

## Assessment of the Accuracy of Matrix Representation with Parsimony Analysis Supertree Construction

OLAF R. P. BININDA-EMONDS<sup>1</sup> AND MICHAEL J. SANDERSON

Section of Evolution and Ecology, University of California at Davis, Davis, California 95616, USA

**Abstract.**—Despite the growing popularity of supertree construction for combining phylogenetic information to produce more inclusive phylogenies, large-scale performance testing of this method has not been done. Through simulation, we tested the accuracy of the most widely used supertree method, matrix representation with parsimony analysis (MRP), with respect to a (maximum parsimony) total evidence solution and a known model tree. When source trees overlap completely, MRP provided a reasonable approximation of the total evidence tree; agreement was usually >85%. Performance improved slightly when using smaller, more numerous, or more congruent source trees, and especially when elements were weighted in proportion to the bootstrap frequencies of the nodes they represented on each source tree (“weighted MRP”). Although total evidence always estimated the model tree slightly better than nonweighted MRP methods, weighted MRP in turn usually outperformed total evidence slightly. When source studies were even moderately nonoverlapping (i.e., sharing only three-quarters of the taxa), the high proportion of missing data caused a loss in resolution that severely degraded the performance for all methods, including total evidence. In such cases, even combining more trees, which had positive effects elsewhere, did not improve accuracy. Instead, “seeding” the supertree or total evidence analyses with a single largely complete study improved performance substantially. This finding could be an important strategy for any studies that seek to combine phylogenetic information. Overall, our results suggest that MRP supertree construction provides a reasonable approximation of a total evidence solution and that weighted MRP should be used whenever possible. [Accuracy; matrix representation; missing data; MRP; phylogenetic supertrees; resolution; taxonomic congruence; total evidence.]

Supertree construction (sensu Sanderson et al., 1998) represents an increasingly popular technique for combining phylogenetic information. Large-scale supertrees already exist for all extant species of the mammalian orders Primates (Purvis, 1995a; Purvis and Webster, 1999) and Carnivora (Bininda-Emonds et al., 1999), for the major clades within the legume subfamily Papilionoideae (Wojciechowski et al., 2000), and for the family-level relationships of all extant mammals (Liu et al., 2001). Furthermore, supertree construction has been identified as the key to producing comprehensive phylogenies for problematic clades (e.g., the kinetoplastid protozoa Trypanosomatidae; Stothard, 2000).

The appeal of supertrees lies in their ability to synthesize many smaller, disparate sources of phylogenetic information into a single more-encompassing, but still well-resolved tree. This is especially true of one supertree method, matrix representation using parsimony (MRP; Baum, 1992; Ragan, 1992; also Brooks, 1981; Doyle, 1992).

In many cases, comprehensive phylogenetic estimates of an entire group cannot otherwise be obtained by conventional phylogenetic methods. For instance, primary analysis or total evidence (sensu Kluge, 1989) requires the combined data to be compatible, whereas taxonomic congruence (sensu Mickevich, 1978) requires that the studies possess the same set of taxa. Supertrees combine the positive aspects of both of the latter two approaches to avoid their individual shortcomings. Like taxonomic congruence, supertree construction utilizes tree topologies and thus allows phylogenetic estimates derived from all possible data sources (which are often incompatible) to be combined—usually retaining good resolution while doing so (Purvis, 1995b). Like total evidence, supertree construction can combine estimates with different sets of terminal taxa to obtain a solution that contains statements of phylogenetic relationship that are not present in any single source study. Overall, supertree construction seems to show great promise for phylogenetic inference and the ultimate goal of estimating the tree of life on the basis of using all available information.

<sup>1</sup>Current address and address for correspondence: Institute of Evolutionary and Ecological Sciences, Leiden University, Kaiserstraat 63, 9516, 2300 RA Leiden, The Netherlands; E-mail: bininda@rulsfb.leidenuniv.nl

Previous research into supertree construction has focused on methodological issues (e.g., different coding methods or optimization criteria to derive the supertree) and its theoretical validity in particular (Baum and Ragan, 1993; Rodrigo, 1993, 1996; Purvis, 1995b; Ronquist, 1996; Bininda-Emonds and Bryant, 1998; Wilkinson et al., 2001). Supertree construction has been strongly criticized (e.g., Rodrigo, 1993, 1996) because the method loses contact with the primary data and thereby essentially throws out some information. This is an admittedly undesirable, but otherwise necessary, feature of the method to be able to combine incompatible data types. What is important is to establish what is the consequence in practice of making this tradeoff inherent to supertree construction. Isolated examples show that MRP supertree construction will occasionally give "wrong" answers (e.g., Purvis, 1995b; Bininda-Emonds and Bryant, 1998; Wilkinson et al., 2001). However, large-scale performance testing of supertree construction has been limited to a meta-analysis based on literature sources, which indicated that MRP supertrees show reasonable similarity to a total evidence solution (Bininda-Emonds and Bryant, 1998), and a single case study comparing a supertree of papilionoid legumes to a phylogeny derived from the chloroplast *matK* gene (Wojciechowski et al., 2000).

