Minority rule supertrees? MRP, Compatibility, and Minimum Flip may display the *least* frequent groups

Pablo A. Goloboff*

Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto Miguel Lillo, Miguel Lillo 205, 4000 SM. de Tucumán, Argentina

Abstract

New examples are presented, showing that supertree methods such as matrix representation with parsimony, minimum flip trees, and compatibility analysis of the matrix representing the input trees, produce supertrees that cannot be interpreted as displaying the groups present in the majority of the input trees. These methods may produce a supertree displaying some groups present in the minority of the trees, and contradicted by the majority. Of the three methods, compatibility analysis is the least used, but it seems to be the one that differs the least from majority rule consensus. The three methods are similar in that they choose the supertree(s) that best fit the set of input trees (quantified as some measure of the fit to the matrix representation of the input trees); in the case of complete trees, it is argued that, for a supertree method to be equivalent to majority rule or frequency difference consensus, two necessary (but not sufficient) conditions must be met. First, the measure of fit between a supertree and an input tree must be symmetrical. Second, the fit for a character representing a group must be measured as absolute: either it fits or it does not fit. In the restricted case of complete and equally resolved input trees, compatibility analysis (unlike MRP and minimum flipping) fulfils these two conditions: it is symmetrical (i.e., as long as the trees have the same taxon sets and are equally resolved, the number of characters in the matrix representation of tree A that require homoplasy in tree B is always the same as the number of characters in the matrix representation of tree B that require homoplasy in tree A) and it measures fit as all-or-none. In the case of just two complete and equally resolved input trees, the two conditions (symmetry and absolute fit) are necessary and sufficient, which explains why the compatibility analysis of such trees behaves as majority consensus. With more than two such trees, these conditions are still necessary but no longer sufficient for the equivalence; in such cases, the compatibility supertree may differ significantly from the majority rule consensus, even when these conditions apply (as shown by example). MRP and minimum flipping are asymmetric and measure various degrees of fit for each character, which explains why they often behave very differently from majority rule procedures, and why they are very likely to have groups contradicted by each of the input trees, or groups supported by a minority of the input trees.

© The Willi Hennig Society 2005.

The idea of combining the results of different phylogenetic analyses by producing a supertree has been proposed by several authors, most recently and notably by Bininda-Emonds (2004a,b; Bininda-Emonds et al., 2002). The most widely used method is MRP ("matrix representation with parsimony"; Baum, 1992; Ragan, 1992; Baum and Ragan, 1993); one of the reasons for using MRP is the belief that, when the input trees display conflicting groups, the

supertree will (often) display the most frequent groups:

> "Many methods have links to conventional consensus techniques. For example, strict and semistrict supertrees are the analogs of their consensus namesakes, MinCut supertree resembles Adams consensus, and MRP performs similarly to fully resolved majority rule consensus" (Bininda-Emonds, 2004b, p. 2)

By "fully resolved" Bininda-Emonds (2004b) means (pers. comm.) a majority rule consensus tree to which compatible groups with a frequency below 50% have been added. Both PAUP* and Phylip calculate these trees; when two contradictory groups have equal

*Corresponding author.
E-mail address: pablogolo@csnat.unt.edu.ar

frequency, one of them is chosen arbitrarily for addition to the tree. Such a ''fully resolved'' consensus is arbitrary; when there are conflicting groups of the same frequency, the result depends on the ordering of the input trees (see Felsenstein, 1993; Farris et al., 1996). The resolution of the majority rule consensus can be improved, non-arbitrarily, using Goloboff et al.'s (2003) frequency difference consensus (or GC measure[1]). Eulenstein et al. (2004) have also compared supertree methods to normal consensus:

> ''MRP, MRF [minimum flipping], and MC [Minimum Cut; Semple and Steel, 2000] supertrees preserve certain consensus properties of their input trees when the input trees include the same taxa. MRF, MRP, and MC supertrees contain the strict and semistrict consensus clusters… This is a desirable property in the sense that consensus properties represent a limiting case for supertree methods when the taxon sets are the same.'' (Eulenstein et al., 2004)

Eulenstein et al. are correct in that the strict and semi-strict consensus clusters will necessarily be present in the MRP, MRF or MC trees, but they neglect to mention that many other clusters, besides the strict and semi-strict ones, may be present in the results for any of those three methods.[2] The additional groups are to be expected, following Bininda-Emonds' (2004b) claim that MRP supertrees are similar to (''fully resolved'') majority rule consensus. However, the analogy proposed by Bininda-Emonds (2004b) is rather surprising, since MRP has long been known (Bininda-Emonds and Bryant, 1998; Bininda-Emonds et al., 1999; Goloboff and Pol, 2002) to produce groups that are contradicted by each of the input trees, which is an impossibility in majority rule trees.

In recent contributions, Wilkinson et al. (2004, in press) have discussed the bias that asymmetric measures of fit introduce in supertree analysis. As they note, both MRP and the recently proposed minimum flipping use asymmetric measures. The present note provides additional examples and discussion of the possible consequences of using asymmetric measures of fit. The examples show that, far from being similar to majority rule consensus trees, the supertrees produced by MRP or minimum flipping may easily display the groups present in a minority of the trees; although less frequently (i.e., under more stringent conditions) the same is true of matrix representation with compatibility.

---

[1]Goloboff et al. (2003) used this measure as a better way to summarize resampling. It is based on calculating the frequency for each group present in the input trees. The frequency difference consensus includes all groups that are more frequent than the most frequent contradictory group. This method is included in the program TNT.

[2]Heuristics aside, the method of Goloboff and Pol (2002), not cited by Eulenstein et al. (2004) will also display all semi-strict consensus clusters, and only those clusters.

As in previous papers (e.g., Bininda-Emonds, 2003b; Eulenstein et al., 2004), I use trees with the same sets of taxa to examine the behaviour of supertree methods. The findings presented here, in showing properties that are necessary (although not sufficient) for a supertree method to behave as a majority rule method, may eventually help define better supertree methods.

All the examples in this paper were analyzed with TNT (Goloboff et al., 2003). TNT has routines to generate matrices representing the trees in memory; these matrices can be subsequently analyzed, using parsimony, cliques, or minimum-flipping. The cliques were implemented by running analyses under implied weights (Goloboff, 1993a), with the weighting function defined to take weight 1 for no homoplasy, and weight 0 for any number of extra steps (TNT allows defining weighting functions of any shape). The minimum flip trees (Chen et al., 2003) were found by heuristic searches, using TBR. These searches were implemented using a macro routine in TNT that calculates the flip distance to a tree (*flipd*), and using the *tbrit* command to generate the TBR rearrangements. The individual examples of minimum flipping were corroborated with Eulenstein et al.'s program, Rainbow (available at http://genome.cs.iastate.edu/supertree/index.html); the results for flipping on random trees were only checked with TNT, which could be done easily with a proper script. In the simulations, the searches for minimum flip trees used Wagner MRP trees as a starting point for TBR branch swapping (with three random addition sequences followed by TBR, saving up to 10 trees per replication); the parsimony or clique searches used 100 random addition sequence wagner trees, each followed by TBR saving up to five trees, and subsequent branch-swapping from the resulting trees saving unlimited numbers of trees.

MRP

Goloboff and Pol (2002) presented a supertree method that was equivalent to the combinable components (or semi-strict) consensus. They also suggested that a conceptual equivalent of the majority rule consensus, even if desirable, might not be definable in the case of supertrees. They showed, with an example (their Fig. 10, reproduced here as Fig. 1), that the trees which support a given group without conflict (when combined with some trees), might be the same trees which contradict the group without conflict (when combined with other trees). In some cases therefore, an individual tree cannot be said to either support or contradict a group, unless all the taxa included and excluded from the group are present in the tree.

