

Increasing Data Transparency and Estimating Phylogenetic Uncertainty in Supertrees: Approaches Using Nonparametric Bootstrapping

BRIAN R. MOORE, STEPHEN A. SMITH, AND MICHAEL J. DONOGHUE

Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520, USA; E-mail: brian.moore@yale.edu (B.R.M.)

Abstract.—The estimation of ever larger phylogenies requires consideration of alternative inference strategies, including divide-and-conquer approaches that decompose the global inference problem to a set of smaller, more manageable component problems. A prominent locus of research in this area is the development of supertree methods, which estimate a composite tree by combining a set of partially overlapping component topologies. Although promising, the use of component tree topologies as the primary data dissociates supertrees from complexities within the underlying character data and complicates the evaluation of phylogenetic uncertainty. We address these issues by exploring three approaches that variously incorporate nonparametric bootstrapping into a common supertree estimation algorithm (matrix representation with parsimony, although any algorithm might be used), including bootstrap-weighting, source-tree bootstrapping, and hierarchical bootstrapping. We illustrate these procedures by means of hypothetical and empirical examples. Our preliminary experiments suggest that these methods have the potential to improve the correspondence of supertree estimates to those derived from simultaneous analysis of the combined data and to allow uncertainty in supertree topologies to be quantified. The ability to increase the transparency of supertrees to the underlying character data has several practical implications and sheds new light on an old debate. These methods have been implemented in the freely available program, tREeBOOT. [Bootstrap-weighted MRP; hierarchical bootstrapping; nodal support; phylogenetic uncertainty; source-tree bootstrapping; supertree; taxonomic congruence; total evidence.]

Supertree methods have attracted attention as a promising alternative to simultaneous analysis for assembling large trees from disparate data sources (e.g., Sanderson et al., 1998; Bininda-Emonds et al., 2003; Bininda-Emonds, 2004a, 2004b). However, two key issues require attention. First, supertrees tend to be dissociated from the primary character data. Conventional phylogenetic analysis provides a topological summary of typically complex patterns of character covariation within the corresponding data matrix. These topological summaries, in turn, provide the raw data for supertree methods, which effectively prevent sub-signals within the component data sets from contributing to the supertree estimate. Consequently, combining a set of source-tree topologies may yield a supertree that is not sanctioned by a simultaneous analysis of the component data sets. We refer to this as the *data-dissociation problem*. Second, it is not clear how best to assess uncertainty in supertree estimates (e.g., Bininda-Emonds, 2003; Wilkinson et al., 2005a; Cotton et al., 2005; Burleigh et al., 2006), which greatly diminishes their utility for systematic and comparative studies.

We address these issues here, considering various means of incorporating nonparametric bootstrapping to both increase the transparency of supertrees to their underlying character data and to gauge uncertainty in supertree estimates. The alternative bootstrap approaches are illustrated by means of hypothetical examples and empirical experiments from the plant group Dipsacales comparing the supertree topologies and bootstrap proportions derived under these alternative bootstrap approaches to those estimated from simultaneous analysis of the component data. Although we focus on matrix representation with parsimony (MRP: Baum, 1992; Ragan, 1992), the approaches we describe are applicable to other supertree methods.

MATERIALS AND METHODS

Supertree Bootstrapping Procedures

Bootstrap-weighted MRP.—Conventional MRP takes as input a set of two or more species phylogenies (i.e., “source trees”) with partially overlapping taxa, which are transposed using additive binary coding (Sokal and Sneath, 1963) to construct a matrix comprising one column (i.e., “matrix element”) for every node in the set of source trees and one row for every terminal taxon in the source-tree superset (Fig. 1). The resulting MRP matrix is then subjected to standard parsimony algorithms to estimate the optimal supertree(s). Under this scheme, each matrix element is arbitrarily assigned an equal weight, such that two conflicting matrix elements will (theoretically) impart an equal influence on the supertree topology, regardless of possible differences in the underlying character support for their corresponding source-tree nodes.

Several authors (e.g., Ronquist, 1996; Bininda-Emonds and Bryant, 1998; Sanderson et al., 1998) have suggested that the correspondence of supertree topologies to the underlying character data might be enhanced by scaling each matrix element by the bootstrap support value of the corresponding source-tree node (Fig. 1). In principle, bootstrap-weighted MRP should arbitrate conflicts among the set of source trees such that the supertree topology is resolved in favor of more strongly supported source-tree nodes. Indeed, this approach appears to improve the accuracy of MRP supertree estimation under simulation (e.g., Bininda-Emonds and Sanderson, 2001). However, this approach does not allow nodal support in the supertree to be evaluated.

Source-tree bootstrapping.—Because a collection of source-tree topologies constitutes the raw data for supertree estimation, these observations might be

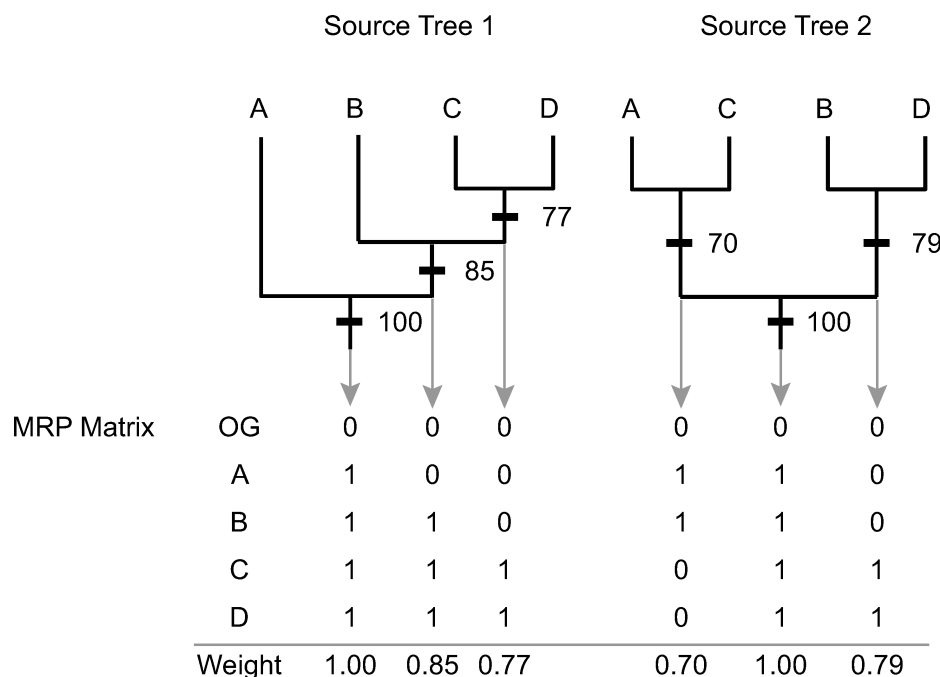


FIGURE 1. Conventional and bootstrap-weighted MRP. Each node in the set of source trees contributes an element to the MRP matrix, which is scaled by the corresponding nodal-support value under the bootstrap-weighting scheme (i.e., the row of weights at the bottom of the matrix). OG = outgroup.

bootstrapped to evaluate the degree of conflict among them (e.g., Daubin et al., 2001; Creevey and McInerney, 2005). In outline, a set of source trees is compiled by randomly and repeatedly drawing (with replacement) from the original set of estimated source trees, constructing an MRP matrix from this bootstrap set of source trees, and finally estimating a supertree from this matrix. This procedure is repeated an arbitrary number of times to generate a bootstrap profile of supertrees, which is then summarized using majority-rule consensus such that the frequencies of nodes in the consensus supertree provide an estimate of the degree of congruence among the original set of source-tree topologies. We refer to this approach as *source-tree bootstrapping*, which is implemented by the following algorithm (Fig. 2):

1. Estimate the optimal tree(s) for each of the m source-tree data matrices.
2. Randomly select from among the m source-tree sets (with $P = 1/m$). If there are multiple trees in the chosen source-tree set (e.g., reflecting equally optimal estimates), randomly select a tree from this set of k trees (with $P = 1/k$). Note that this is statistically equivalent to reciprocally weighting the chosen tree (by $1/k$) when averaged over many draws.
3. Iterate step (2) m times.
4. Construct an MRP matrix from the set of m randomly selected source trees.
5. Estimate the optimal supertree(s) from the bootstrap MRP matrix and append the resulting supertree(s) to a file.
6. Iterate steps (2) through (5) r times (e.g., 1000 replicates).

7. Summarize the bootstrap profile of supertrees using majority-rule consensus. As in conventional bootstrap consensus, multiple trees derived for a given bootstrap replicate are reciprocally weighted. The frequency of clades in this consensus supertree provides an estimate of the topological consistency among the set of source trees that is analogous to standard bootstrap proportions (e.g., Felsenstein, 1985).

We note that source-tree bootstrapping can also be based directly on previously estimated phylogenies (obviating step 1) when source-tree data matrices are unavailable for reanalysis.