Through simulation, we examine the accuracy of supertree construction in relation to a known model tree and an inferred total evidence solution. Variables examined include the size, number, and degree of incongruence between source trees; the effects of taxon sampling (i.e., the degree to which source trees overlap in their taxon sets); and some suggested modifications to the basic MRP coding procedure. We limit our analyses to MRP supertree construction because it represents the most widely used supertree technique to date—no doubt in part because of its widespread applicability.

## METHODS

Our basic protocol involved evolving nucleotide sequences down known model trees that were constructed according to a branching process (see below). The resulting matrix was subdivided into equal sized data partitions, each of which was analyzed individ-

ually to produce single source trees subsequently combined as an MRP supertree. Simultaneous parsimony analysis of all partitions yielded a total evidence solution. We then compared all possible pairs of trees (i.e., supertree, total evidence tree, and model tree) to assess the accuracy of MRP supertree construction. Each set of simulated model parameters was replicated 1,000 times.

### *Generating the Model Tree and Primary Data*

We constructed model trees, using the default parameters of the YULE\_C procedure in the program *r8s* (written by M.J.S.; available from <http://loco.ucdavis.edu/r8s/r8s.html>). This procedure obtains trees according to a stochastic Yule birth process conditioned on a fixed number of tips and a fixed time between the root of the tree and the present (Ross, 1996). Among other things, this model ensures that the age distribution of the nodes in the trees is invariant to taxon number. Rates of character evolution were modeled to depart from the assumption of a molecular clock. Specifically, branch durations from the Yule model were multiplied by branch-specific rates of evolution that are the product of an overall tree-wide substitution rate and a random normal variate (drawn from a distribution with a mean of 1.0, standard deviation of 0.5, and truncated outside of [0.1, 2.0]).

Nucleotide sequences were evolved on the model trees according to a standard Markov model as implemented in the program *Seq-Gen* v1.1 (Rambaut and Grassly, 1997). We used a relatively standard model of evolution, namely, a Kimura 2-parameter model with a transition/transversion ratio of 2.0, different average rates of substitution across the tree (see below), and site-to-site rate heterogeneity (with a shape parameter of 0.5) (Swofford et al., 1996). The size of the primary data matrix equaled 500 nucleotides multiplied by the number of source trees.

### *Derivation of MRP Supertrees and Total Evidence Solutions*

To obtain the MRP supertree, we divided the primary data matrix into equal-sized data partitions of 500 nucleotides according to the number of source trees that were to be combined. Because *Seq-Gen* evolves each site independently, we could select contiguous

blocks of nucleotides to obtain random partitions of the data.

The sequences for each data partition were then analyzed by using weighted parsimony (with transitions weighted twice as heavily as transversions in a step matrix approach) to derive a set of source trees. Use of weighted parsimony greatly improves performance over unweighted parsimony (Hillis et al., 1994) and so represents a best case scenario for our phylogenetic reconstructions, given that the weighting scheme exactly matched the model under which the data were generated. Analysis was done with PAUP\* v4.0b2 (Swofford, 1999), using a heuristic search strategy with a random addition sequence (25 replicates), TBR branch swapping, and 1,000 MAXTREES. Source trees were taken to be the strict consensus of all equally most-parsimonious solutions and were saved into a single tree file to be coded for the MRP analysis. Briefly, MRP encodes the topology of a tree as a series of binary elements, one for each (informative) node in the tree. Taxa that are derived from a given node are scored as 1, those that are not so derived but are still present on the tree are scored as 0, and all other taxa are scored as ?. The matrix representations for all source trees are then combined into a single matrix that is analyzed by parsimony. Given a tree file with descriptions of all the source trees, r8s can produce a NEXUS formatted matrix (see Maddison et al., 1997) for analysis in PAUP\* (as can PAUP\*). Determination of the supertree in PAUP\* used the same heuristic search strategy as described above for the source trees. The supertree was taken to be the strict consensus of all equally most-parsimonious solutions.

We obtained the total evidence solution through the combined analysis of all data partitions. Weighted parsimony analysis followed the search strategy described above for each individual partition (i.e., the strict consensus of all equally most-parsimonious solutions derived from a heuristic search).

#### *Comparing Tree Topologies*

To assess the accuracy of supertree construction, we checked the three pairwise combinations of the actual model tree and the supertree and total evidence estimates of it. The metrics we used compared tree topologies either in isolation (partition met-

ric and consensus fork index) or taking the underlying simulated sequence data into account (Kishino–Hasegawa test; Kishino and Hasegawa, 1989). Although the consensus fork index (CFI; Colless, 1980, 1981) is strictly speaking a “consensus object invariant” (Day and McMorris, 1985), it can also be adapted as a tree comparison metric (e.g., Swofford, 1991).