Bininda-Emonds (2004b) provided no discussion or refutation of the argument of Goloboff and Pol (2002).
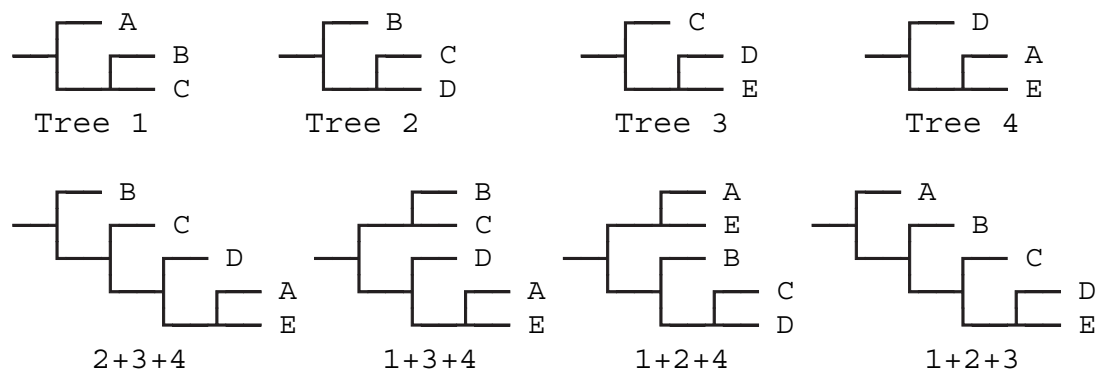
Fig. 1. A case where the number of trees supporting a given group (or triplet resolution) cannot be counted. The group ADE is implied by tree combinations $2+3+4$ and $1+3+4$, but tree combinations $1+2+4$ and $1+2+3$ contradict ADE. The same applies for the resolution of the ACD triplet; tree combinations $2+3+4$ and $1+3+4$ imply C(AD), but tree combinations $1+2+4$ and $1+2+3$ imply A(CD). This shows that a given input tree can support a group (or triplet resolution) when combined with some trees, but contradict it when combined with others.

In his 2003 paper (Bininda-Emonds, 2003a) he misrepresented their position:

> "it seems unreasonable, in practice, to expect a clade in a supertree, especially the larger clades, to not be contradicted by any source tree (as advocated by Goloboff and Pol, 2002)." (Bininda-Emonds, 2003a, p. 846)

Goloboff and Pol (2002) never maintained such a thing; rather, they proposed the only method, to date, that could reliably identify whether a given group is or is not supported or contradicted by any (combination of) input tree(s). These uncontradicted groups provide, no doubt, only a baseline for subsequent resolution, but they first need to be identified; no other method besides Goloboff and Pol's can do that. Goloboff and Pol (2002, p. 523) considered very reasonable "the idea that groups recovered more frequently in independent analyses are more likely to be real phenomena", but pointed out that even so "designing a method that displays groups that are recovered more often in phylogenetic studies when the taxon subsets for each tree are very different may not be possible." Goloboff and Pol (2002) suggested – without formalization – a way in which frequency of recovery of groups could perhaps be taken into account, and noted that (strictly speaking) this was not equivalent to a majority rule consensus tree. In fact, Goloboff and Pol's claim (that majority rule supertree methods may not be possible) concerned clusters, but the claim is equally true as far as the resolution of conflicting triplets (i.e., three-taxon statements) goes. Figure 1, from Goloboff and Pol (2002), provides an example. Consider the relationships between A, C and D in the four input trees. The statement (C(AD) is entailed by the combination of trees $2+3+4$ and $1+3+4$. None of the trees therefore seems to contradict (C(AD)). However, different combinations ($1+2+4$ and $1+2+3$) entail the alternative resolution (A(CD)). Each of the trees can entail one resolution of the ACD triplet when combined

with some trees, but a different resolution when combined with others. In such cases, going back to the original trees and checking fit for triplets is, simply, devoid of meaning.

Contrary to Bininda-Emonds' (2004b) claims, MRP comes so far from producing majority rule supertrees that the results of MRP can be very different from a majority rule consensus, even in the case of trees with identical taxon sets. Consider the set of 10 trees, shown in Fig. 2. Tree A appears in nine copies, while tree B appears only once. Trees A and B differ only in the placement of four taxa (R, T, V, and X). The frequency of each of the groups in tree A is 90%. MRP analysis, however, produces six of the groups in the single tree (marked), even when those groups are directly contradicted by nine out of the 10 input trees. The reason for this is that the many groups from tree B in which R, T, V and X are included outnumber the few groups (even if multiplied by nine copies) from tree A to which R, T, V and X do not belong in the supertree. However, membership to those groups in the MRP tree is simply a parallelism, and results from R, T, V, and X belonging to larger groups in tree B; for example, R, T, V, and X are part (in tree B) of the group KLRTVX, but the existence of group KLRTVX is not in itself reason to believe in the existence of a group RTVX; as discussed by Goloboff and Pol (2002) belonging to a larger group is not in itself evidence of monophyly.

The idea that MRP and majority rule consensus perform similarly is therefore grossly mistaken; the results can be radically different. If there is any logical justification for the MRP results in the present case, it has not yet been provided by any defendant of MRP. Perhaps the results could be justified considering that, if each and every one of the groups in topologies A and B had been found to be supported by the same number of entirely homoplasy-free characters in the individual data
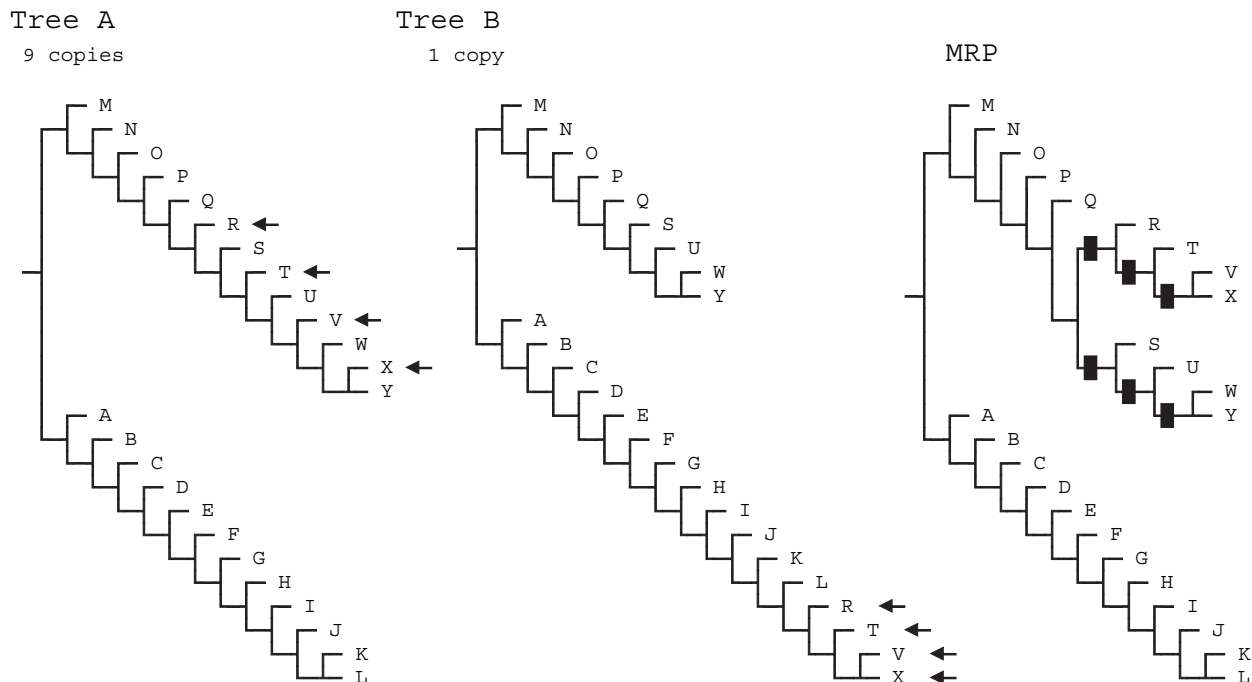
Fig. 2. A case where MRP displays groups (marked) supported by a single input tree and contradicted by nine. Trees A and B differ only in the placement of four taxa, marked with arrows.

sets, then combining the original matrices would produce the same result as the MRP supertree.[3] There are two problems with this defence. First, people who use supertrees use them to combine trees resulting from real matrices (where characters do have homoplasy), or even trees that result from no matrix at all. Second, this uses as reference for the "correct" results the results of a combined analysis ("supermatrix"); if what justifies the result is a supermatrix, then there is no reason for the supertree approach. A convincing justification of the results of Fig. 2 therefore must use some other arguments. None has been offered so far.