Hierarchical bootstrapping.—Rather than drawing from point estimates of each source tree (as implemented in source-tree bootstrapping), we might instead exploit the set of bootstrap profiles for each of the source trees. That is, we could bootstrap from the set of source-tree bootstrap profiles (e.g., Cotton and Page, 2002; Page, 2004). In outline, this approach entails randomly and repeatedly sampling (with replacement) a set of source trees from their respective bootstrap profiles, constructing an MRP matrix from this bootstrap set of source trees, and finally estimating the optimal supertree(s) from this matrix. This procedure is repeated an arbitrary number of times to generate a bootstrap profile of supertrees, which is then summarized by majority-rule consensus such that the frequencies of nodes in the consensus supertree provide an estimate of the level of congruence among the original set of source-tree data matrices. We refer to this approach as *hierarchical bootstrapping*, which is implemented by the following algorithm (Fig. 3):

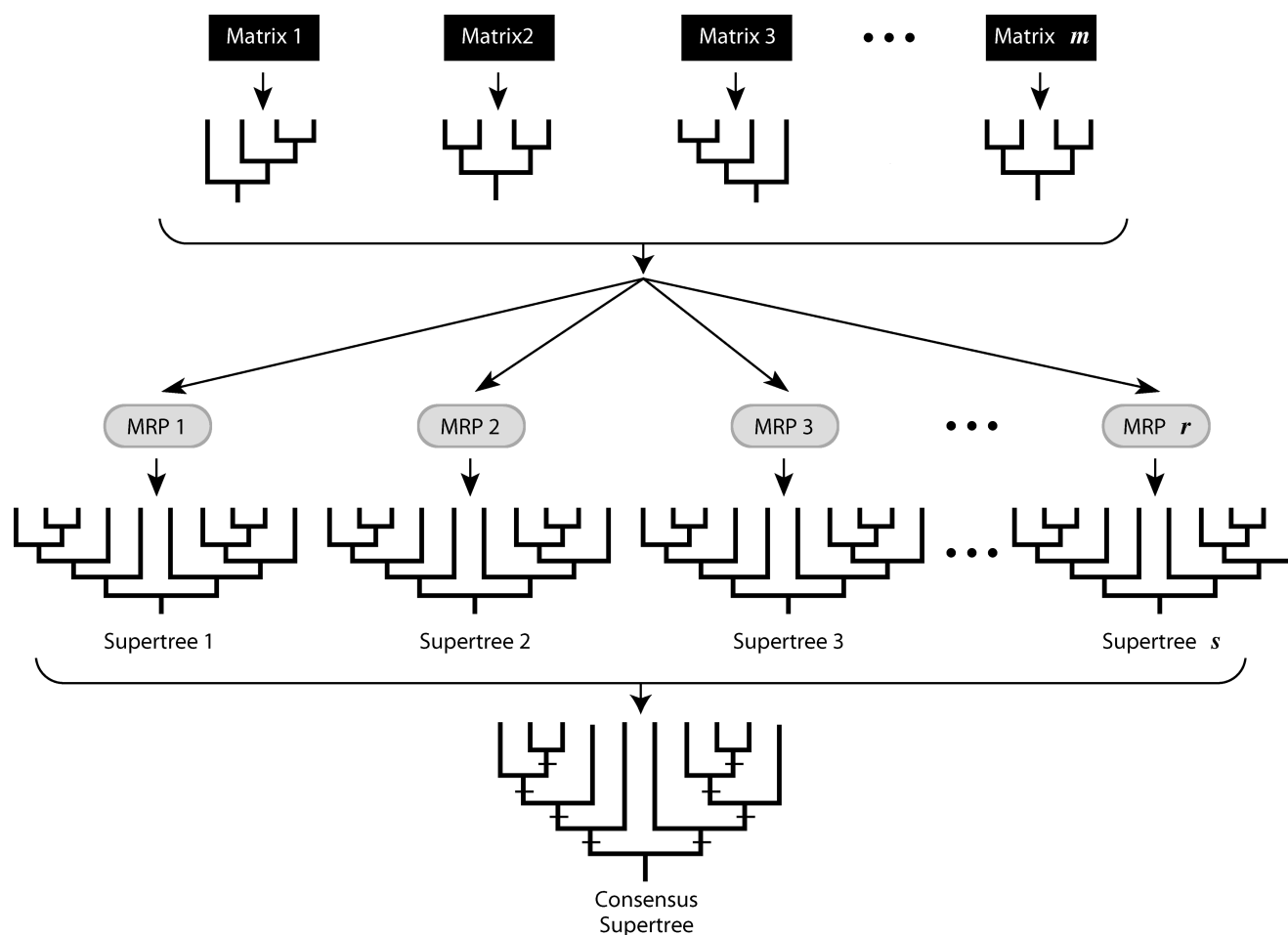


FIGURE 2. Source-tree bootstrapping. A set of m random draws (with replacement) are made from the set of source trees. Each set of source trees is then translated into an MRP matrix, from which a supertree is estimated and appended to an array. This procedure is replicated r times. The resulting array of s supertrees is summarized by majority-rule consensus such that the frequencies of nodes in the consensus supertree provide an estimate of the topological congruence among the set of source-tree topologies. This approach is referred to as *EPT-based source-tree bootstrapping* when samples are drawn from the entire set of equally parsimonious source trees, and as *MRC-based source-tree bootstrapping* when samples are drawn from the set of majority-rule consensus of those equally parsimonious source trees. See text for details.

1. Generate n bootstrap pseudo-matrices for each of the original m source-tree data matrices and estimate the optimal tree(s) for each replicate.
2. Randomly draw a tree from the first source-tree bootstrap profile (with $P = 1/n$). If more than one tree is associated with the selected replicate (e.g., reflecting a set of equally parsimonious trees), randomly select a tree from among the set of k trees (with $P = 1/k$). This is statistically equivalent to reciprocally weighting the chosen tree (by $1/k$) when averaged over many draws.
3. Iterate step (2) for each of the m source-tree bootstrap profiles.
4. Construct an MRP matrix from the set of m randomly selected source trees.
5. Estimate the optimal supertree(s) from the bootstrap MRP matrix and append the resulting supertree(s) to a file.
6. Iterate steps (2) through (5) r times (e.g., 1000 replicates).
7. Summarize the bootstrap profile of supertrees using majority-rule consensus. As in conventional bootstrap consensus, multiple trees derived for a given bootstrap replicate are reciprocally weighted. The frequencies of clades in the consensus supertree provide an estimate of their underlying character support that is analogous to standard bootstrap proportions (e.g., Felsenstein, 1985).

A Simple Experimental Design to Explore Supertree Bootstrapping Procedures

We present results from hypothetical examples and empirical data to address two questions: (1) Is it possible for bootstrap procedures to improve the performance of MRP? and (2) Is it possible to reflect the underlying character support in the estimated supertree topology? We pursue these questions by comparing the similarity of the supertree topologies and/or bootstrap proportions derived under the various bootstrap approaches to those

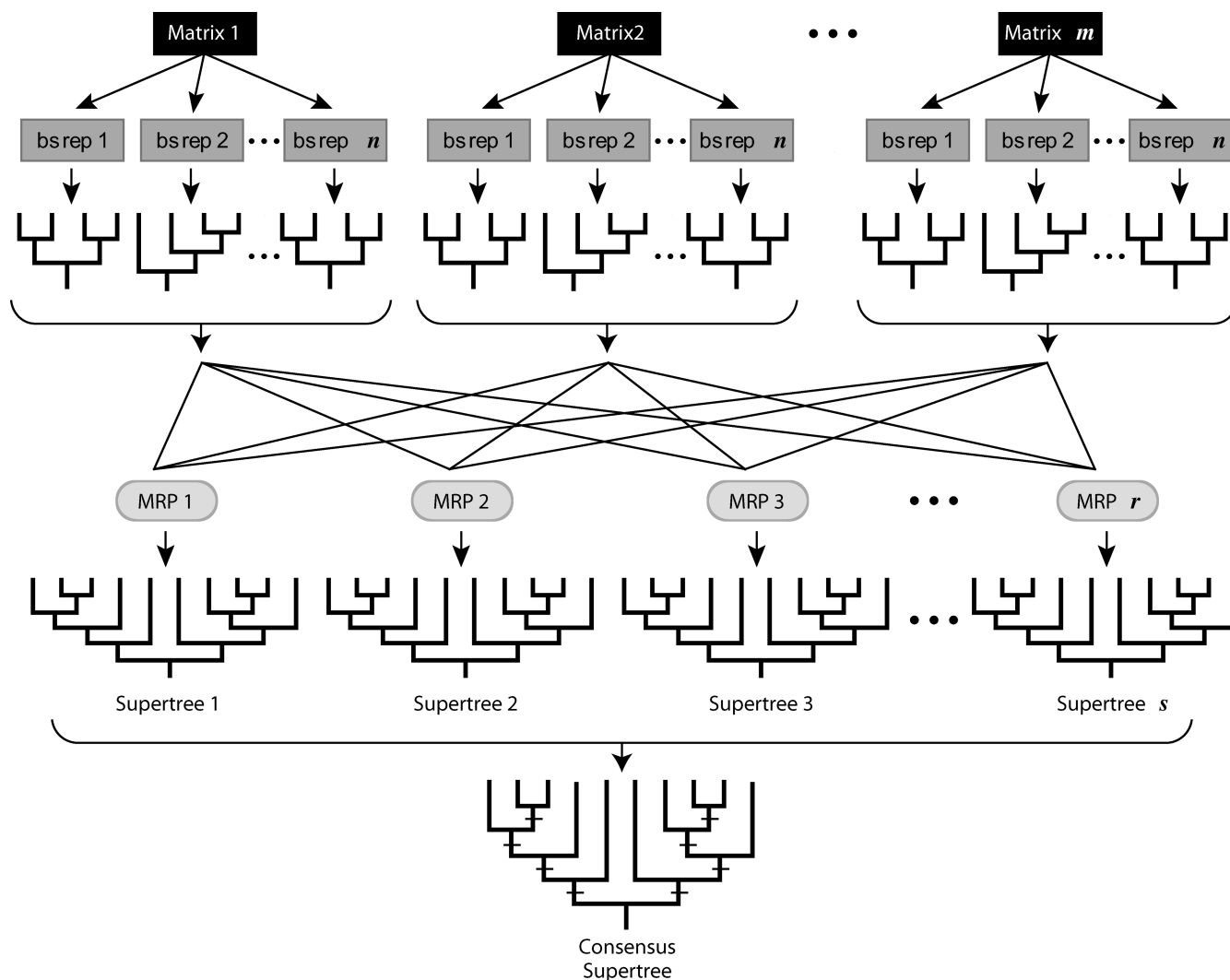


FIGURE 3. Hierarchical bootstrapping. This procedure is similar to that of source-tree bootstrapping (illustrated in Fig. 2), but differs in two important respects. First, each replicate set of source trees is drawn randomly from the set of m source-tree bootstrap profiles, such that each source-tree profile contributes one tree to each bootstrap replicate (with $P = 1/n$). Second, the frequencies of clades in the consensus supertree rendered under this approach provide an estimate of the underlying character support within the set of source-tree data matrices that is analogous to conventional bootstrap proportions. See text for details.

based on a simultaneous analysis of the primary data sets and also to estimates derived with a number of other conventional supertree methods

Some aspects of our experimental design warrant comment. First, our analyses involve source trees with the same set of terminal taxa for each component data set. This contrasts, of course, with the normal application of supertree methods, which are typically used to combine source trees with only partially overlapping taxa. However, our approach factors out the potentially confounding effects introduced by differential taxonomic overlap and, thus, provides a baseline for the future study of such cases. Second, we compare estimates derived under the various bootstrap procedures to those obtained from simultaneous analyses of the primary character data. Although the true value of the target parameter is generally

unknown for empirical inferences, we suspect that most systematists would generally be satisfied if supertree methods performed as well (or as poorly) as simultaneous analysis. Accordingly, the topologies and bootstrap proportions derived by simultaneous analysis provide the references against which the “accuracy” of various bootstrap approaches are empirically evaluated.