Both the partition metric ( $d_S$ ; Robinson and Foulds, 1981) and CFI are based on the number of common bipartitions between two trees. They differ in how they treat polytomies. The former views them as “hard” (see Maddison, 1989) so that the same polytomy on two trees is held to be “correct”. In contrast, our adaptation of the CFI essentially views polytomies as “soft”, thereby holding two identically placed polytomies as “incorrect”. This distinction is important, depending on the reference tree. In comparing the MRP supertree to the total evidence tree, we ideally want to reproduce any polytomies on the latter exactly. Thus,  $d_S$  is the appropriate choice for this comparison. In contrast, the model tree is always strictly bifurcating, and low values for  $d_S$  can occur only if either the supertree or total evidence tree is well resolved but disagrees strongly with the model tree. If either test tree is poorly resolved,  $d_S$  tends toward 50%. Thus, CFI is more informative when the model tree is the reference tree because it counts as incorrect both polytomies and clades of differing membership on the estimated trees. Values of  $d_S$  were standardized for tree size (by dividing through by  $2n - 6$ , where  $n$  = number of taxa; Steel and Penny, 1993) and subtracted from 1.0 to create a similarity metric equivalent to CFI.

Two trees that possess different topologies may still not be *significantly* different in the context of the underlying data. To examine this possibility, we used the Kishino–Hasegawa test to assess supertree accuracy relative to the simulated molecular sequences. We restricted our analyses to the nominal variable of whether or not a significant difference (i.e.,  $P \leq 0.05$ ) was detected for a given replicate. Because the topologies of both the supertree and the total evidence tree, and arguably the model tree, were selected a posteriori, the Kishino–Hasegawa test as implemented in PAUP\* is invalid in this instance (Swofford et al., 1996; Goldman et al., 2000). Specifically, Goldman et al. (2000) point out that the underlying

distribution of the test assumes that either test tree is equally likely to be more parsimonious than the other. Because the total evidence tree will usually be the most optimal given the data (taking into account the uncertainty of the heuristic search algorithm), this assumption is ordinarily violated whenever it forms one of the test trees. We correct for this by applying the Kishino–Hasegawa test as a one-tailed test (i.e.,  $P/2 \leq 0.05$ ; Goldman et al., 2000).

#### *Variables under Examination*

We examined several variables that potentially affect supertree performance: the coding method used, the size and number of source trees, the degree of incongruence among source trees, and the degree to which the source trees overlap (i.e., effect of nonidentical taxon sets among source trees). Multivariate ANOVAs were used to determine whether any of these factors, individually or in combination, significantly affected the accuracy of the supertrees in comparison with either reference tree (i.e., the total evidence or model trees).

Several modifications to the basic (“standard”) MRP procedure described above exist. In an attempt to minimize perceived redundancy in the basic coding method, Purvis (1995b) suggested that only taxa forming the direct sister group to the node under examination be scored as 0; all remaining taxa formerly scored as 0 should be scored as missing instead. In another proposed modification, Bininda–Emonds and Bryant (1998) argued from first principles that 0s in the MRP matrix should not be used as evidence to cluster taxa because they denote a lack of membership in a clade; instead, perhaps the MRP matrix should be analyzed with reversals (i.e.,  $1 \rightarrow 0$  transformations) prohibited. Finally, as has been widely suggested (Baum, 1992; Purvis, 1995b; Ronquist, 1996; Bininda–Emonds and Bryant, 1998), the individual elements in the MRP matrix should be weighted in proportion to the evidential support of the nodes they represent to retain greater contact with the raw data. We refer to these modifications as Purvis, irreversible, and weighted MRP, respectively.

r8s produces both standard and Purvis MRP codings. The latter were analyzed exactly in the same way as the former. For irreversible MRP, we specified that rever-

sals be prohibited (Camin–Sokal parsimony) during the parsimony analysis in PAUP\*. To obtain a weighted MRP supertree, we performed a fast heuristic bootstrap analysis (Felsenstein, 1985; see Swofford, 1999) of 1,000 replicates in PAUP\* for each data partition subsequent to the basic (weighted) parsimony analysis. The fully resolved majority rule bootstrap tree (Margush and McMorris, 1981) was saved to a tree file with bootstrap proportions saved as branch lengths. r8s can distill branch length information from a tree description to derive a MRP matrix with an associated character weighting set for analysis in PAUP\* in the manner described above for standard MRP. Thus, all four versions of MRP supertree were derived from the same set of simulated data.

Of the remaining variables, source tree number (that is, the number of data partitions of 500 bp each) was set to either 2 or 10, and source tree size was 8, 16, or 32 taxa per set of replicates. We mimicked increasing incongruence among source trees by increasing the rate of evolution for the simulated sequence data in Seq-Gen, thereby making it less likely that two partitions would give the same source tree. Average rates of evolution, representing the number of substitutions per site measured along a path from the root to any tip of the tree, were increased from the default of 0.1 (“no incongruence”) to 0.5 (“low”), 1.0 (“medium”), or 1.5 (“high”). These runs were performed on source trees with 16 taxa. Admittedly, this procedure may not model incongruence adequately or realistically. For instance, a key source of incongruence in practice is combining source trees representing different phylogenetic histories (e.g., gene trees vs. species trees), whereas our source trees are ultimately derived from a single model tree. However, there is no ideal way to model incongruence quantitatively, especially with different model trees. Minimally, our procedure yielded source trees that were increasingly different, and significantly so, as the rate of evolution was increased, as measured by both  $d_s$  and CFI (results not shown).