Another possible defence of MRP is that the situation shown in Fig. 2 is unlikely to arise in real data and larger trees. Given that for large input trees, it is virtually impossible to check (visually) whether minority or spurious groups are present in the MRP, the claim cannot be tested. When the input trees present significant differences, the situation is far from exceptional. Consider the results of 100 triplets of trees of the same 25 taxa, with one of the trees appearing twice, and the other tree having a random topology. The average number of groups in the MRP that did not appear in any of the input trees was 4.40 (94% of the cases had at least one such group). The average number of groups

appearing in the least frequent tree and the MRP tree, but not in the duplicate input tree, was 0.76 (51% of the cases had at least one such group). The situation is, clearly, not very unlikely. Bininda-Emonds (2003a) has suggested that the spurious groups (which he calls "novel"), shown to occur only for pairs of trees, are less likely to occur in the more realistic case where there are numerous input topologies. As shown below, only having more than two input trees (complete, binary) makes it possible for compatibility supertrees to display groups in the minority of the trees; likewise, having more than two input trees or topologies will no doubt make MRP more prone to displaying groups in the minority of the trees, not less.

Compatibility and symmetry

Goloboff and Pol (2002, p. 522) had suggested that analysing the matrix that represents the input trees using cliques would often produce results that take into account how frequently a group is supported better than MRP. This is so because, under the compatibility criterion, reversals or parallelisms (that is, absence from some groups, or membership to larger groups, in the input trees) do not provide support for groups. Thus, spurious groups (i.e., groups contradicted by each and every one of the input trees, common under MRP) are not possible under compatibility analysis. Ross and

---

[3]An anonymous reviewer pointed out that a ms. in press by Bryant makes this same point; I have not yet seen the paper published.

Rodrigo (2004) examined the performance of clique analysis for simulated input trees; they baptized the method as MRC. MRC often behaves much better than MRP. For the case shown in Fig. 2, it is enough to have two copies of tree A for MRC to produce exactly the same results as the majority rule consensus. The generation of triplets (one tree duplicated) or doublets of random trees and subsequent analysis with MRC produced results that were always identical with majority rule or frequency difference consensus.[4]

Consideration of the properties of compatibility makes it clear why those results are obtained. First, when the input trees are binary and complete, compatibility is (as noted by Wilkinson et al., in press) a symmetric measure. For any two (complete, binary) trees A and B, the number of characters in the matrix representation of tree A that require homoplasy in tree B is always the same as the number of characters in the matrix representation of tree B that require homoplasy in tree A. Second, cliques measure fit with an all-or-none criterion; for binary trees, the cliques are equivalent to simply counting the number of shared groups. Therefore, given two trees with some conflicting group(s), whenever some changes in one tree improve the fit in some characters, by necessity the fit of the contradicting characters will be worsened by an equal amount; the two original trees will therefore be among the set of optimal trees (possibly together with other trees). The MRC results cannot support any group that is not in both trees (when there are only two, binary trees). Since the measure is symmetric, when one of the trees appears in more copies than the other, the results will also by necessity be identical to the more frequent topology.

In the case of parsimony, such a symmetry does not exist, and fit is measured on a step-by-step basis. The lack of symmetry implies that, given two types of topologies in conflict, fitting tree A onto the matrix for tree B may produce a better fit than fitting tree B onto the matrix for tree A; if that is the case, MRP may prefer the groups of tree A, even if topology A is *less* frequent than topology B. This is precisely what the examples in the preceding section show. On the other hand, the fact that parsimony measures fit on a step-by-step basis makes it possible that (when two single trees are in conflict) changing a tree saves in some characters more steps than are added to other characters, and thus a tree that is better than both the original trees may exist. This is what produces the spurious or "novel" groups, well recognized in the MRP literature. The problems with MRP could have been predicted by simply considering that parsimony is not a symmetric measure between two trees, and that parsimony measures fit by degrees, unlike

compatibility. When there are more than two trees, there is even more room for the supertree to be different from the majority rule consensus.

The discussion above shows that MRC is really an equivalent of the frequency difference consensus when there are only two trees. However, the equivalence between MRC and the frequency difference consensus is not general. Compatibility is symmetric only when the input trees have the same taxon sets and are completely resolved. When that is not the case, MRC may be asymmetric and easily produce groups that are not found in the majority of the trees, even for only two types of topology. Goloboff and Pol (2002) provided a case (their Fig. 2) with different numbers of taxa. Figure 3 provides a case with the same numbers of taxa, where the MRC supertree displays several groups contradicted by six trees and displayed by four. Other examples show that, for more than two trees, the conditions of symmetry and absolute fit are not sufficient for a method to be equivalent to frequency difference consensus. Consider the three (complete, binary) trees of Fig. 4; note taxon X with two very different positions in the tree. The group J + K appears in only one tree, and is contradicted by two groups which are equally frequent (K + X and K + L). The frequency difference for J + K is then zero, and the group is not present in the frequency difference consensus (see Fig. 4). However, there are only two compatibility trees; the number of groups shared with each of the input trees (which increases as the number of characters with homoplasy decreases, as the trees are complete and binary) is shown below the two compatibility trees. The group J + K occurs in the compatibility trees because of interactions between the three types of topology. The group K + X, which occurs in one tree, is contradicted by the group B–L (which occurs in two trees); the group K + L, which occurs in one tree, is contradicted by the group I + L (which occurs in two trees). Then, the compatibility tree displays the groups B–L and I + L, instead of the groups K + X or K + L. The group J + K is then free to occur in the compatibility trees. Perhaps the presence of J + K in the results could be justified as the best summary of the three input trees: the group J + K is contradicted as frequently as it is supported, but the groups which contradict it are contradicted more often than they are supported. Concluding J + K may then be reasonable. Whether such a result is desirable or not, is a question that should be further scrutinized.[5] What is clear, however, is that

---

[4] The two methods are equivalent when there are only two types of topologies.