General considerations.—We deliberately set out to identify a set of data partitions with an identical set (and relatively small number) of terminal taxa that conflicted strongly on the relationships among a subset of those taxa. This might seem an odd strategy with which to explore the behavior of supertree estimation methods, which more typically involve partial overlap among a collectively large number of taxa. Nevertheless, our approach offered several advantages: (1) complete

taxonomic overlap among component data sets controlled for potential biases in supertree methods associated both with differential taxonomic overlap and with differential source-tree size (e.g., Purvis, 1995; Ronquist, 1996; Bininda-Emonds and Bryant, 1998); (2) the identical taxon sample among the component data sets also decreased the incidence of (and potential bias associated with) missing data in the simultaneous analyses of the component data sets (to which the various supertree estimates were compared); and (3) the relatively small number of taxa allowed a degree of experimentation that would have otherwise been computationally prohibitive (which should help identify the relevant subset of issues to be explored by future studies of larger, more complex data sets), while still enabling relatively thorough analyses in all of these experiments (which should reduce the incidence of spurious and potentially confounding estimates stemming from insufficiently rigorous phylogenetic analyses). Unless stated otherwise, all of the analyses described below were performed using parsimony as the optimality criterion (with all characters weighted equally) with PAUP* v. 4.0 (Swofford, 2002). The data sets referred to are available from <http://www.phylodiversity.net/bmoore/resources.html>.

Delineation of component matrices.—A global data set was compiled from recent studies of Dipsacales phylogeny that included one morphological and nine DNA sequence partitions (8407 characters with no missing data) for 25 taxa. Sequence partitions were aligned using ClustalX (Thompson et al., 1997) and adjusted manually where necessary. We performed an exhaustive series of pairwise partition homogeneity tests ($N = 45$) to identify compatibility issues, which failed to detect significant conflict among the set of partitions. As a baseline, we first subjected the entire matrix to exact and bootstrap analyses. The exact analysis was performed using the branch-and-bound algorithm, and the bootstrap analysis entailed heuristic searches of 10^3 replicate data sets, each search involving 10^3 random-taxon addition starting trees that were subjected to TBR branch swapping. We then performed a set of analyses to identify conflicting components within the global matrix: 11 data partitions (comprising the 10 individual partitions and the combined DNA alignment) were each subjected to exact and bootstrap analyses (executed as described above). These analyses identified three components (morphol-

ogy, *atpB*, and the *rbcL* coding region) that conflicted strongly with the simultaneous estimate primarily regarding relationships within two clades: Dipsacaceae and Linneaeae. These conflicting components were then combined under the following four partition-set scheme: a single 2-partition set (all DNA | morphology), two 3-partition sets (set A: morphology | *atpB* | DNA complement 1; set B: morphology | *rbcL* | DNA complement 2), and one 4-partition set (morphology | *atpB* | *rbcL* | DNA complement 3). Properties of the component data sets are summarized in Table 1.

Estimation of component phylogenies.—The four partition sets comprised a total of 7 unique component data sets, each of which was subjected to a second round of analyses. Exact analyses were performed by branch-and-bound, and conventional bootstrap analyses entailed exact searches of 10^3 replicate data sets. This resulted in three sets of trees for each component of the four partition sets: equally parsimonious trees from the exact analyses were summarized using both strict and majority-rule consensus, and the profile of trees from the conventional bootstrap analyses were summarized (as is standard practice) by majority-rule consensus. Properties of trees estimated from the 7 component data sets are summarized in Table 1.

Combination of component phylogenies using bootstrap supertree methods.—Component trees for each of the four partition sets were combined under three main bootstrap approaches: bootstrap-weighted MRP, source-tree bootstrap MRP, and hierarchical bootstrap MRP.

The bootstrap-weighted MRP analyses entailed encoding each set of component trees as an MRP matrix and scaling the resulting matrix elements by the bootstrap values of their corresponding source-tree nodes (using tREeBOOT; see below). The resulting scaled-MRP matrix was then subjected to an exact (branch-and-bound) analysis. Equally optimal supertrees were summarized by strict consensus.

Source-tree bootstrap analyses were performed on two sets of trees: one set of analyses was based on the majority-rule consensus of the set of equally parsimonious trees from exact searches of the component data matrices (i.e., MRC-based source-tree bootstrapping), and the second set of analyses was based on the entire set of those equally parsimonious trees (i.e., EPT-based source-tree bootstrapping). Source-tree

TABLE 1. Properties of component data sets and their associated trees.

Partition	N_C^a	N_{PIC}^b	N_{EPT}^c	2-Source-tree partition set	3-Source-tree partition set A	3-Source-tree partition set B	4-Source-tree partition set
morph	97	82	77	1	1	1	1
<i>atpB</i>	979	296	12		2		2
<i>rbcL</i>	1428	110	2			2	3
All DNA	8310	1862	1	2			
DNA comp 1	5904	1456	1		3		
DNA comp 2	7332	1566	4			3	
DNA comp 3	6882	1752	1				4

^aNumber of characters. ^bNumber of parsimony-informative characters. ^cNumber of equally optimal trees.

bootstrapping involved randomly sampling (with replacement) from the (MRC or EPT) set of source trees until a bootstrap sample equal in size to the original had been generated. An MRP matrix was then derived from each bootstrap sample. In the case of EPT-based source-tree bootstrapping, sampled trees were reciprocally weighted according to the number of equally parsimonious trees to which they belonged (e.g., whenever 1 of the 12 equally parsimonious trees for the *atpB* data set was randomly drawn, it was assigned a weight of $1/12$), which entails scaling the set of matrix elements for that tree. For every analysis, 10^3 source-tree bootstrap MRP replicates were generated, which were combined as a batch file. Each MRP matrix was subjected to an heuristic search in which 10^3 starting trees were generated by random-taxon addition and subjected to TBR branch swapping. The optimal supertree(s) found for each of the 10^3 replicate MRP matrices were appended to a file, and the resulting supertree bootstrap profile was then summarized by majority-rule consensus. As in conventional bootstrap consensus, reciprocal weighting of equally parsimonious (super) trees was used.

Hierarchical bootstrapping analyses involved randomly sampling (with replacement) a single tree from each of the bootstrap profiles for the set of source trees. An MRP matrix was then derived from this bootstrap sample of trees. This process was iterated 10^3 times, and the resulting MRP matrices were then combined as a batch file. Each MRP matrix was subjected to an heuristic search in which 10^3 starting trees were generated by random-taxon addition and subjected to TBR branch swapping. The optimal supertree(s) found for each of the 10^3 replicate MRP matrices were appended to a file, and the resulting supertree bootstrap profile was then summarized by majority-rule consensus using reciprocal weighting. A second series of hierarchical bootstrap analyses was performed as above, except that each tree in the bootstrap sample was scaled by the proportion of parsimony-informative characters within its component data matrix.

To assess the effect of sampling intensity on the bootstrap supertree estimates, we repeated all of the source-tree and hierarchical bootstrap analyses described above using 10^4 and 10^5 replicates. The results obtained from these larger bootstrap samples were essentially indistinguishable from those based on 10^3 replicates, which are the source of the results presented in this paper.

Combination of component phylogenies using conventional supertree methods.—For the purpose of comparison, we also combined the set of source trees for each partition set using several conventional supertree algorithms, including standard MRP (Baum, 1992; Ragan, 1992), matrix representation with flipping (MRF; Chen et al., 2002, 2003; Burleigh et al., 2004), MinCut (MC; Semple and Steel, 2000), and modified MinCut (MC*; Page, 2002). Our rationale for choosing these four methods from the growing pool of available supertree approaches (currently comprising more than a dozen alternatives; e.g., Wilkinson et al., 2005b; Bininda-

Emonds, 2004a) deserves comment: MRP is the most frequently used method (e.g., Bininda-Emonds, 2004b), MinCut and modified MinCut methods share the unique (and highly desirable) property of running in polynomial time on species number (e.g., Semple and Steel, 2000; Page, 2002), and MRF has typically outperformed other methods under simulation (e.g., Bininda-Emonds and Sanderson, 2001; Chen et al., 2002; Eulenstein et al., 2004).

MRP and MRF analyses require that the set of source trees first be encoded as an MRP matrix (accomplished using tREeBOOT). MRP matrices were subjected to exact (branch-and-bound) analyses. MRF analyses involved a series of heuristic searches (using Rainbow v.b1.3; Chen et al., 2004): in each search, 10^3 starting trees were generated by random taxon addition, and each starting tree was then subjected to 10^6 branch-swap perturbations. To reduce the probability of entrapment in a local optima, we performed replicate heuristic searches with TBR, SPR, and NNI branch-swapping algorithms. The MC and MC* analyses were performed using Supertree 0.3 (Page, 2002). Searches resulting in multiple, equally optimal supertrees were summarized using strict consensus. The entire set of supertree analyses was repeated using each of the three sets of source trees inferred by the component analyses (see *Estimation of Component Phylogenies*).

RESULTS

Hypothetical Examples

As predicted, scaling MRP matrix elements by the bootstrap proportions of the corresponding source-tree nodes in the two hypothetical examples (Figs. 4 and 5) allows more strongly supported nodes to exert a proportionate influence of the inferred supertree topology. However, this approach may either increase or decrease the correspondence of the supertree to the phylogeny inferred by a simultaneous analysis of the component matrices: bootstrap-weighting improved upon the MRP estimate in one example (Fig. 5) but resulted in a supertree not sanctioned by a simultaneous analysis of the data in the other (Fig. 4).