Finally, we deleted taxa from each partition to simulate the effect of nonidentical taxon sets among source trees. Except for the first (outgroup) taxon, which we retained to provide an unambiguous root for all trees, each taxon had a fixed probability of being deleted from each partition (25%, 50%, or 75%). Thus,

source trees varied in size according to a binomial probability distribution (with an expected mean of the total number of taxa multiplied by the inclusion probability), and the same taxon could be deleted from more than one partition. If a taxon was deemed to be absent from a given partition, it was either removed from the initial parsimony analysis (supertree) or its nucleotides were replaced with missing data for that partition (total evidence). Because parsimony analysis cannot be performed with fewer than four taxa, all deletion runs were performed on source trees with 32 taxa to minimize the likelihood of all but three or fewer taxa being deleted. If the latter condition still occurred, we recast the replicate. We did not verify whether the data partitions fulfilled the minimum requirement for supertree construction—that each source tree shares at least two taxa with the rest of the source trees combined—because this was unlikely to not occur and would affect both supertree construction and total evidence analyses similarly.

## RESULTS

For each factor (source tree size, incongruence between source trees, and degree of taxon overlap), we initially describe the trends for when two source trees are combined using standard MRP as a baseline for comparison. Thereafter, we describe the effect of combining more source trees or using different coding methods. Unless otherwise specified, all significant differences were highly significant (i.e.,  $P < 0.0001$ ).

### *Effect of Source Tree Size*

A multivariate ANOVA revealed that coding method, source tree size, and source tree number, both individually and in most combinations thereof, had a significant effect on supertree accuracy with respect to either the total evidence tree or the model tree. The only exceptions were the lack of a coding  $\times$  size interaction when compared to the total evidence tree by  $d_S$  ( $P = 0.6636$ ) and of a coding  $\times$  size  $\times$  number interaction when compared to the model tree by CFI ( $P = 0.8865$ ). Contingency tables also revealed highly significant differences attributable to each individual factor (interactions could not be analyzed) with respect to whether a Kishino-Hasegawa test indicated a

significant difference between the supertree and either reference tree.

In particular, supertree accuracy decreased significantly as the source trees became larger. In practical terms, however, most of the differences were small. Two source trees of eight taxa combined by using standard MRP showed 94.8% similarity with respect to the total evidence tree; this decreased to only 91.6% when the source trees were four-fold larger, at 32 taxa (Table 1). Equivalent values in comparison with the model tree were 87.1% and 79.9%, respectively (Table 2). A more obvious decrease in performance was detected by the Kishino-Hasegawa test. For eight taxa, 31.3% of the 1,000 simulation runs detected a significant difference between the supertree and the total evidence tree. This value jumped to 71.3% with 32 taxa (Table 3). Analogous numbers for comparisons with the model tree were slightly better: 30.6% and 57.4%, respectively (Table 4).

Combining 10 source trees significantly improved the accuracy of the supertree in all cases, although the same negative trend with increasing tree size remained. When 10 source trees of 32 taxa were combined using standard MRP, the supertree showed 95.2% similarity with the total evidence tree and 91.4% similarity with the model tree (Tables 1 and 2, respectively). Both values were better than their respective best case scenarios using only two source trees (i.e., of eight taxa). Improvement was not as marked with the Kishino-Hasegawa test: between 11% and 14% of the replicates indicated significant differences with either reference tree for eight taxa and 46% to 60% for 32 taxa (Tables 3 and 4).

Except for (bootstrap) weighted MRP, the behavior of the remaining coding methods was generally indistinguishable statistically from standard MRP. Weighted MRP, in contrast, performed significantly better than the nonweighted methods for nearly all cases. The only exceptions to this finding were when the total evidence tree was used as the reference tree, but only when  $d_S$  was used to measure accuracy (Table 1); the Kishino-Hasegawa test revealed a significantly improved performance by weighted MRP. Significant differences detected by the latter test always occurred in  $<30\%$  of the replicates, and usually  $<20\%$ , for weighted MRP (Table 3). The trend is more definite when comparing with the model tree: Weighted

TABLE 1. Difference in topology between various MRP supertrees and the total evidence solution of the same data set as measured by  $d_s$  (presented as a similarity metric). Values presented are means with standard deviations appearing below each mean in parentheses ( $n = 1,000$ ). For each coding method, 2 or 10 source trees were combined for each factor. Size values represent the number of taxa in the source tree. Incongruence values represent the average rate of evolution in number of substitutions per site for the simulated data generated with Seq-Gen. Deletion values represent the probability that a given taxon was missing from each source tree partition.