[5] The resolution produced by compatibility beyond a simple frequency difference consensus may be well justified in the case of binary trees, but using compatibility (instead of frequency differences) to summarize the results of resampling is a bad idea: resampling normally produces incompletely resolved trees, and thus compatibility becomes asymmetric.
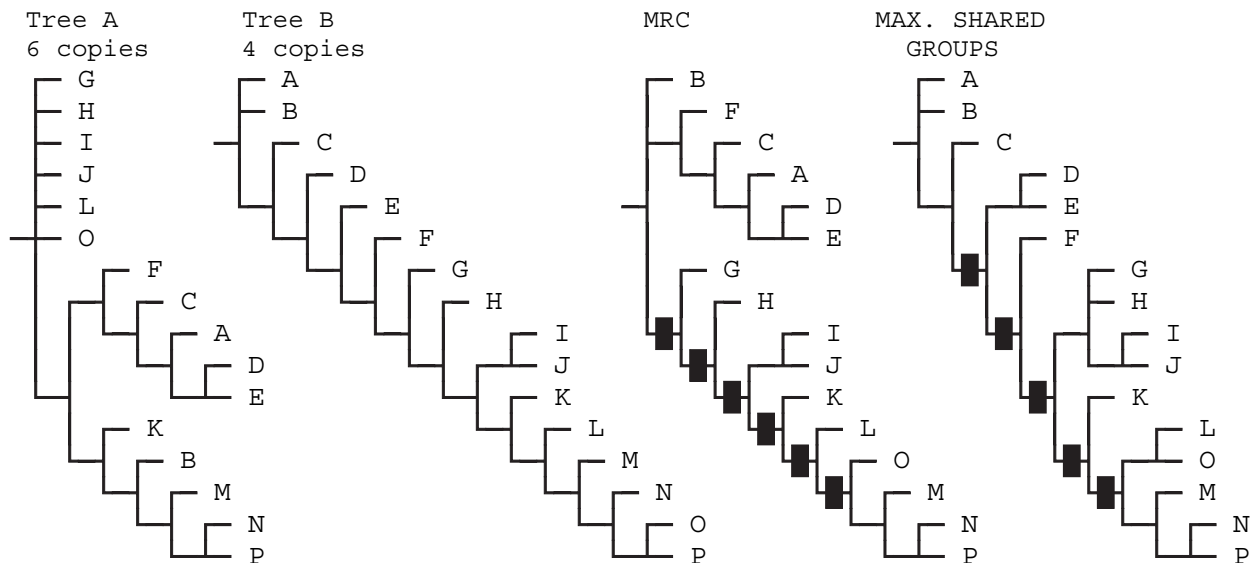
Fig. 3. A case where MRC (matrix representation with compatibility) displays groups (marked) supported by four trees and contradicted by six. Choosing the supertree that has the maximum number of nodes compatible with the input trees produces a result different from MRC, also displaying some groups contradicted by the majority of trees.

MRC produces results which are not equivalent to frequency difference consensus—they will display each and every one of the groups in the frequency difference consensus, but may display some additional groups. This situation (a group contradicted by a second group with equal frequency, but with the second group contradicted by a third, even more frequent one) can happen only when there are more than two distinct trees. Note that when there are more than three trees, this problem may become worse and produce supertrees that contradict the majority rule consensus. An example is shown in Fig. 5, where a group present in 6 out of 11 input trees (54%) is contradicted by the supertree. The example makes it clear that (when there are more than three trees) symmetry and absolute fit are not sufficient conditions for a supertree method to recover the groups in the majority of the input trees.

As noted above, when the input trees are binary and complete, choosing supertrees based on compatibility is equivalent to choosing them based on the number of groups shared. The number of (compatible) groups shared by the (candidate) supertree and each of the input trees could perhaps be used instead of compatibility, when the input trees have different taxon sets and different degrees of resolution. Candidate supertrees could be examined heuristically (say, with branch-swapping, as in Eulenstein et al., 2004), calculating the number of groups for each of a series of candidate supertrees examined during TBR. In a test implementation of this method, I simply counted the number of groups in a semi-strict consensus of the candidate

supertree and each of the input trees (pruning the supertree to have the same taxon set as the respective input tree); the tree(s) which have the greatest sum of number of groups are chosen as supertrees. Such a method, for the case of Fig. 2 in Goloboff and Pol (2002), produces a tree identical to the MRC tree (i.e., as the MRP tree, shown in Goloboff and Pol, 2002, Fig. 2, but with *Atrax* as part of a basal trichotomy. This has two groups (*Hexathele + Scotinoecus*, and *Ischnothele + Hexathele + Scotinoecus*) which are contradicted by one tree and supported by another. The equivalence of this method with MRC is not general, however. For example, for the case shown in Fig. 3, choosing the supertree(s) with the maximum number of compatible groups with each of the input trees, also displays some groups present in the minority of the trees (see Fig. 3), but these are not the same ones found by MRC. In some cases, however, this measure may perform better than MRC, because it is not affected by asymmetries. Having a single copy of each of those trees produces better results than compatibility, at least for the example shown; a tree displaying only the (uncontradicted) group IJ is found, as in a frequency difference consensus. The test implementation of this method used here is very slow, as it simply combines, using the scripting language of TNT, a series of calculations of semi-strict consenses (for the candidate supertree, and each of the input trees) subsequently counting the number of nodes; better implementations are possible, of course, and they would be required for proper testing of this method, which seems promising.

INPUT TREES

FREQUENCY DIFFERENCES

Tree 1

Tree 2

Tree 3

COMPATIBILITY TREES

CONSENSUS OF MRC TREES

Tree 1

Tree 2

Groups shared:
tree 1: 9
2: 3
3: 8
_____
Total: 20
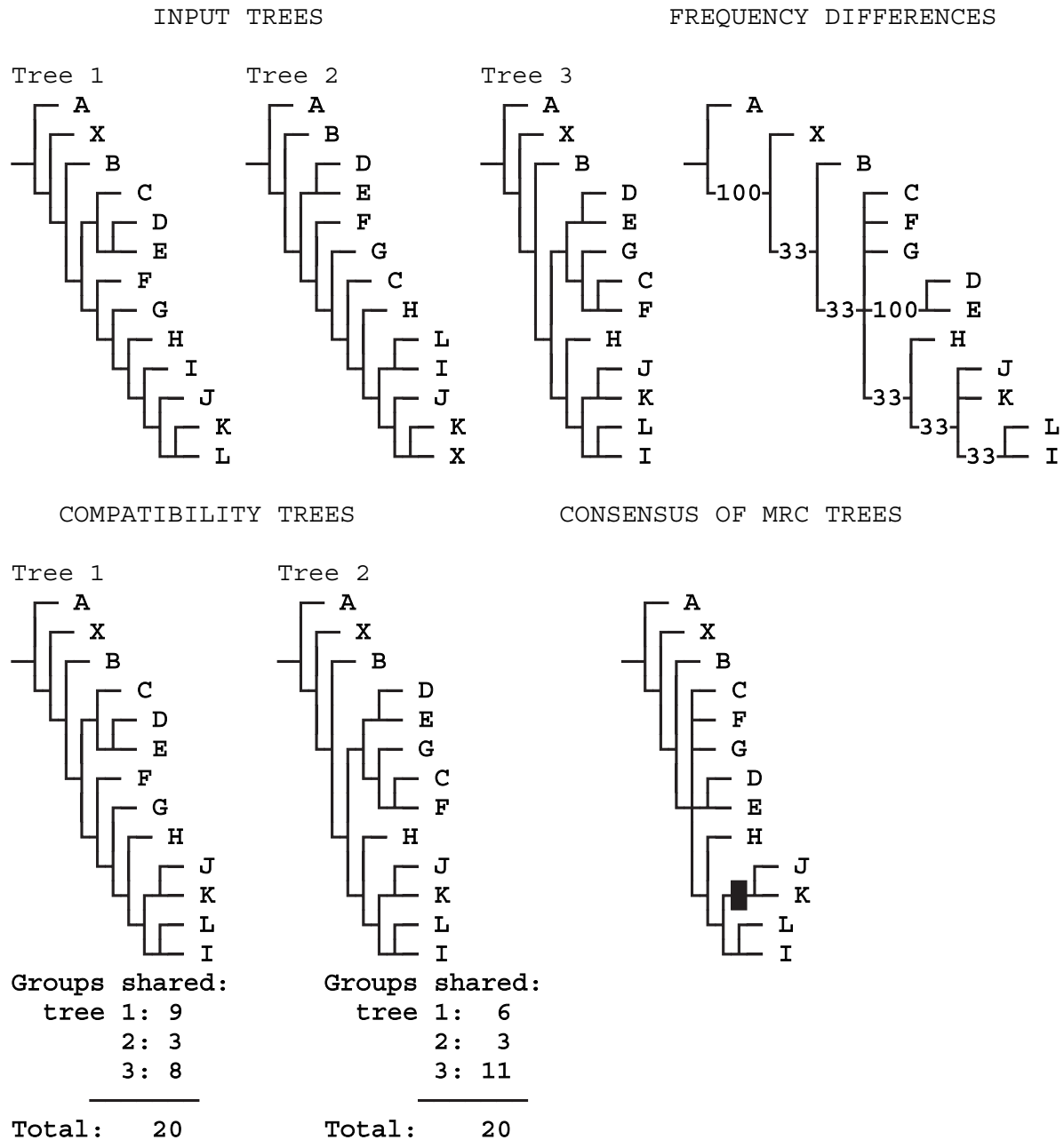
Groups shared:
tree 1: 6
2: 3
3: 11
_____
Total: 20

Fig. 4. A case where MRC displays more resolution than the frequency difference consensus (the additional node is marked). Choosing the supertree that has the maximum number of nodes compatible with the input trees is equivalent in this case, since the input trees are complete and binary.