Source-tree bootstrapping performed poorly in both hypothetical examples. This is somewhat unsurprising given that this approach should allow more frequent source-tree topologies to contribute proportionately to the inferred supertree topology. In these examples, the conflicting source-tree topologies occur in equal frequency, such that the method is unlikely to improve the supertree estimate. In the first example, the supertree inferred using source-tree bootstrapping is identical to one of the source trees (Fig. 4), whereas the supertree inferred in the second example is different from either source tree (presumably owing to anomalies of the MRP algorithm). In both examples, the bootstrap values for the single correctly inferred node (subtending taxa BCD) was inflated relative to the values derived from simultaneous analysis.

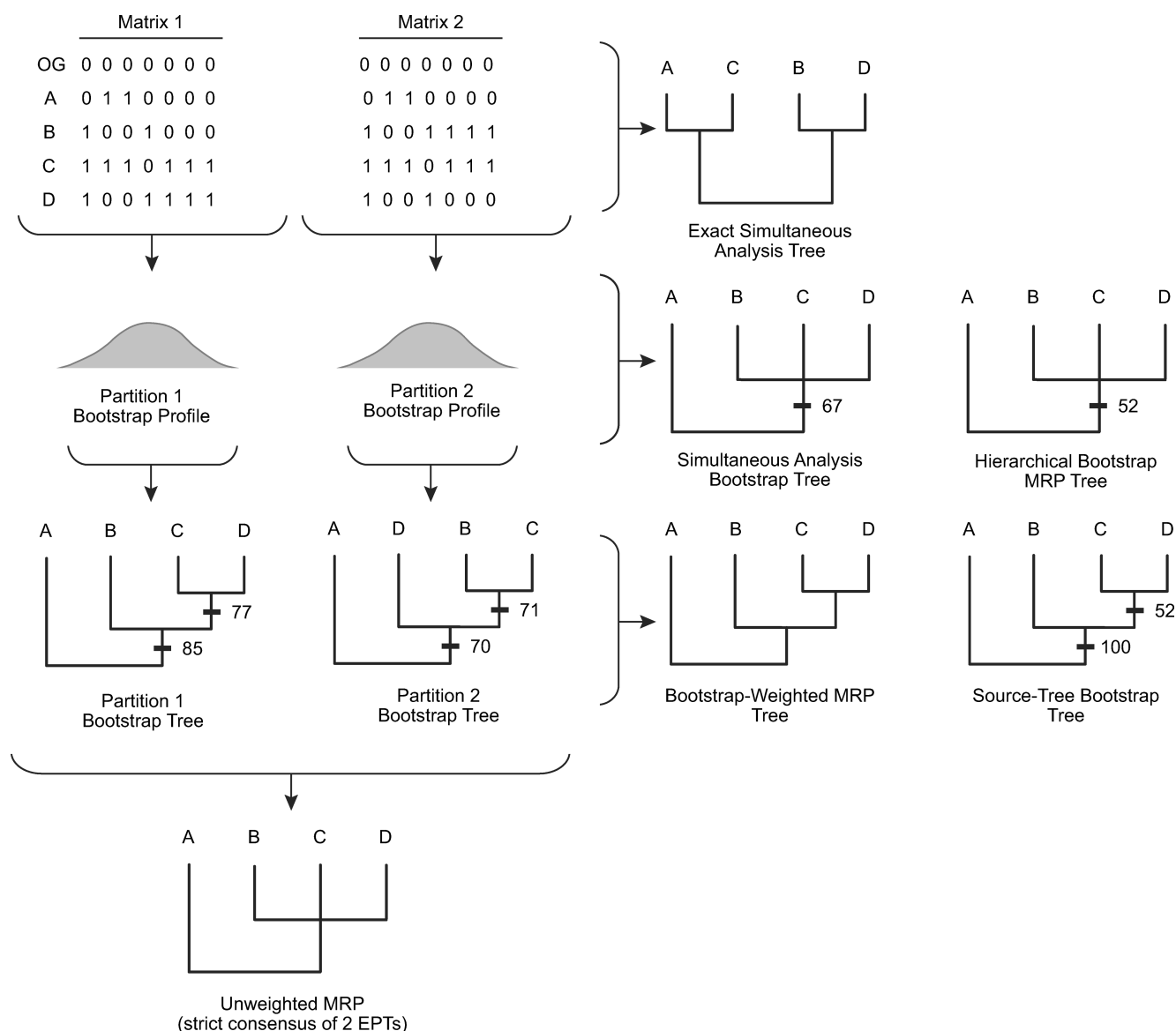


FIGURE 4. A hypothetical example illustrating the proposed bootstrap approaches (modified from Barrett et al., 1991; see also Pisani and Wilkinson, 2002). The example was originally contrived to illustrate the phenomenon of “weak-signal enhancement”: congruent sub-signals in the two component matrices combine to become the dominant signal when the component matrices are analyzed simultaneously. The two trees derived from separate analyses differ from that sanctioned by the simultaneous analysis. Although the latter topology is not recovered by the supertree bootstrapping approaches, neither is it recovered by a conventional bootstrap analysis. The estimate based on hierarchical bootstrapping is identical to the simultaneous bootstrap tree (i.e., the target tree). In this example, the strict consensus of the two equally parsimonious trees found by unweighted MRP is also identical to the target tree, but the estimate derived by bootstrap-weighted MRP is not (but is identical to the source tree with the higher bootstrap proportions).

Hierarchical bootstrapping fared better in the hypothetical examples. The supertree topology inferred using this approach is identical to that derived from a bootstrap analysis of the combined source-tree data matrices in the first example (Fig. 4), whereas the supertree inferred in the second example was compatible with (but less resolved than) the simultaneous bootstrap estimate (Fig. 5). Apparently, the failure to recover the target topology in the second example reflects the inability of the hierarchi-

cal bootstrap to adequately gauge the weight of evidence in the conflicting source-tree bootstrap profiles. The bootstrap can be interpreted as summarizing two aspects of a given data matrix: the “quantity” of data (the weight of evidence) and the “quality” of the data (the level of homoplasy). The latter component is partially manifest by the number of equally parsimonious trees derived from the replicate data sets (the scatter of the bootstrap profile), but the former term is poorly represented by the profile.

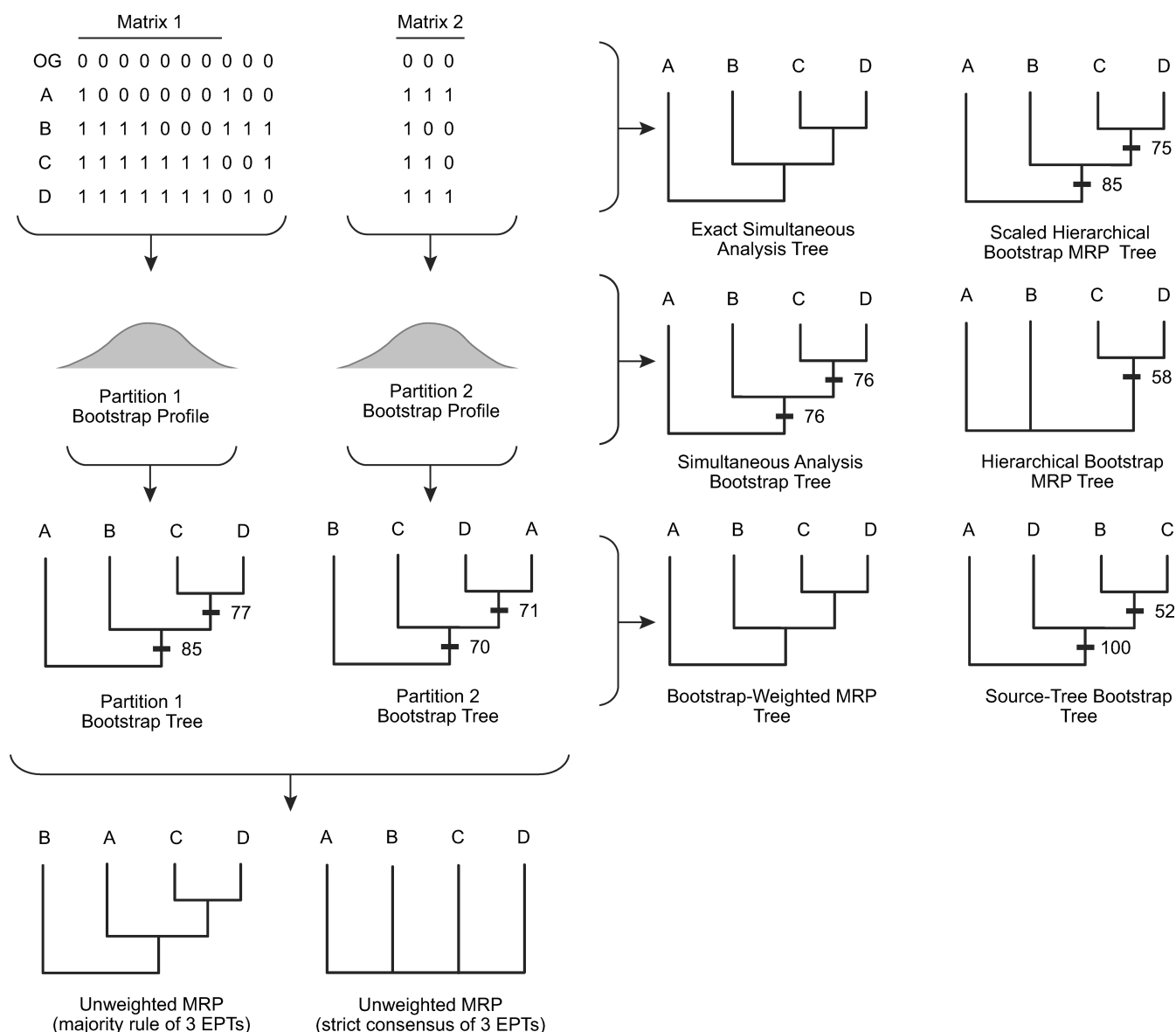


FIGURE 5. A second hypothetical example illustrating additional aspects of the proposed bootstrap approaches. Bootstrap analyses of the two data partitions in this example produce bootstrap proportions similar to those of the first example (Fig. 4) despite differences in the weight of character evidence between the matrices. The hierarchical bootstrap tree is compatible with (but less resolved than) the simultaneous bootstrap tree (i.e., the target tree), but converges on the target tree when scaled to reflect the relative weight of evidence in the two source-tree data matrices. In contrast to the first example, the tree based on unweighted MRP in this case differs from the target tree, but the bootstrap-weighted MRP tree is identical to the target (however, it is also still identical to the source tree with the higher bootstrap proportions).