Factor	Value	Standard MRP		Purvis MRP		Irreversible MRP		Weighted MRP	
		2	10	2	10	2	10	2	10
Size	8	0.948	0.973	0.946	0.971	0.948	0.969	0.961	0.970
		(0.083)	(0.067)	(0.087)	(0.071)	(0.082)	(0.070)	(0.079)	(0.076)
	16	0.926	0.962	0.923	0.958	0.927	0.954	0.935	0.960
		(0.070)	(0.051)	(0.071)	(0.055)	(0.068)	(0.057)	(0.070)	(0.055)
	32	0.916	0.952	0.914	0.946	0.916	0.944	0.916	0.953
		(0.053)	(0.042)	(0.054)	(0.046)	(0.053)	(0.044)	(0.052)	(0.043)
Incongruence (16 taxa)	None (0.1)	0.926	0.962	0.923	0.958	0.927	0.954	0.935	0.960
		(0.070)	(0.051)	(0.071)	(0.055)	(0.068)	(0.057)	(0.070)	(0.055)
	Low (0.5)	0.900	0.944	0.896	0.937	0.902	0.933	0.915	0.942
		(0.076)	(0.068)	(0.082)	(0.076)	(0.073)	(0.076)	(0.086)	(0.074)
	Medium (1.0)	0.882	0.931	0.877	0.920	0.884	0.913	0.896	0.925
		(0.088)	(0.079)	(0.090)	(0.091)	(0.085)	(0.085)	(0.095)	(0.085)
High (1.5)	0.856	0.925	0.849	0.913	0.860	0.909	0.880	0.923	
	(0.097)	(0.080)	(0.101)	(0.091)	(0.095)	(0.089)	(0.100)	(0.084)	
Deletion (32 taxa)	None (0.1)	0.916	0.952	0.914	0.946	0.916	0.944	0.916	0.953
		(0.053)	(0.042)	(0.054)	(0.046)	(0.053)	(0.044)	(0.052)	(0.043)
	Low (0.25)	0.924	0.941	0.913	0.937	0.924	0.934	0.901	0.940
		(0.057)	(0.044)	(0.064)	(0.046)	(0.057)	(0.047)	(0.064)	(0.047)
	Medium (0.50)	0.934	0.911	0.924	0.901	0.925	0.905	0.889	0.912
		(0.071)	(0.059)	(0.074)	(0.063)	(0.083)	(0.063)	(0.089)	(0.065)
High (0.75)	0.982	0.927	0.980	0.923	0.952	0.741	0.950	0.912	
	(0.049)	(0.065)	(0.051)	(0.065)	(0.102)	(0.181)	(0.087)	(0.079)	

MRP always significantly outperformed the remaining MRP methods and was at least on a par with, but usually slightly outperformed, total evidence (Tables 2 and 4).

#### *Effect of Source Tree Incongruence*

Again, all the individual factors (coding method, incongruence, and source tree number) and most combinations thereof significantly influenced the values of all metrics for both reference trees. The only exceptions were for  $d_s$  when comparing with the total evidence tree for the two interactions coding  $\times$  homoplasy ( $P = 0.0983$ ) and coding  $\times$  size  $\times$  number ( $P = 0.6057$ ).

As indicated by combining two source trees (of 16 taxa) using standard MRP, supertree construction became less accurate with increasingly incongruent source trees. Again, the decrease, although significant, was generally small. The percent of clades shared between the supertree and total evidence tree decreased from 92.6% with "no" incongruence (rate = 0.1) to only 85.6% with "high" incongruence (rate = 1.5) (Table 1). When using the model tree as the reference tree, accuracy was lower and the de-

crease in performance with increasingly incongruent source trees was a little larger. The analogous values are 83.0% and 72.5%, respectively (Table 2). The Kishino-Hasegawa test again showed a large decrease (e.g., 57.6–87.7% of the replicates yielded significant differences compared with the total evidence solution), although the trend was not linear. The smaller value was obtained with no incongruence; with any substantial level of incongruence (i.e., rate  $\geq 0.5$ ), the values were all  $>80\%$  (Table 3). We observed a similar pattern when the model tree was used as the reference tree, albeit with a slightly improved performance (Table 4).

Although the same trends were apparent, combining 10 source trees did much to ameliorate the negative effects of increasing incongruence among them. Even when 10 "highly" incongruent source trees were combined by using standard MRP, the supertree showed the same resemblance to the total evidence solution (92.5%) as when only two congruent source trees were combined (Table 1). The situation was even better with respect to the model tree (86.4% vs. 83.0%, respectively; Table 2). A similar improvement was noted for the Kishino-Hasegawa test,

TABLE 2. Difference in topology between various trees inferred from a simulated DNA data set (MRP supertrees and the total evidence solution) and the model tree on which the data set is based as measured by the CFI. Values presented are means with standard deviations appearing below each mean in parentheses ( $n = 1000$ ). For each coding method, 2 or 10 source trees were combined for each factor (see Table 1).