Minimum flip trees: SPR outperforming TBR?

The method of minimum flipping was introduced by Chen et al. (2003). The idea is to find the trees which require the minimum number of changes of matrix entries for the matrix to represent the tree perfectly (see Eulenstein et al., 2004, for details). Each change to a terminal taxon is counted separately, as it "represents an irreducible item of error" (Eulenstein et al., 2004, p. 300).

The calculation of the flip distances for each tree examined during branch swapping is, in the absence of shortcuts similar to those for parsimony (e.g., Goloboff, 1993b, 1996), rather costly. Eulenstein et al. (2004) report that 5 h were required to calculate a supertree for 96 taxa. They do not specify number of addition sequences or trees saved, but Rainbow (the program they used for their simulations) has a default of five replications saving up to 10 trees per replication; they probably used SPR instead of TBR (see

Tree 1
5 copies

Tree 2
3 copies

Tree 3
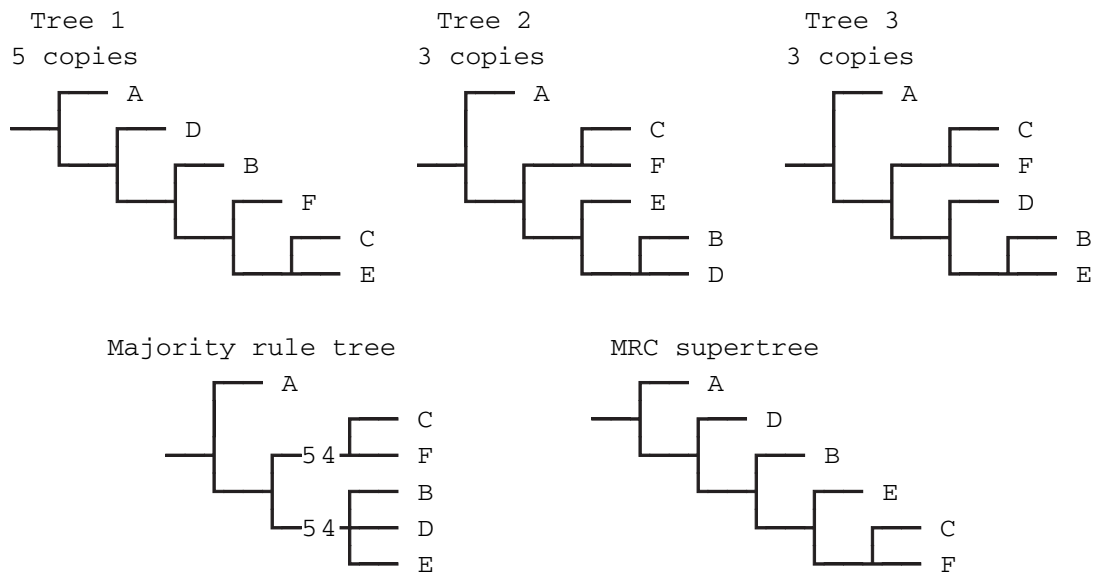3 copies

Majority rule tree

MRC supertree

Fig. 5. A case where MRC produces a supertree which displays groups that contradict the majority rule tree, even if the input trees are complete and binary.

below). As Chen et al. (2003) discussed, one of the reasons for the supertree approach was that finding most parsimonious trees was thought to be computationally difficult for large data sets, but even a thorough parsimony analysis of a "supermatrix" of 96 taxa would take hundreds of times shorter than a superficial analysis under minimum flipping (see Goloboff, 1999, 2000; Nixon, 1999). It is possible that the speed of the algorithms for minimum flipping could be improved in the future, but even so, serious problems with the method remain.

Eulenstein et al. (2004) omitted an important detail of how their program, Rainbow, calculates the distance between the matrix representing the input trees and the matrix representing the (super)tree that is being evaluated: the matrix representing the (super)tree includes the terminals as well. Eulenstein et al.'s Fig. 1.iv shows how to do such evaluation, but includes representation of only the internal nodes. However, the only way to get the numbers of flips calculated by Rainbow is by including the terminal nodes as well. Eulenstein et al. do not discuss why terminals should be included; they may have included them because they produce lower (better) scores, and so including them gives the superficial impression that better results are being obtained. However, the very fact that fit is modified by including those "autapomorphies" suffices to make flipping suspect as a measure of the agreement between trees. The flip scores are lower when the terminals are included because cases of characters representing groups with only two taxa can then automatically be accounted for with just one flip. That comes at the cost, however, of

implying a different maximum possible numbers of flips for groups of different numbers of taxa, and this can introduce distortions in the method (besides other problems; see next section). For comparability, my implementation of flip distances in TNT includes terminals, to match the results from Rainbow, the program used by Eulenstein et al.

One of most surprising findings of Eulenstein et al. is that of the three swapping algorithms (NNI, SPR or TBR) that their program can use to calculate the minimum flip tree, SPR performs best. Since TBR examines each and every rearrangement examined by SPR, there are many situations where SPR would get stuck at a given score, but TBR would be able to find better trees. If SPR usually produced a better agreement with the model tree in Eulenstein et al.'s simulations, this would mean that trees of a worse score produce a better agreement with the model tree, and then this would automatically make the optimality criterion invalid and self-contradictory (according to Eulenstein et al.'s stated goal of recovering the model tree).

