Department of Earth Sciences, University of Bristol, UK.

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

Pisani, D., A. M. Yates, M. C. Langer, and M. J. Benton. 2002. A genus-level supertree of the Dinosauria. Proc. R. Soc. Lond. B 269:915–921.

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 representation of trees. Mol. Phylogenet. Evol. 1:53–58.

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

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

Ronquist, F. 1996. Matrix representation of trees, redundancy, and weighting. Syst. Biol. 45:247–253.

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. 2000. A supertree method for rooted trees. Discrete Appl. Math. 105:147–158.

Slowinksi, J., and R. D. M. Page. 1999. How should species trees be inferred from molecular sequence data? Syst. Biol. 48:814–825.

Steel, M. A. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. J. Classif. 9:91–116.

Steel, M. A., A. W. M. Dress, and S. Böker. 2000. Simple but fundamental limitations on supertree and consensus tree methods. Syst. Biol. 49:363–368.

Swofford, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts.

Thorley, J. L., and R. D. M. Page. 2000. RadCon: Phylogenetic tree comparison and consensus. Bioinformatics 16:486–487.

Thorley, J. L., and M. Wilkinson. 2000. The RadCon manual 1.1.2. Bristol University, UK. http://darwin.zoology.gla.ac.uk/~thorley/manual/manual.htm.

Thorley, J. L., and M. Wilkinson. 2003. A view of supertree methods. Pages 185–193 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.

Thorley, J. L., M. Wilkinson, and M. A. Charleston. 1998. The information content of consensus trees. Pages 91–98 in Advances in data science and classification (A. Rizzi, M. Vichi, and H.-H. Bock, eds.). Springer, Berlin.

Wilkinson, M. 1994a. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. Syst. Biol. 43:343–368.

Wilkinson, M. 1994b. Three-taxon statements: When is a parsimony analysis also a clique analysis? Cladistics 10:221–223.

Wilkinson, M., J. L. Thorley, D. T. J. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree for the Platyhelminthes? Pages 292–301 in interrelationships of the Platyhelminthes (D. T. J. Littlewood and R. A. Bray, eds.). Chapman-Hall, London.

Wilkinson, M., J. L. Thorley, D. Pisani, F.-J. Lapointe, and J. O. McInerney. 2004. Some desiderata for liberal supertrees. Pages 227–246 in Phylogenetic supertrees: Combining information to reveal the Tree of Life. (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.