Factor	Value	Standard MRP		Purvis MRP		Irreversible MRP		Weighted MRP		Total evidence	
		2	10	2	10	2	10	2	10	2	10
Size	8	0.871	0.948	0.870	0.946	0.876	0.949	0.927	0.959	0.913	0.958
		(0.137)	(0.092)	(0.139)	(0.093)	(0.131)	(0.090)	(0.104)	(0.081)	(0.115)	(0.081)
		0.830	0.926	0.824	0.922	0.831	0.927	0.902	0.947	0.883	0.941
Incongruence (16 taxa)	16	(0.103)	(0.073)	(0.108)	(0.076)	(0.102)	(0.073)	(0.086)	(0.062)	(0.092)	(0.065)
		0.799	0.914	0.792	0.907	0.800	0.910	0.875	0.939	0.852	0.931
		(0.078)	(0.057)	(0.082)	(0.061)	(0.077)	(0.056)	(0.064)	(0.047)	(0.070)	(0.052)
Deletion (32 taxa)	None (0.1)	0.830	0.926	0.824	0.922	0.831	0.927	0.902	0.947	0.883	0.941
		(0.103)	(0.073)	(0.108)	(0.076)	(0.102)	(0.073)	(0.086)	(0.062)	(0.092)	(0.065)
		0.786	0.896	0.786	0.887	0.793	0.895	0.879	0.921	0.861	0.910
Incongruence (16 taxa)	Low (0.5)	(0.111)	(0.086)	(0.114)	(0.092)	(0.107)	(0.088)	(0.090)	(0.075)	(0.100)	(0.065)
		0.756	0.871	0.756	0.857	0.765	0.868	0.859	0.900	0.844	0.890
		(0.114)	(0.097)	(0.120)	(0.105)	(0.113)	(0.098)	(0.098)	(0.088)	(0.104)	(0.093)
Deletion (32 taxa)	High (1.5)	0.724	0.864	0.724	0.851	0.739	0.861	0.844	0.893	0.819	0.882
		(0.121)	(0.095)	(0.127)	(0.108)	(0.120)	(0.099)	(0.105)	(0.089)	(0.111)	(0.094)
		0.799	0.914	0.792	0.907	0.800	0.910	0.875	0.939	0.852	0.931
Incongruence (16 taxa)	None (0.1)	(0.078)	(0.057)	(0.082)	(0.061)	(0.077)	(0.056)	(0.064)	(0.047)	(0.070)	(0.052)
		0.682	0.887	0.657	0.880	0.684	0.883	0.765	0.923	0.728	0.913
		(0.123)	(0.062)	(0.129)	(0.066)	(0.120)	(0.064)	(0.105)	(0.052)	(0.113)	(0.055)
Deletion (32 taxa)	Low (0.25)	0.444	0.717	0.419	0.696	0.454	0.713	0.523	0.791	0.482	0.768
		(0.173)	(0.112)	(0.171)	(0.118)	(0.167)	(0.113)	(0.167)	(0.096)	(0.166)	(0.098)
		0.227	0.145	0.221	0.135	0.250	0.246	0.254	0.193	0.236	0.180
Incongruence (16 taxa)	High (0.75)	(0.176)	(0.100)	(0.171)	(0.097)	(0.176)	(0.120)	(0.191)	(0.111)	(0.178)	(0.107)

TABLE 3. Difference in topology between various MRP supertrees and the total evidence solution of the same data set as measured by the Kishino-Hasegawa metric. Values presented are the proportion of 1,000 replicates in which a significant difference was detected ( $P \leq 0.05$ ) between the two trees in the context of the simulated DNA data set. For each coding method, 2 or 10 source trees were combined for each factor (see Table 1).

Factor	Value	Standard MRP		Purvis MRP		Irreversible MRP		Weighted MRP	
		2	10	2	10	2	10	2	10
Size	8	0.272	0.115	0.271	0.123	0.263	0.123	0.054	0.038
	16	0.486	0.323	0.505	0.327	0.490	0.320	0.110	0.084
	32	0.643	0.561	0.695	0.572	0.649	0.582	0.192	0.153
Incongruence (16 taxa)	None (0.1)	0.486	0.323	0.505	0.327	0.490	0.320	0.110	0.084
	Low (0.5)	0.811	0.353	0.778	0.371	0.800	0.363	0.146	0.090
	Medium (1.0)	0.842	0.393	0.830	0.435	0.842	0.452	0.139	0.112
	High (1.5)	0.868	0.379	0.856	0.433	0.864	0.408	0.141	0.114
Deletion (32 taxa)	None (0)	0.643	0.561	0.695	0.572	0.649	0.582	0.192	0.153
	Low (0.25)	0.686	0.611	0.756	0.659	0.680	0.630	0.435	0.169
	Medium (0.50)	0.586	0.744	0.645	0.816	0.598	0.765	0.577	0.335
	High (0.75)	0.171	0.739	0.190	0.773	0.275	0.886	0.272	0.669

with supertrees derived from 10 source trees being significantly different from either the total evidence or model tree ~50% and 30% less of the time, respectively, than when only two source trees were used (Tables 3 and 4).

The effect of using different coding methods was again complicated, although the general pattern remained that weighted MRP outperformed both the remaining MRP methods and total evidence. With two source trees, using weighted MRP significantly improved the estimate of the total evidence tree (as measured by  $d_s$ ) relative to the non-weighted MRP methods. With 10 source trees, however, standard and weighted MRP performed equally well, and slightly but significantly better than the other methods (Table 1). Weighted MRP always outperformed the other MRP methods, which were indistinguishable, by the Kishino-Hasegawa test, with significant differences from the total evidence tree occurring in <20% of the replicates (Table 3). The ability to reconstruct the model tree depended somewhat on the degree of incongruence among source trees. Weighted MRP was always slightly, but significantly, better than total evidence regardless of the metric, and both were always significantly better than the remaining MRP coding methods. As the level of incongruence increased, the nonweighted MRP methods tended to become significantly different from one another with no apparent pattern (Tables 2 and 4). Again, although these differences were often statistically significant, absolute differences were small, particularly when 10 source trees were combined.