The superior performance of SPR would invalidate flipping, however, only if SPR and TBR are both properly implemented. An examination of the program Eulenstein et al. (2004) used for their simulations shows that the TBR branch-swapper outputs trees of highly suboptimal score, even when starting to swap from trees of optimal score. This is so even for perfectly congruent data, for which trees of zero flips exist. Consider the case of two perfectly pectinate trees with 26 taxa (A–Z). Using these two (identical, fully resolved) trees as an input for Rainbow, and swapping with TBR (a single
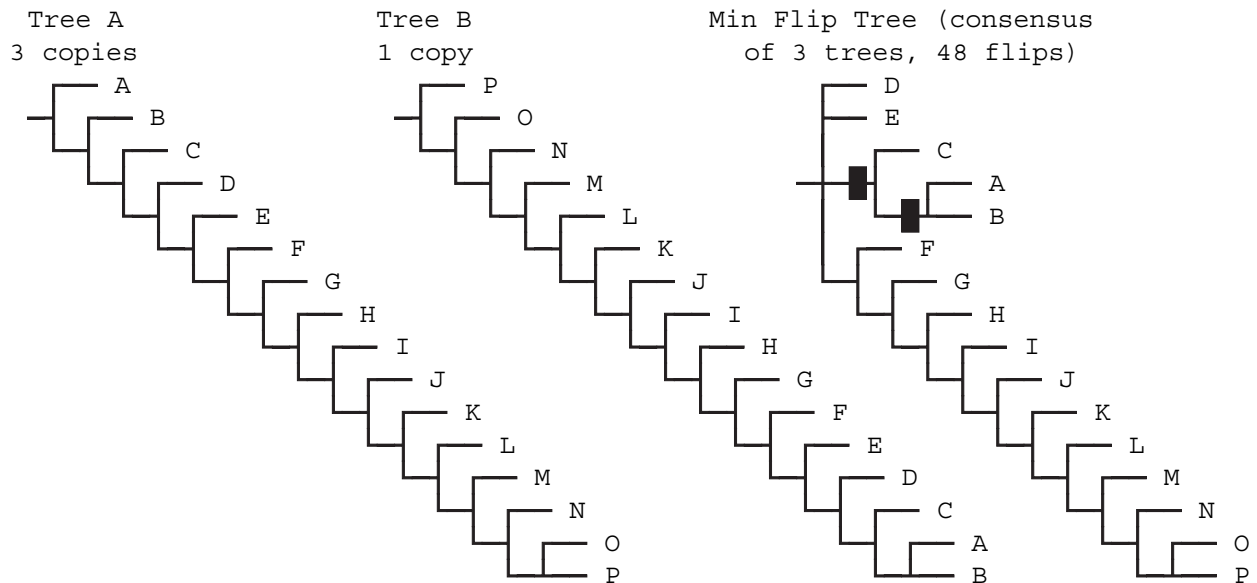
Fig. 6. A case where the Minimum Flip Tree displays groups (marked) supported by one tree and contradicted by three. Trees A and B differ only in their rooting.

replication, saving up to 250 trees[6]) found trees of an average score of 56.47, not 0. The initial tree had a score of 0 (because for such a clean problem, the sequential addition of taxa easily finds the optimal tree). The trees saved subsequently did appear (under visual inspection) to be TBR rearrangements of the original tree, but their scores had obviously been miscalculated; the scores ranged from 0 to 242. The consensus of the 250 trees found by Rainbow was totally unresolved, although the input trees were two identical, perfectly resolved trees. Only the SPR swapper of Rainbow produced the expected, totally resolved supertree.[7]

That Eulenstein et al.'s TBR branch-swapper performs so poorly is at least comforting in one sense: it makes perfectly clear why the simulations using SPR outperformed those using TBR. The problems of Eulenstein et al.'s branch-swapper also illustrate the dangers of testing programs using only complex data sets (like their 48 and 96-taxon simulated data), where it is not evident what result to expect; a very simple case of two identical trees provides a better test, in this case. Using simulations has also been the approach most commonly used to test supertree methods, like MRP (see citations in Bininda-Emonds, 2004b), but simple examples—like those used in this paper—are easier to interpret.

---

[6]I saved only 250 trees because saving much larger numbers of trees often crashes the program. The random seed used in that run was 1086986491, but other seeds have the same problem.

[7]The NNI swapper of Rainbow also has problems, given that a partially resolved supertree is produced in the present case.

Minimum flip trees: conceptual problems

Eulenstein et al. themselves pointed out (2004, p. 300) that the flip distance is not a symmetric measure. To make things worse, although flipping does not count "homoplasy" as such, it does consider that the character representing a given group has varying degrees of fit in the supertree, just like parsimony. Therefore, minimum flip trees may well display groups that are contradicted by each of the input trees, or display groups that are present in the minority of the trees. Figure 6 is an example, with three copies of tree A, and a single copy of tree B. Note that trees A and B differ only in their rooting. The minimum flip tree contains two groups (marked) which are present in one tree and contradicted by three. The groups present in the minimum flip tree may be groups in the minority of the trees, or groups present in no tree at all; Fig. 7 shows an example where the minimum flip tree for three input trees displays eight groups (out of 22) contradicted by each of the input trees.

As in the case of MRP, minimum flipping is rather prone to create groups found in none of the input trees, or in the least numerous topologies, as shown by an examination of 100 triplets of trees for the same 25 taxa (two random topologies, one of those in a double copy, the other in single copy). The average number of groups found in the minimum flip tree but in neither of the input trees was 1.77 (76% of the cases had at least one such group). The average number of groups found in the minimum flip tree and the single tree, but not the duplicate tree, was 0.55 (39% of the cases had at least one such group).
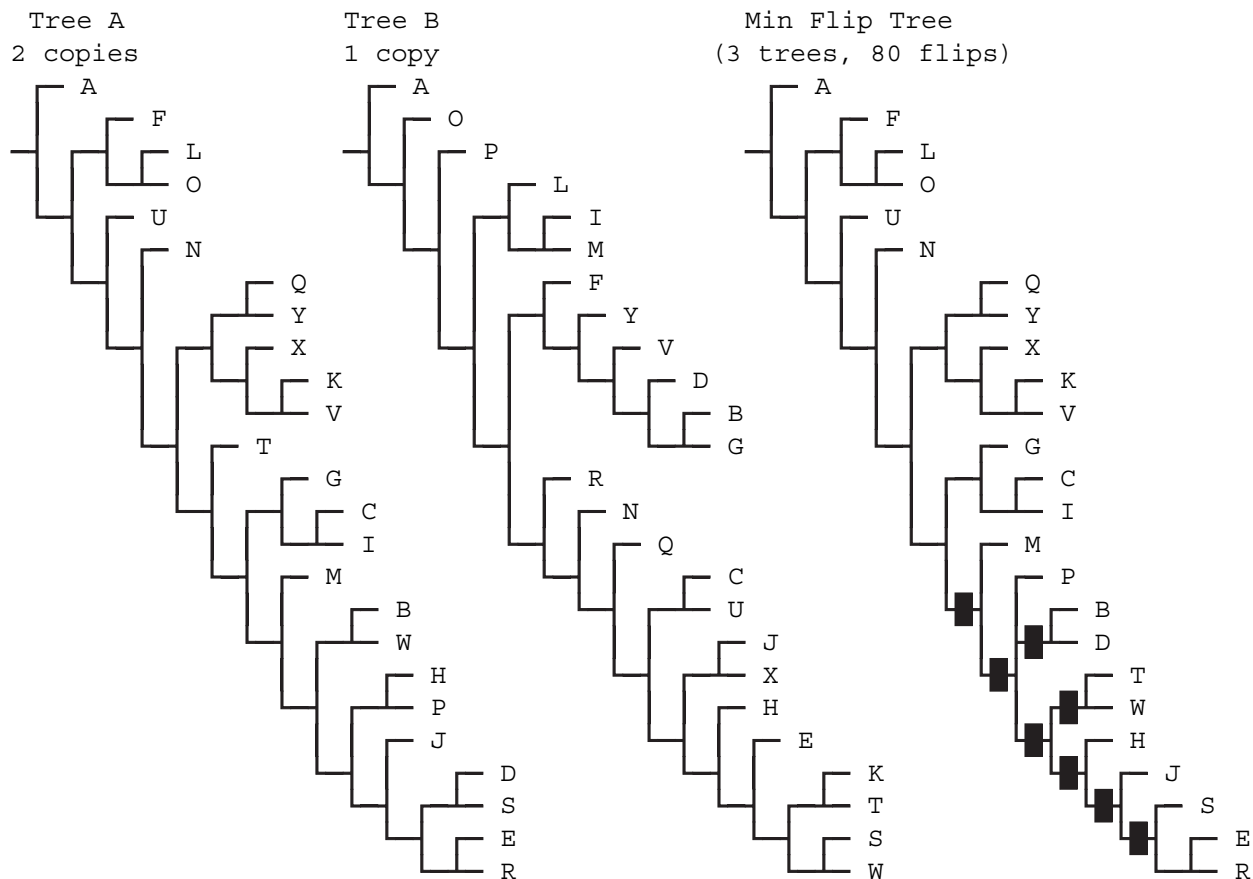
Fig. 7. A case where the Minimum Flip Tree displays groups (marked) found in none of the input trees.