Consequently, matrix 2 in the second example exerts a disproportionate influence on the supertree inferred by hierarchical bootstrapping. We might attempt to compensate for this by scaling the hierarchical bootstrap profiles by some measure proportional to the weight of evidence in each source-tree data matrix. For example, although admittedly crude, the sampled source trees could be scaled by the proportion of parsimony informative characters in the corresponding data matrices. Application of this scaled hierarchical bootstrap procedure to the

second example increased the correspondence of the resulting supertree topology and bootstrap proportions to those estimated from a bootstrap analysis of the combined data (Fig. 5).

An Empirical Example: Dipsacales

Topological distance.—The similarity of alternative supertree estimates to the topology inferred by a simultaneous analysis of the component data is summarized

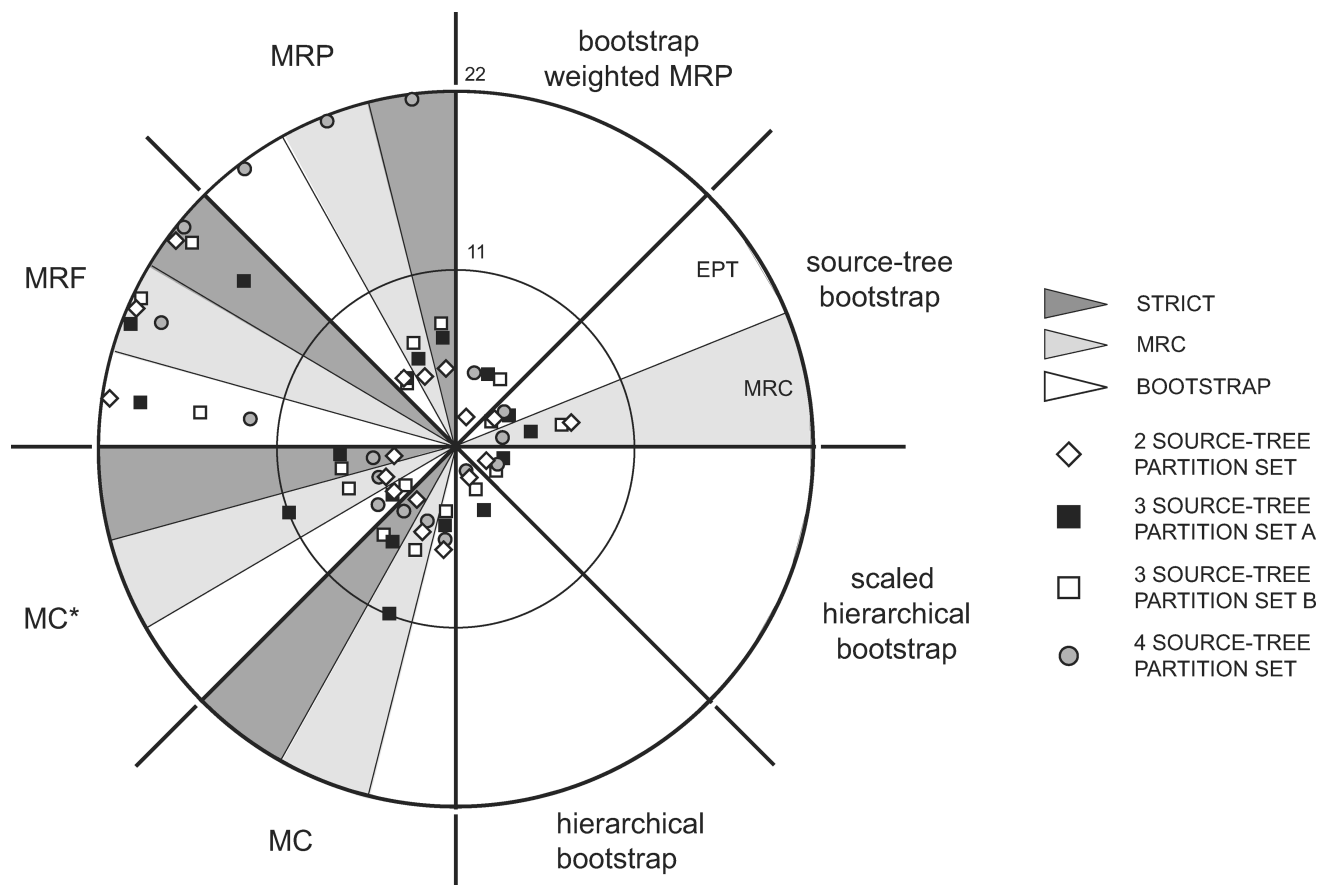


FIGURE 6. Topological similarity of supertree estimates to a target phylogeny based on a simultaneous analysis of the combined dataset for the plant group, Dipsacales. The target tree is located at the center of the graph, and each supertree is plotted within (sub)sectors for each of the supertree estimation methods. The radial distance from each supertree to the center of the graph reflects its topological distance to the target tree in units of NNI (nearest-neighbor interchange) branch swaps. Abbreviations for supertree methods: MC = MinCut; MC* = modified MinCut; MRP = matrix representation with parsimony; MRF = matrix representation with flipping. Abbreviations for sets of source trees from which supertree estimates were derived: EPT = set of equally parsimonious source trees; MRC = set of majority-rule consensus source trees; STRICT = set of set of strict consensus source trees; BOOTSTRAP = set of bootstrap source trees. The degree of resolution for each supertree estimate is listed in Table 2.

in Figure 6. Supertrees were estimated using the four bootstrapping approaches described above (bootstrap-weighted MRP, source-tree bootstrapping, hierarchical bootstrapping, and scaled hierarchical bootstrapping) and four conventional supertree estimation methods: MRP, MinCut, modified MinCut, and matrix representation with flipping (MRF).

Whenever possible, each supertree method was applied to three alternative sets of source-tree topologies: equally parsimonious trees resulting from exact parsimony searches of the component data sets were summarized as both majority-rule and strict consensus trees, and the profiles of trees generated by conventional bootstrap analyses of each component data set were summarized as majority-rule consensus trees.

The entire set of analyses (eight methods by three source-tree sets) was performed under each of four data partitioning schemes: the combined data set was parsed into one set of two components (the 2-partition set),

two sets of three components (3-partition sets A and B), and one set of four component matrices (the 4-partition set). Equally parsimonious supertrees derived from these analyses were summarized using strict consensus.

The results are summarized on a circular "dartboard" plot, which is divided into eight primary sectors (one for each supertree method), and some of these are further subdivided into three sub-sectors (one for each set of source trees). The target (simultaneous analysis) topology is located at the center of the plot. Each supertree estimate is plotted as a point within the appropriate (sub)sector, such that the radial distance from each point to the center of the graph corresponds to the topological distance of that supertree to the target tree. Topological distances were calculated using the Robinson-Foulds index (Robinson and Foulds, 1981) and were normalized such that the values reflect the number of NNI branch swaps between trees. Note that the distance between the various supertree

TABLE 2. Phylogenetic resolution¹ of supertree estimates under alternative methods.

Supertree method	2-Source-tree partition set ^A	3-Source-tree partition set A ^B	3-Source-tree partition set B ^C	4-Source-tree partition set ^D
MC	83	79	79	92
	79	79	79	92
	79	88	83	92
MC*	83	79	79	83
	79	83	79	92
	79	88	83	88
MRF _{NI}	4	4	4	4
	4	4	4	4
	4	4	4	4
MRF _{SPR}	79	75	79	4
	79	75	88	4
	79	75	79	79
MRF _{TBR}	4	25	8	4
	4	4	4	17
	4	13	38	71
MRP	79	75	79	4
	79	83	79	4
	79	83	79	4
MRP*	100	100	96	100
STBS _{MRC}	96	92	79	83
STBS _{EPT}	96	96	96	92
HBS	92	83	83	83
HBS*	92	92	92	100

¹Percent resolution was calculated as $k/(N-1)$, where k is the number of nodes in a tree of N species: this implicitly assumes that the underlying tree is strictly dichotomous (i.e., that all polytomies are “soft” *sensu* Maddison, 1989).

Abbreviations: MC = MinCut; MC* = modified MinCut; MRF_i = matrix representation with flipping (where the subscript indicates the branch-swapping algorithm); MRP = matrix representation with parsimony; MRP* = bootstrap weighted matrix representation with parsimony; STBS_{EPT} = source-tree bootstrap based on the set of equally parsimonious source trees, STBS_{MRC} = source-tree bootstrap based on the set of majority-rule consensus of equally parsimonious source trees, HBS = hierarchical bootstrap, HBS* = scaled hierarchical bootstrap.

The shading scheme corresponds to that used in Figure 6: Cells with dark shading pertain to estimates derived from the set of strict consensus source trees, medium shading indicates estimates derived from the set of majority-rule consensus source trees, and un-shaded cells indicate estimates derived from the set of bootstrap source trees.

^A2-Source-tree partition set = all DNA | morphology; ^B3-source-tree partition set A = morphology | atpB | DNA complement 1; ^C3-source-tree partition set B = morphology | rbcL | DNA complement 2; ^D4-source-tree partition set = morphology | atpB | rbcL | DNA complement 3.

estimates on these plots does not reflect their topological disparity.

Three points warrant comment. First, estimates obtained using the various bootstrapping approaches tended to outperform those derived with conventional supertree methods (Fig. 6). That is, they produced supertrees that more closely resembled the topology derived by simultaneous analysis. To some extent, this result is conservative, as the strict consensus of equally optimal supertree estimates derived from conventional methods tended to be highly unresolved (owing to conflict among the equally optimal supertrees; Table 2). Second, the relative performance of the conventional supertree methods in these experiments was somewhat surprising. For example, the poor performance of MRF in our empirical experiments contradicts previous simulation studies in which MRF consistently outperformed other supertree approaches (e.g., Chen et al., 2002, 2003; Eulenstein et al., 2004). This apparent anomaly may stem from our combination of strongly conflicting trees with identical terminal taxa. Finally, the hierarchical bootstrapping estimates tended to outperform the other bootstrap supertree approaches.