#### *Effect of Nonidentical Taxon Sets*

Multivariate ANOVAs again revealed that all individual factors (coding method, degree of taxon overlap, and source tree number) and combinations thereof significantly affected values of both  $d_s$  and CFI when using either reference tree. Similarly, each individual factor had a highly significant influence on the nominal Kishino-Hasegawa variables (for either reference tree) as indicated by contingency tables.

The baseline for comparison of two source trees (of 32 taxa) combined by using standard MRP revealed opposing trends depending on the reference tree. When compared with the total evidence solution, accuracy increased slightly, but significantly, as the taxon overlap in the source trees decreased: from 91.6% similarity with complete taxon overlap to 98.2% similarity when a taxon had a 75% probability of being deleted from a given partition (Table 1). The Kishino-Hasegawa test mirrored this result, although the improvement in performance was not linear and increased markedly beyond a 50% taxon deletion probability (Table 3). In contrast, when compared with the model tree, a severe drop in performance occurred as taxon overlap between the source trees decreased. Unlike any trends for the previous factors, this decrease was both significant and large. With complete taxon overlap, the supertree and model tree showed 79.9% similarity. This dropped to 22.7% at the highest probability level of taxon deletion of 75% (Table 2). Moreover, the Kishino-Hasegawa test revealed that the inferred MRP supertree was usually significantly different



TABLE 4. Difference in topology between various trees inferred from a simulated DNA data set (MRP supertrees and the total evidence solution) and the model tree on which the data set is based as measured by the Kishino–Hasegawa metric. Values presented are the proportion of 1,000 replicates in which a significant difference was detected ( $P \leq 0.05$ ) between the two trees in the context of the simulated DNA data set. For each coding method, 2 or 10 source trees were combined for each factor (see Table 1).

Factor	Value	Standard MRP		Purvis MRP		Irreversible MRP		Weighted MRP		Total evidence	
		2	10	2	10	2	10	2	10	2	10
Size	8	0.250	0.118	0.250	0.122	0.230	0.119	0.015	0.028	0.056	0.063
	16	0.403	0.290	0.423	0.298	0.404	0.290	0.039	0.093	0.107	0.165
	32	0.493	0.401	0.545	0.444	0.504	0.429	0.088	0.142	0.191	0.263
Incongruence (16 taxa)	None (0.1)	0.403	0.290	0.423	0.298	0.404	0.290	0.039	0.093	0.107	0.165
	Low (0.5)	0.738	0.436	0.705	0.441	0.708	0.407	0.129	0.241	0.237	0.312
	Medium (1.0)	0.782	0.505	0.747	0.545	0.769	0.490	0.122	0.280	0.250	0.371
Deletion (32 taxa)	High (1.5)	0.835	0.540	0.809	0.540	0.817	0.482	0.159	0.315	0.290	0.396
	None (0)	0.493	0.401	0.545	0.444	0.504	0.429	0.088	0.142	0.191	0.263
	Low (0.25)	0.693	0.456	0.795	0.501	0.685	0.474	0.312	0.106	0.497	0.231
	Medium (0.50)	0.919	0.748	0.943	0.806	0.911	0.769	0.769	0.316	0.867	0.474
	High (0.75)	0.943	0.996	0.953	0.997	0.917	0.728	0.918	0.990	0.944	0.994

from the model tree with respect to the underlying simulated sequence data. Significant differences were found in 57.4% (complete overlap) to >90% of the replicates (deletion probability >50%) (Table 4).

The otherwise positive effects of increasing the number of source trees witnessed for other factors were manifested only at the higher levels of taxon overlap, regardless of the comparison metric or reference tree. Up to and including the 50% deletion probability level, combining more source trees improved the accuracy of the supertree estimate with respect to either reference tree. However, the increases in performance were not as substantial as for the other factors. For example, as measured by either  $d_S$  (total evidence solution) or CFI (model tree), it was always better to combine two completely overlapping source trees than to combine 10 source trees on which any given taxon had a 50% probability of being missing. Moreover, for highly nonoverlapping source trees (75% taxon deletion probability), using 10 source trees produced a more inaccurate estimate than using only two source trees, indicating a large decrease in performance for the former (Tables 1 and 2). Similar trends obtained from the Kishino–Hasegawa test: The beneficial effect of combining more source trees decreased as the extent of taxon overlap decreased. At high proportions of overlap (deletion probabilities of  $\leq 25\%$ ), significant differences with respect to either reference tree occurred less frequently than when two source trees were combined. However, at greater deletion probabilities, this pattern reversed itself such that combining two source trees yielded greater accuracy (Tables 3 and 4). In particular, supertrees derived from 10 highly nonoverlapping source trees were virtually always significantly different (99.7%; Table 4) from the model tree. Finally, as measured by both  $d_S$  and the Kishino–Hasegawa test (Tables 1 and 3, respectively), the estimate of the total evidence solution worsened slightly, but significantly, as taxon overlap between the 10 source trees decreased. This trend was the opposite of when two source trees were combined.