Additionally, counting every single instance of a cell change separately, is bound to create problems. Parsimony avoids doing this (because the counts of *ad hoc* hypotheses of homoplasy would otherwise not be independent; see Farris, 1983 and De Laet, in press, for discussion). Differences in the placement of a given group will be considered by minimum flipping as more significant, or less, depending on the number of members of the group. As an example, consider the two input trees of the top row in Fig. 8, which produce a single (completely resolved) flip tree. That the two trees shown are best combined by that tree of 28 flips is subject to discussion, but the real problem arises when we add nine more taxa, closely related to Q and connected to the same node to which Q connects, forming a polytomy (creating trees 1′ and 2′, in the bottom row). The new trees are, aside from the addition of some taxa, perfectly equivalent to the trees in the top row. However, every item of error counted for Q as a single one is now counted as 10 (by virtue of being "an irreducible item of error"). Thus, the solution produced by minimum flipping is different, and the minimum flip tree(s) define an almost totally unresolved supertree. In cases like the present one, both MRP and MRC correctly detect that nothing has changed, and they produce entirely equivalent solutions in both cases. Another case of strong bias is shown in Fig. 9. Although Eulenstein et al. (2004), based on their simulations, claim that minimum flipping is a viable alternative to MRP, a closer examination suggests that minimum flipping is even more problematic than MRP.

Observations, explanations, and conclusions

Some of the critics of MRP have considered that the minimization of homoplasy makes no sense when the "characters" represent groups. Bininda-Emonds (2004a), in his response to Gatesy et al. (2004), stated that the interpretation of incongruence as homoplasy in a supertree analysis is not necessary. He says that "the principle of parsimony makes no statement regarding either homoplasy or incongruence having to be biologically interpretable" (2004a, p. 357). He doesn't examine similar, but more carefully worded criticisms. For example, Goloboff and Pol (2002) did not criticize the
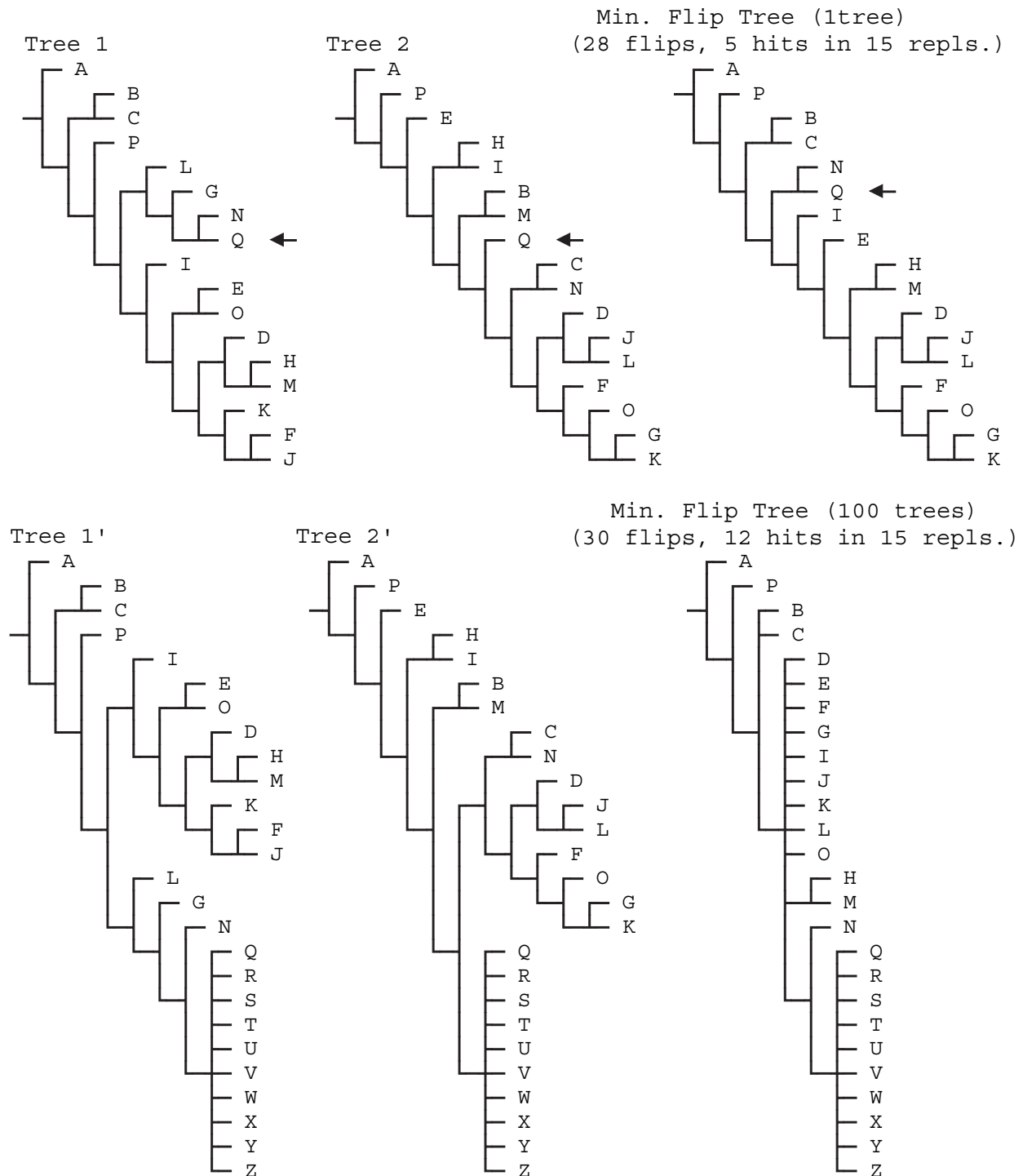
Fig. 8. A case showing the dependence of Minimum Flip Trees to group size. When the group formed by Q alone (top row) is replaced by the larger but equivalent group, the resolution of the minimum flip tree changes (bottom row).

fact that homoplasy in supertree analyses has no biological meaning; instead, in discussing a group A + B appearing in an MRP supertree but contradicted by all the input trees, Goloboff and Pol (2002) said that "A and B are grouped because the "0" in characters 2

and 3 of the matrix can then be accounted for with a single step; that is, A and B are grouped because in input tree 1 they do *not* belong to groups CDEF and DEF" (p. 515, italics in the original). Goloboff and Pol (2002) also analyzed that same example with irreversible

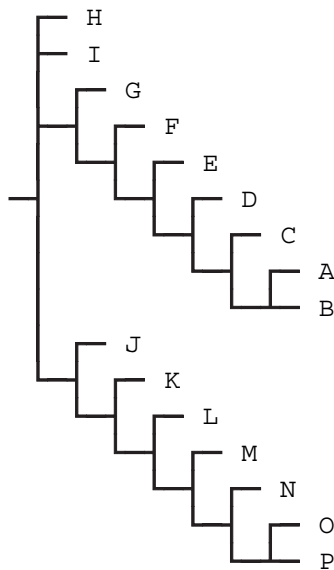Min Flip Tree
(3 trees, 32 flips)

Fig. 9. A case where Minimum Flip shows a strong size bias. The input trees are the same as those in Fig. 6, with a single copy each. In every case in which a group on one of the trees conflicts with a group of larger size in the other tree, the smaller group appears in the Minimum Flip Tree, despite both groups being equally frequent in the input trees.

parsimony (which, according to Bininda-Emonds and Bryant, 1998; could avoid that problem). The (irreversible) MRP tree still displayed the group A + B, now because both A and B are part of larger groups (ABCDEF, ABDEF, and ABF; see Goloboff and Pol, 2002, Fig. 1). Being part of a group ABCDEF, again, provides in itself no grounds for concluding the existence of a group A + B. Defending MRP would require explaining the reasons to consider the exclusion from some groups and the shared membership to some larger groups as grounds for grouping.