Bootstrap proportions.—We evaluated the ability of the alternative bootstrapping procedures to quantify phylo-

genetic uncertainty in supertrees by plotting the bootstrap proportions estimated by these methods against those obtained by simultaneous bootstrap analysis (Fig. 7). A total of four bootstrapping approaches were compared. Source-tree bootstrapping was applied in two ways: by drawing from the entire set of equally parsimonious trees for each component matrix (EPT-based source-tree bootstrapping) and by sampling from the set of majority-rule summaries of those trees (MRC-based source-tree bootstrapping). We also evaluated both scaled and unscaled variants of the hierarchical bootstrap. Comparisons were made for each of the four data partition sets.

Two generalities emerge from these empirical experiments. First, all four bootstrap approaches produced estimates of nodal support in the supertree that tended to be inflated relative to those based on simultaneous bootstrap analysis. Second, the four bootstrap supertree approaches are not equally biased but were consistently ranked as follows: MRC-based source-tree bootstrapping was the most inflated, followed by EPT-based source-tree bootstrapping, with scaled and unscaled hierarchical bootstrapping producing the least biased estimates of supertree nodal support values.

A more detailed look at *Dipsacaceae*.—Here, we focus on a subset of the empirical data—relationships and nodal support values estimated within the *Dipsacaceae* clade—to illustrate in more detail how the various bootstrapping approaches cope with this locus of phylogenetic conflict (Fig. 8). Results of the simultaneous analysis strongly support the topology (V(T(S(P,D)))) for *Valerianaceae*, *Triplostegia*, *Scabiosa*, *Pterocephalus*, and *Dipsacus*. The four partition sets each include one large partition comprising the bulk of the data (i.e., all DNA, and DNA compliments 1, 2, and 3) that strongly supports the topology found in the simultaneous analysis, and one to three smaller partitions (i.e., morphology, rbcL, and/or atpB) that partially conflict with the topology found in the simultaneous analysis. For instance, these smaller partitions all provide varying support for (D(S,P)) but conflict somewhat on the position of T: morphology places it with V, atpB places it with (D,S,P), and rbcL is equivocal, placing it in a polytomy with V and (D(S,P)) (Fig. 6).

Supertrees were estimated for each of the four sets of source-tree topologies using six estimation methods organized in five columns in Figure 8; from left to right: MRP, bootstrap-weighted MRP, MRC-based source-tree bootstrapping, EPT-based source-tree bootstrapping, hierarchical bootstrapping, and scaled hierarchical bootstrapping.

In general, the inferred supertree topology increasingly resembles the simultaneous analysis tree from left to right across the row. Supertree estimates under MRP and MRC-based source-tree bootstrapping tend to simply resolve conflict in favor of the most frequent source-tree topology, whereas bootstrap-weighted MRP tends to favor the resolution sanctioned by the largest sum of bootstrap proportions among the conflicting source-tree nodes. When no clear topological majority exists among the set of source trees (e.g., the 2-partition

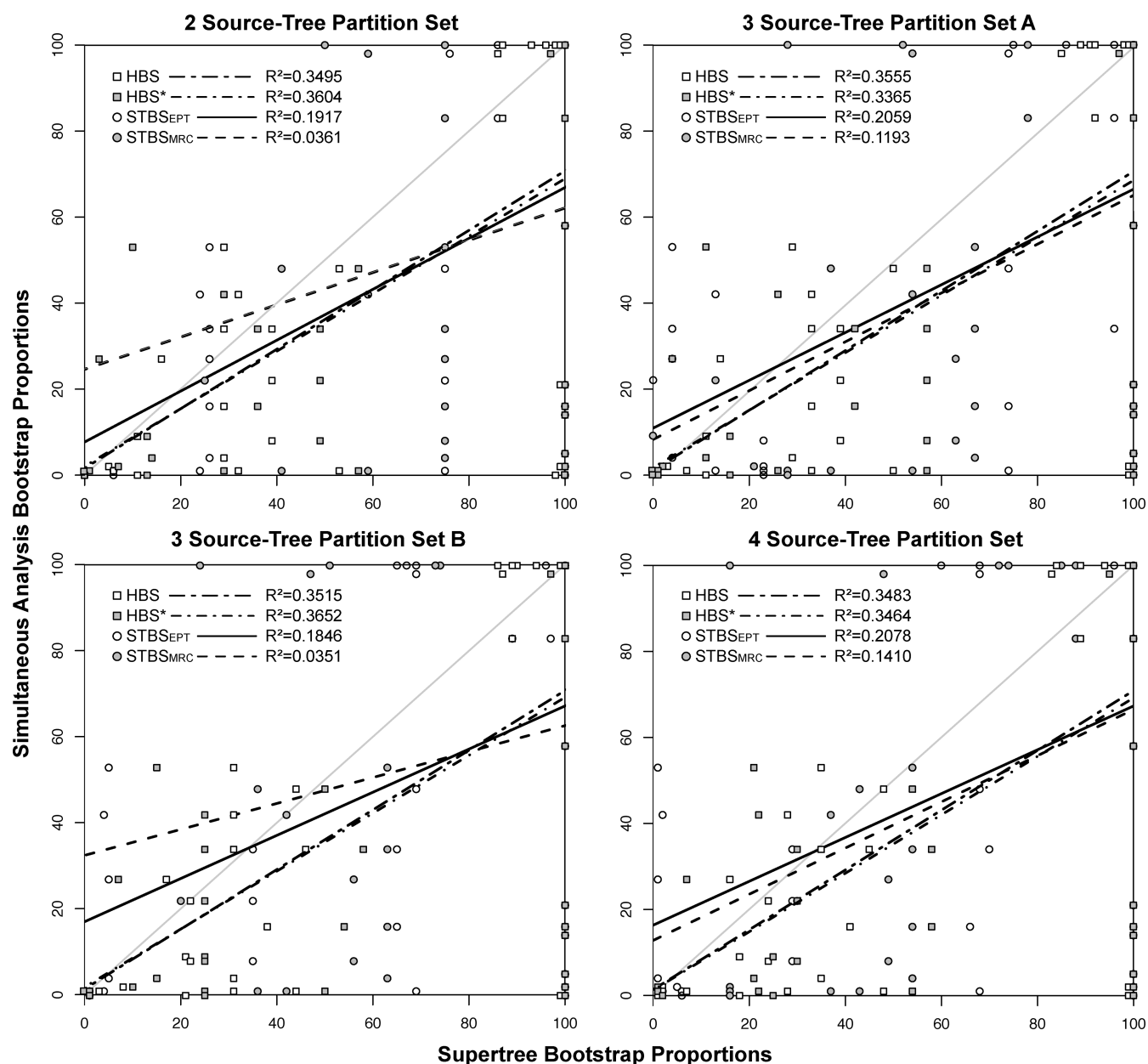


FIGURE 7. Correspondence between bootstrap proportions based on a simultaneous analysis and those estimated by the four bootstrapping approaches for each of the four data partition sets. A line was fitted through each set of points using linear regression. Unbiased estimates of supertree bootstrap proportions would fall along the diagonal. Abbreviations for the bootstrapping procedures: HBS = hierarchical bootstrap, HBS* = scaled hierarchical bootstrap; STBS_{EPT} = source-tree bootstrapping based on the set of equally parsimonious source trees; STBS_{MRC} = source-tree bootstrapping based on the set of majority-rule consensus of equally parsimonious source trees.

set in Fig. 8; see also Figs. 4 and 5), MRP tends to summarize the conflict as a polytomy, bootstrap-weighted MRP will resolve the conflict in favor of the more strongly supported source-tree nodes, and MRC-based source-tree bootstrapping stochastically chooses among the conflicting topologies. Overall, these three methods essentially behaved like conventional consensus techniques in these analyses. By contrast, EPT-based source-tree bootstrapping and both forms of hierarchical bootstrapping (scaled and unscaled) recovered the

simultaneous analysis topology over all data partition sets.

Although they performed similarly with respect to recovering the target topology, the latter four bootstrap approaches differed in their ability to approximate the bootstrap proportions estimated by simultaneous analysis. Again, the estimated supertree bootstrap proportions increasingly approximated those derived by simultaneous bootstrap analysis from left to right across the row, with scaled hierarchical bootstrapping performing best.

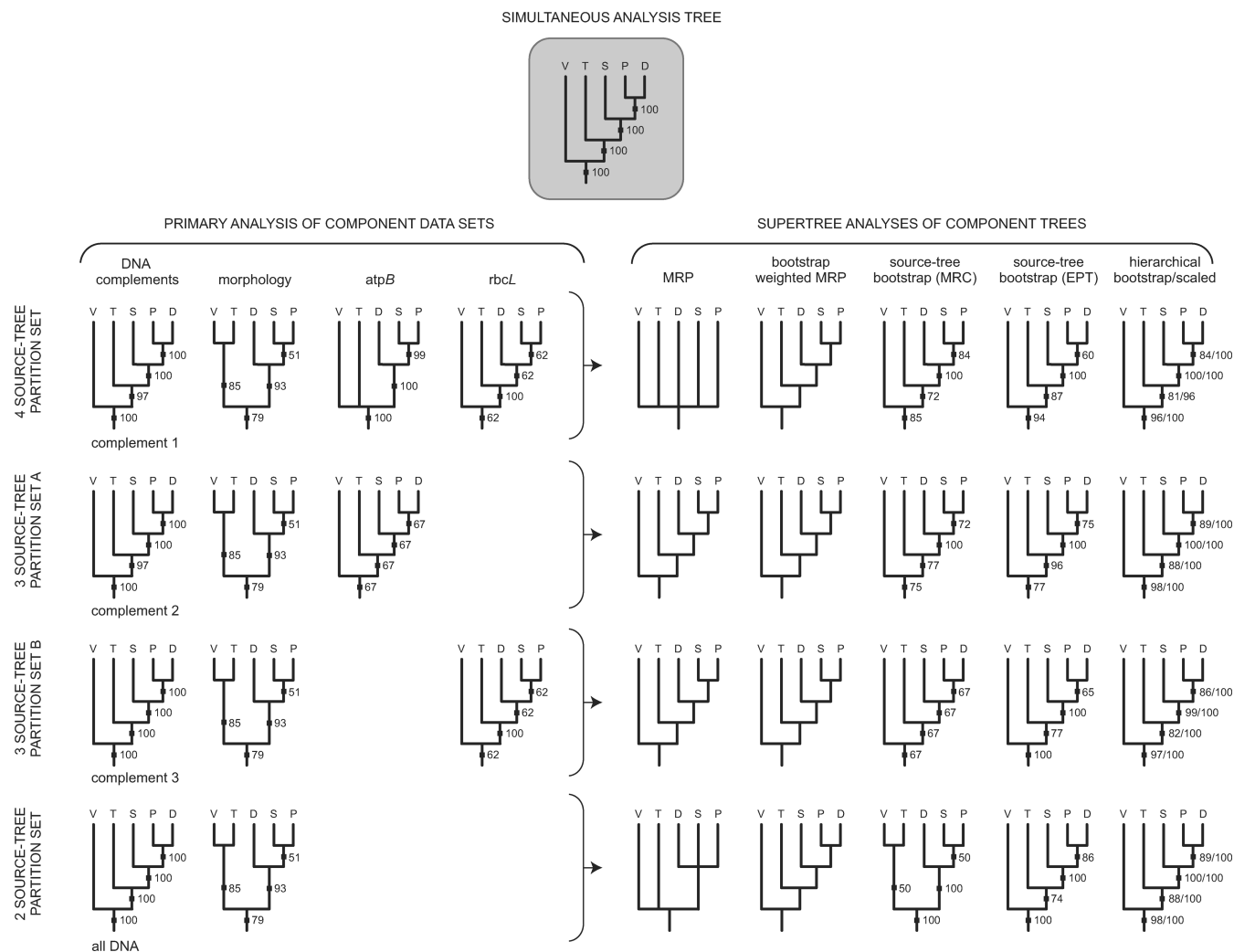


FIGURE 8. Dipsacaceae as a locus of phylogenetic conflict. The phylogeny estimated from a simultaneous analysis of the combined data is depicted at the top of the figure, and those based on conventional analyses of the various data partitions associated with the four partition schemes are arranged in rows on the left side of the figure. Each of the resulting four sets of source trees were variously combined using six supertree methods arranged in five columns in the adjacent rows to the right side of the figure. Abbreviations for taxon names: V, Valerianaceae; T, *Triplotelia*; D, *Dipsacus*; S, *Scabiosa*; P, *Pterocarpus*.

DISCUSSION

Nonparametric Bootstrapping, Supertree Topology, and Phylogenetic Uncertainty

Bootstrap-weighted MRP has the potential to improve the accuracy of supertree estimation by relaxing the implicit (and unrealistic) assumption that all source-tree nodes are equally supported by the underlying character data. However, this approach will only enhance the accuracy of supertree topologies to the extent that an equally unrealistic assumption holds: that bootstrap proportions are comparable across component data sets. As illustrated in the hypothetical and empirical examples, bootstrap support values are matrix specific (Hillis and Bull, 1993; Efron et al., 1996). Consequently, bootstrap-weighted MRP can in some cases decrease the correspondence between supertree topologies and those based on simultaneous analysis. In any event, bootstrap-weighted

MRP is less than ideal because it cannot quantify phylogenetic uncertainty in estimated supertree topologies.

Source-tree bootstrapping provides an effective means of assessing the level of conflict among observations relevant to conventional supertree estimation methods: the sample of source-tree topologies. Under certain conditions, this approach may increase the correspondence between supertree and simultaneous tree topologies and provide estimates of the phylogenetic uncertainty in supertrees. However, as a solution to the data-dissociation problem, source-tree bootstrapping is essentially a half measure. Because source-tree bootstrapping only reaches back to the topological summaries of the component matrices, much of the complexity in the underlying data will remain unavailable to the supertree inference, which is apt to confound estimates of supertree topology and nodal support. For example, source-tree bootstrapping is expected to perform

poorly under weak-signal enhancement scenarios (Barrett et al., 1991); i.e., when congruent sub-signals in two or more component matrices are manifest as homoplasy when the data sets are analyzed separately but emerge as the dominant signal when the component matrices are combined in a simultaneous analysis (Fig. 4). Furthermore, the reliance of this approach on source-tree topologies is also partially responsible for the observed bias in the estimated bootstrap proportions: ignoring much of the noise within the component data matrices causes bootstrap proportions to be correspondingly inflated (Figs. 7 and 8).

The failure of this approach to adequately address the data dissociation problem also explains the discrepancy between the results obtained with EPT-based versus MRC-based source-tree bootstrapping. Our experiments suggest that supertree topologies and nodal support values derived using source-tree bootstrapping more closely approximate those estimated by simultaneous analysis when it draws from the entire set of equally parsimonious source trees rather than from a set of consensus summaries of those equally parsimonious trees (i.e., EPT-based consistently outperformed MRC-based source-tree bootstrapping in our analyses). This is largely due to the relationship between the level of homoplasy within a data matrix and the number of equally parsimonious trees estimated from it. Analyses of homoplasy-rich data sets tend to produce a greater number of equally parsimonious trees than those derived from relatively signal-rich data sets. As in conventional bootstrapping, each tree in the bootstrap sample is reciprocally weighted by the number of equally parsimonious trees of which it is an instance, such that more noisy data sets exert a weaker influence on the estimated supertree. For example, combining the two source-tree partitions for *Dipsacaceae* involves source trees estimated from two partitions: all DNA and morphology, for which one and 77 equally parsimonious trees were found, respectively (Fig. 8 and Table 1). When source-tree bootstrapping is used to estimate a supertree from the two conflicting majority-rule consensus trees, the supertree is (by chance) identical to the morphological tree. By contrast, when source-tree bootstrapping is applied to the entire set of equally parsimonious trees for both partitions, the supertree overwhelmingly favors the DNA tree.

Accordingly, the advantage of EPT-over MRC-based source-tree bootstrapping stems from the fact that the number of equally parsimonious trees associated with the component matrices allows some of the complexity of the underlying source-tree data to seep through. Although EPT-based source-tree bootstrapping improves upon MRC-based source-tree bootstrapping (and other conventional supertree methods), it is surpassed by supertree methods that more accurately describe and directly incorporate complexity in the underlying character data (such as hierarchical bootstrapping, which relies on conventional bootstrapping to assay complexity in the data). Furthermore, practical considerations are apt to complicate the use of source-tree bootstrapping as a means of improving the accuracy of—and/or assessing

uncertainty in—supertree estimates. For example, application of source-tree bootstrapping to a set of source trees with moderate to low levels of taxonomic overlap could result in a bootstrap sample of trees with no overlapping taxa, in which case no meaningful supertree could be estimated. Second, it may not be possible to construct a majority rule consensus supertree when source-tree bootstrapping generates a profile of supertrees that differ in their taxon sets (O. R. P. Bininda-Emonds, personal communication). Clearly, these are potentially serious limitations of source-tree bootstrapping that cannot be addressed given our experimental design.

Although source-tree bootstrapping may prove to be of limited utility for improving the performance of—or assessing phylogenetic uncertainty in—supertree estimates, it may nevertheless find useful application to other problems. For example, because it effectively assesses conflict among a set of topologies, source-tree bootstrapping might be used to evaluate the level of conflict in the gene-tree/species-tree problem (e.g., Burleigh et al., 2006).

More so than the other approaches considered here, hierarchical bootstrapping is able to accommodate complexity in the underlying data. Conceptually, the hierarchical bootstrap is analogous to spectral decomposition: the primary bootstrap provides a prism that arrays complex patterns of character covariation in the component data matrices, allowing weaker signals to be reconstituted as spectra within the bootstrap profile, which provides an opportunity for a greater diversity of trees to contribute to the supertree estimate. The supertree algorithm (here MRP) provides a lens that focuses randomly sampled replicate sets of spectra along the bootstrap supertree profile. It is presumably because the hierarchical bootstrap reaches much closer to the primary character data that estimates of topology and nodal support derived under this approach corresponded most closely to those inferred by simultaneous analysis. As suggested by the consistently inflated bootstrap values, however, even the hierarchical bootstrap apparently fails to completely capture the complexity of the underlying character data.

Of course, it cannot be stressed too strongly that these insights were obtained from a limited set of empirical analyses performed under somewhat simplified conditions (complete taxonomic overlap among all source trees, etc.). Additional empirical and simulation studies will be necessary to assess the extent to which the findings obtained in the present study might be generalized to more complex supertree estimation scenarios.

Modularity, Portability, and Practicality

As mentioned previously, the bootstrapping approaches described here are not specifically tied to the MRP method, but instead are inherently modular procedures that could be interpolated with any supertree algorithm. Neither are the bootstrap procedures intrinsically reliant upon source-tree bootstrap profiles. For example, the hierarchical-bootstrapping and source-tree

bootstrapping approaches could be ported to a Bayesian framework simply by drawing from the set of source-tree posterior densities to estimate the posterior probabilities of clades in the supertree (this is an area of current research by J. Huelsenbeck and colleagues; personal communication to BRM). In light of apparent discrepancies between bootstrap proportions and posterior probabilities (e.g., Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Erixon et al., 2003; Yang and Rannala, 2005), it may be desirable to have both parametric and nonparametric estimates of uncertainty in supertrees.

Given that supertrees are valued for their relative computational efficiency, one possible reaction to the proposed bootstrap procedures is that they seem prohibitively expensive. We would address this concern as follows. First, the performance of the bootstrap procedures is likely to scale with the attendant computational burden, such that a greater computational investment is likely to pay dividends in the form of more accurate supertree estimates. Second, although MRP (like many parsimony-based phylogenetic inference algorithms) is NP-complete (Graham and Foulds, 1982), recent developments in supertree estimation have led both to heuristic parsimony-based algorithms that are fixed-parameter tractable (Chen et al., 2002) and to more efficient algorithms with polynomial running times (e.g., Semple and Steel, 2000; Page, 2002); use of these algorithms could greatly ameliorate the computational burden of the proposed bootstrapping approaches. Finally, the supertree bootstrapping approaches are inherently suited to distribution in a parallel processing environment.

The methods described in this paper have been implemented by SAS in the freely available program, tREeBOOT (Smith and Moore, in press). This command-line Java application (which will run on Macintosh OS X, Windows, Linux, and Unix operating systems) can perform bootstrap-weighting, source-tree bootstrapping, and hierarchical bootstrapping with MRP, MinCut, and modified MinCut supertree algorithms. The program enables source-tree bootstrapping and hierarchical bootstrapping approaches based both upon bootstrap profiles (as described herein) and also posterior probability densities of source trees to estimate posterior probabilities of nodes in the supertree (as mentioned above). The tREeBOOT distribution bundle may be obtained at <http://blackrim.org/programs.html>.