Bininda-Emonds' argument, however, has a more fundamental problem than just ignoring some punctual criticism made against MRP. Parsimony, when applied to real data, produces the most explanatory arrangement of taxa. The way in which such an explanation is achieved is not trivial: grouping taxa which share some similarity allows the attribution of that similarity to a common cause (common ancestry). If some other taxon with the same character occurs far away in the tree, the two independent groups still explain the similarity within the members of each group (Farris, 1983). When the "character" simply represents membership to a group, the situation is radically different, for "membership" is what explains observations, not an observation itself. More properly, the conclusion that some taxa form a monophyletic group is required (by parsimony)

to best explain the observations. Using MRP to summarize the results of different analyses amounts to finding the arrangement of taxa that provides the best explanation of the conclusions of those analyses, not the best explanation of observations. The mechanism by which observations are to be explained need not be the same as the mechanism by which conclusions are to be summarized. Bininda-Emonds (2004a, p. 356) himself suggested this much when he proposed that supertree methods (i.e., methods to summarize conclusions) must be judged according to criteria different from the criteria used to judge methods of character analysis (i.e., methods to explain observations). But by Bininda-Emonds' own admission, that parsimony is justified in one case does not mean it is justified in the other. If taxa A and B were (as in the above example) part of a larger group in some of the input trees, and the best arrangement in the supertree has the other members of that group far away... why is the fact that most researchers had considered that A and B are part of a larger group supposed to be best summarized by saying that most researchers considered that A and B form a group? The conclusion of monophyly of the larger group is violated, but the violation is not mitigated by placing some of the members of the group in separate subgroups; this problem also pervades minimum flipping methods (although minimum flipping measures deviation from perfect agreement differently, it still measures it as having degrees, and asymmetrically). In this sense, compatibility analysis of the matrix representing the input trees is (if not without problems) more logical than MRP or minimum flipping, as it is an all-or-none criterion. As noted above, that the researchers must have relied on some character(s) to conclude monophyly of the larger group to which A and B belong is not grounds to say that then placing A and B together will probably better explain those characters; if that argument is to be advanced, then supertrees should be given up altogether, except as a (poor) proxy for a proper analysis of the original data.

Bininda-Emonds' reasoning on the use of parsimony to summarize conclusions is exactly analogous to the reasoning behind the use of parsimony to analyze three taxon statements, which was often justified (e.g., Platnick, 1993) on the grounds that the individual statements of relationships (that is, conclusions!) had to be "explained". Similarly illogical is the use of parsimony in biogeography, where it has been proposed as a way of identifying areas of endemism (Morrone, 1994), and has been strongly criticized (Szumik et al., 2002; Szumik and Goloboff, 2004). The minimization of "homoplasy", as a technique, can be applied to any problem one wishes, but the claim that it is only a way "to find the solution with the minimum amount of incongruence" (Bininda-Emonds, 2004a, p. 357) relies on nothing but the technical aspect of the problem.

While I have nothing against technical treatments in themselves, they are of no help in providing a better understanding of the problems they address unless coupled with a deeper justification.

Acknowledgements

References

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

Baum, B., Ragan, M., 1993. A reply to A.G. Rodrigo's. 'A comment on Baum's method for combining phylogenetic trees'. Taxon 42, 637–640.

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

Bininda-Emonds, O., 2003b. MRP supertree construction in the consensus setting. In: Janowitz, M., Lapointe, F.-J., McMorris, F.R., Mirkin, B., Roberts, F.S. (Eds) DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Bioconsensus. American Mathematical Society, Providence, RI, Vol. 61, pp. 231–242.

Bininda-Emonds, O., 2004a. Trees versus characters and the supertree/supermatrix 'paradox'. Syst. Biol. 53, 356–359.

Bininda-Emonds, O., 2004b. The evolution of supertrees. Trends. Ecol. Evol. 19, 315–322.

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

Bininda-Emonds, O., Gittleman, J., Purvis, A., 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. 74, 143–175.

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

Chen, D., Diao, L., Eulenstein, O., Fernández-Baca, D., Sanderson, M., 2003. Flipping: a supertree construction method. In: Janowitz, M., Lapointe, F., McMorris, F., Mirkin, B., Roberts, F. (Eds). DIMACS Series in Discrete Mathematics and Theoretical Computer Science,

Vol. 61, Bioconsensus. American Mathematical Society, Providence, RI, pp. 135–160.

De Laet, J., in press. Parsimony and the problem of inapplicables in comparative data. In: Albert, V. (Ed.), Parsimony, Phylogeny, and Genomics. Oxford University Press.

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

Farris, J., 1983. The logical basis of phylogenetic analysis. In: Platnick, N., Funk, V. (Eds.), Advances in Cladistics, Vol. 2. Columbia University Press, New York, pp. 7–36.

Farris, J., Albert, V., Källersjö, M., Lipscomb, D., Kluge, A., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics, 12, 99–124.

Felsenstein, J., 1993. Phylip, Version 3.5c. Department of Genetics, University of Washington, Seattle, WA.

Gatesy, J., Baker, R., Hayashi, C., 2004. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. Syst. Biol. 53, 342–355.

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

Goloboff, P., 1993b. Character optimization and calculation of tree lengths. Cladistics, 9, 433–436.

Goloboff, P., 1996. Methods for faster parsimony analysis. Cladistics, 12, 199–220.

Goloboff, P., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics, 15, 415–428.

Goloboff, P., 2000. Techniques for analysis of large data sets. In: De Salle, R., Wheeler, W., Giribet, G. (Eds.), Techniques in Molecular Systematics and Evolution. Birkhauser-Verlag, Basel, pp. 70–79.

Goloboff, P., Farris, J., Nixon, K., 2003. TNT: Tree Analysis using New Technology. Program and documentation, available at http://www.zmuc.dk/public/phylogeny/tnt.

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

Morrone, J., 1994. On the identification of areas of endemism. Syst. Biol. 43, 438–441.

Nixon, K., 1999. The parsimony ratchet, a new method for rapid pasimony analysis. Cladistics, 15, 39–50.

Platnick, N., 1993. Character optimization and weighting: differences between the standard and three-taxon approaches to phylogenetic inference. Cladistics, 9, 267–272.

Ragan, M., 1992. Phylogenetic inference based on matrix representation of trees. Mol. Phyl. Evol. 1, 51–58.

Ross, H., Rodrigo, A., 2004. An assessment of matrix representation with compatibility in supertree construction. In: Bininda-Emonds, O. (Ed.), Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life, Vol. 3. Kluwer Academic, The Netherlands, pp. 35–63.

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

Szumik, C., Cuezzo, F., Goloboff, P., Chalup, A., 2002. An optimality criterion to determine areas of endemism. Syst. Biol. 51, 806–816.

Szumik, C., Goloboff, P., 2004. Areas of endemism: an improved optimality criterion. Syst. Biol. 53, 968–977.

Wilkinson, M., Cotton, J., Creevey, C., Eulenstein, O., Harris, S., Lapointe, F., Levasseur, C., McInerney, J., Pisani, D., Thorley, J., in press. The shape of supertrees to come. Syst. Biol., in press.

Wilkinson, M., Thorley, J., Pisani, D., Lapointe, F., McInerney, J., 2004. Some desiderata for liberal supertrees. In: Bininda-Emonds, O. (Ed.), Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life. Dress, A. (Series Ed.), Computational Biology, Vol. 3. Kluwer Academic, The Netherlands, pp. 227–246.