Implications and Applications

Although additional research is clearly needed to explore the generality of our findings, the various bootstrap approaches appear to have promise as a means for both improving the accuracy of supertree estimation and for assessing the uncertainty in supertree estimates. These preliminary results have practical and conceptual implications. First, the ability to gauge uncertainty in supertree estimates would greatly enhance their appeal in evolutionary studies—e.g., of character evolution, historical biogeography, coevolution, diversification rates—by al-

lowing the accommodation of phylogenetic uncertainty. For example, estimates of character evolution could be evaluated over the bootstrap profile of supertrees to estimate the confidence interval on the inference.

Second, although simulation studies have begun the important work of characterizing the behavior of a few supertree methods (e.g., Bininda-Emonds and Sanderson, 2001; Chen et al., 2002; Eulenstein et al., 2004), many methods remain poorly characterized (e.g., Wilkinson et al., 2001; Gatesy et al., 2002; Goloboff and Pol, 2002; Pisani and Wilkinson, 2002), an issue compounded by the frenetic pace at which new supertree methods are being proposed (e.g., Bininda-Emonds, 2004a). Furthermore, simulation studies of these methods constitute a nontrivial undertaking. Such studies provide meaningful results to the extent that conclusions are based on a thorough exploration of a simulated parameter space and that this space adequately approximates what the method will encounter in the analysis of real data. For this reason, simulation studies of the supertree inference problem are notoriously difficult owing to the high dimensionality of the relevant parameter space. However, if we are willing to accept the premise that the level of accuracy achieved by simultaneous analysis provides a reasonable benchmark against which the performance of supertree methods can be compared, then the experimental design adopted in the present study could provide an empirical counterpart to simulation studies for exploring the relative performance of available supertree methods.

As a final point, we note that one of the examples presented here (Fig. 1; based on Barrett et al., 1991) was previously used by Pisani and Wilkinson (2002) to demonstrate the correspondence of MRP estimates to those obtained by conventional consensus methods, prompting their conclusion that MRP should be classified as a “taxonomic congruence” method. Other authors have countered that MRP—by virtue of permitting the combination of otherwise incompatible data types—corresponds to a “total evidence” method (e.g., Bininda-Emonds and Bryant, 1998). As we have shown, however, MRP can behave either as a consensus method or—by interpolating various bootstrapping approaches—can approximate simultaneous analysis.

We believe that the conventional total evidence/taxonomic congruence dualism oversimplifies the association between character data and (super) tree topology, effectively rendering it as a binary relationship. Instead, the degree of correspondence between character data and tree topology appears to fall along a spectrum: this continuum is bracketed by direct consensus methods at one end and by simultaneous analysis of the primary data at the other, with conventional and bootstrap supertree approaches falling between these two extremes. Of the various approaches considered here, hierarchical bootstrapping appears to hold the most promise both to increase the correspondence of supertree topologies to the primary character data and to evaluate phylogenetic uncertainty of supertree estimates.

ACKNOWLEDGEMENTS

We offer sincere thanks to Olaf Bininda-Emonds, John Huelsenbeck, Rod Page, and Mike Sanderson for thoughtful discussion of the ideas presented in this paper, to the audience at the botany meeting held in Snowbird, Utah (August, 2004), for their input, and to Deborah Ciszek, Olaf Bininda-Emonds, François-Joseph Lapointe, Rod Page, and Mike Sanderson for providing valuable reviews. We are particularly indebted to Mary Moore for stimulating discussion, assisting with data analysis and figures, and for copyediting an earlier draft of this manuscript. Funding for this research was provided by a Natural Science and Engineering Research Council (NSERC) postgraduate scholarship to BRM.

REFERENCES

- Alfaro, M., S. Zoller, and F. Lutzoni. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol. Biol. Evol.* 20:255–266.
- Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. *Syst. Zool.* 40:486–493.
- 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.
- Bininda-Emonds, O. R. P. 2003. Novel versus unsupported clades: Assessing the qualitative support for clades in MRP supertrees. *Syst. Biol.* 52:839–848.
- Bininda-Emonds, O. R. P. 2004a. Phylogenetic supertrees: Combining information to reveal the Tree of Life. *Computational biology, volume 4*. Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Bininda-Emonds, O. R. P. 2004b. The evolution of supertrees. *Trend. Ecol. Evol.* 19:315–322.
- Bininda-Emonds, O. R. P., and H. N. Bryant. 1998. Properties of matrix representation with parsimony analysis. *Syst. Biol.* 47:497–508.
- Bininda-Emonds, O. R. P., J. L. Gittleman, and M. Steel. 2003. The (super) tree of life: Procedures, problems, and prospects. *Annu. Rev. Ecol. Syst.* 33:265–289.
- Bininda-Emonds, O. R. P., and M. J. Sanderson. 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Syst. Biol.* 50:565–579.
- Burleigh, J. G., A. Driskell, and M. J. Sanderson. 2006. Supertree bootstrapping methods for assessing phylogenetic variation among genes in genome scale data. *Syst. Biol.* 55:426–440.
- Burleigh, J. G., O. Eulenstein, D. Fernández-Baca, and M. J. Sanderson. 2004. MRF supertrees. in pages 65–85. *Phylogenetic supertrees: Combining information to reveal the Tree of Life* (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.
- Chen, D., L. Diao, O. Eulenstein, D. Fernández-Baca, and M. J. Sanderson. 2003. Flipping: A supertree construction method. *DI-MACS Ser. Discrete Math. Theoret. Comput. Sci.*, 61:135–160.
- Chen, D., O. Eulenstein, D. Fernández, and M. J. Sanderson. 2002. Supertrees by flipping. In *Proceedings of COCOON 2002*. Springer-Verlag LNCS.
- Chen, D., O. Eulenstein, and D. Fernández-Baca. 2004. Rainbow: A toolbox for phylogenetic supertree construction and analysis. *Bioinformatics* 20:2872–2873.
- Cotton, J. A., and R. D. M. Page. 2002. Going nuclear: Vertebrate phylogeny and gene family evolution reconciled. *Proc. R. Soc. Lond. B.* 269:1555–1561.
- Cotton, J. A., C. S. C. Slater, and M. Wilkinson. 2006. Discriminating supported and unsupported relationships in supertrees using triplets. *Syst. Biol.* 55:345–350.
- Creevey, C. J., and J. O. McInerney. 2005. CLANN: Investigating phylogenetic information through supertree analyses. *Bioinformatics* 21:390–392.
- Cummings, M. P., S. A. Handley, D. S. Myers, D. L. Reed, A. Rokas, and K. Winka. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* 52:477–487.
- Daubin, V., M. Gouy, and G. Perrière. 2001. Bacterial molecular phylogeny using supertree approach. *Genome Informatics* 12:155–164.
- Douady, C., F. Delsuc, Y. Boucher, W. Doolittle, and E. Douzery. 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Mol. Biol. Evol.* 20:248–254.
- Efron, B., E. Halloran, and S. Holmes. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Nat. Acad. Sci. USA* 93:13429–13434.
- Erixon, P., B. Sennblad, T. Britton, and B. Oxelman. 2003. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst. Biol.* 52:665–673.
- Eulenstein, O., D. Chen, J. G. Burleigh, D. Fernández-Baca, and M. J. Sanderson. 2004. Performance of flip-supertree construction with a heuristic algorithm. *Syst. Biol.* 53:1–10.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- Gatesy, J., C. Matthee, R. DeSalle, and C. Hayashi. 2002. Resolution of a supertree/supermatrix paradox. *Syst. Biol.* 51:652–664.
- Goloboff, P. A., and D. Pol. 2002. Semi-strict supertrees. *Cladistics* 18:514–525.
- Graham, R. L., and L. R. Foulds. 1982. The unlikelyhood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Math. Biosci.* 60–133–142.
- Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method of assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–192.
- Maddison, W. P. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5:365–377.
- Page, R. D. M. 2002. Modified MinCut supertrees. In *WABI 2002* (R. Guigó and D. Gusfield, eds). LNCS 2452:537–551.
- Page, R. D. M. 2004. Taxonomy, supertrees, and the Tree of Life pages 247–265. In *Phylogenetic supertrees: Combining information to reveal the Tree of Life*, (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.
- Penny, D., and M. D. Hendy. 1985. The use of tree comparison metrics. *Syst. Zool.* 34:75–82.
- Piaggio-Talice, R. 2003. Quartet suite v 1.0. <http://genome.cs.iastate.edu/supertree/downloads/download.html>.
- Pisani, D., and M. Wilkinson. 2002. Matrix representation with parsimony, taxonomic congruence, and total evidence. *Syst. Biol.* 51:151–155.
- Purvis, A. 1995. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251–255.
- Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1:53–58.
- Robinson, D. F., and L. R. Foulds. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.
- 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.
- Smith, S. A., and B. R. Moore. 2006. tREeBOOT: Assessing phylogenetic uncertainty in supertrees. *Bioinformatics* (in press).
- Swofford, D. L. 2002. *PAUP*: Phylogenetic analysis using parsimony (*and other methods)*, version 4.0. Sinauer Associate, Sunderland, Massachusetts.
- Thompson, J. D., J. Plewniak, F. Jeanmougin, F., and D. G. Higgins. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24:4876–4882.
- Wilkinson, M., J. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F.-J. Lapointe, C. Levasseur, J. O. McInerney, D. Pisani, and J. L. Thorley. 2005b. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Syst. Biol.* 54: 419–431.
- Wilkinson, M., D. Pisani, J. A. Cotton, and I. Corfe. 2005a. Measuring support and finding unsupported relationships in supertrees. *Syst. Biol.* 54:823–831.
- Wilkinson, M., J. L. Thorley, D. T. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree for the Platyhelminthes? Pages 292–301. In *Interrelationships of the Platyhelminthes* (D. T. Littlewood and R. A. Bray, eds.) Chapman-Hall, London.

First submitted 18 August 2005; reviews returned 7 October 2005;

final acceptance 19 July 2006

Associate Editor: Olaf Bininda-Emonds