The alternative MRP methods displayed the same trends as standard MRP noted above: generally high accuracy with respect to the total evidence solution (Tables 1 and 3), but noticeably less accuracy with respect to the model tree (Tables 2 and 4).

We focus on comparisons with the model tree where all methods suffered severely (i.e.,  $d_S$  or CFI < 50%; Kishino–Hasegawa > 90%), namely, those beyond the 25% or 50% deletion probabilities for 2 or 10 source trees, respectively. This list included weighted MRP, which was otherwise more resistant to the negative effects of increasing source tree size and incongruence. However, whereas the nonweighted supertree methods were largely indistinguishable statistically, weighted MRP always performed slightly better, if still poorly, at severe taxon deletion levels. The improved performance of weighted MRP was generally more conspicuous at the lower taxon deletion levels for either the tree comparison metrics (Tables 1 and 2) or the Kishino–Hasegawa tests (Tables 3 and 4). Again, total evidence outperformed nonweighted MRP methods; however, it too was affected adversely by decreasing overlap between source trees. For instance, as measured by CFI, the similarity between the total evidence and model trees was only 18.0% (10 source trees) or 23.6% (2 source trees) when there was a 75% taxon deletion probability (Table 3). Moreover, total evidence was always slightly outperformed by weighted MRP at estimating the model tree. One exception to the above trends was the behavior of irreversible MRP when 10 source trees were combined with a 75% taxon deletion probability. Although this method was otherwise statistically indistinguishable from the remaining nonweighted MRP methods, it was significantly worse at estimating the total evidence solution (e.g.,  $d_S = 74.1\%$ ; Table 1) but significantly better at reconstructing the model tree (e.g., CFI = 24.6%; Table 3) in comparison with the remaining methods.

## DISCUSSION

MRP is necessarily an approximation of a total evidence solution, intended primarily for those situations when the latter technique cannot be applied (e.g., incompatible data types). At least in theory, MRP supertree construction might also be expected to be a reasonable approximation. Matrix representation is well grounded in basic graph and network theory, where there is an exact one-to-one correspondence between an individual tree and its matrix representation (Poincaré, 1901; Ponstein, 1966; Ragan, 1992). In such cases, parsimony represents

the “most efficient method” to recover the original tree (Baum and Ragan, 1993:638) because as a “perfect parsimony” problem, solutions can be found in polynomial time scaling according to the number of taxa (Gusfield, 1991, 1997). The justification is less clear when multiple source trees are combined and parsimony must necessarily be viewed as a heuristic (Baum and Ragan, 1993). However, there is reason for optimism. By recoding the nodes of a single source tree as a set of consistent elements (analogous to single homology statements), each individual matrix can be viewed as representing the primary signal from the original data stripped of any internal homoplasy (Bininda-Emonds et al., 1999). Correspondence with the original data can be improved by weighting the matrix elements according to some measure of evidential support derived from the raw data (Ronquist, 1996).

In the majority of cases, we have shown that MRP does provide a good approximation of a total evidence solution and, to a lesser degree, of the model tree. Moreover, this result holds regardless of the type of MRP method used. Despite theoretical arguments as to their respective validities (see Purvis, 1995b; Ronquist, 1996; Bininda-Emonds and Bryant, 1998), the different variants of MRP supertree construction were largely indistinguishable. Some evidence for this exists in the literature. Bininda-Emonds and Bryant (1998) found that standard and irreversible MRP gave largely similar answers in a small-scale meta-analysis of 19 literature sources and when applied to the carnivore supertree of Bininda-Emonds et al. (1999). Two versions of a supertree for all extant species of primate, one coded with Purvis MRP (Purvis, 1995a) and a revised version using standard MRP (Purvis and Webster, 1999), were found to be highly congruent (Purvis and Webster, 1999). In the current study, only weighted MRP was noticeably and statistically better than other MRP methods. In most situations, it yielded a tree topology that was absolutely closer to the total evidence or model trees and also was significantly different from those trees less often given the underlying data. In fact, weighted MRP usually outperformed a total evidence analysis of the raw data in reconstructing the model tree. This result may stem from the decreased amount of homoplasy in an MRP matrix. MRP matrices for individual source trees

are necessarily completely congruent; homoplasy can occur only between source trees. In contrast, incongruence can occur both within and between studies in a total evidence analysis. A weighted MRP analysis retains most of the important information from a total evidence analysis (i.e., differential support among inferred nodes) without a lot of the attendant noise. Together, these two properties may explain the improved performance of weighted MRP relative to both nonweighted methods (which lack information about differential support) and total evidence.

Therefore, we suggest that weighted MRP be used whenever equivalent metrics of evidential support can be obtained for all source trees (see Bininda-Emonds and Bryant, 1998). However, we recognize that the necessary information is usually unavailable. For this reason, none of the supertrees that currently exist in the literature (e.g., Purvis, 1995a; Bininda-Emonds et al., 1999; Wojciechowski et al., 2000) make use of weighted MRP. Fortunately, nonweighted MRP methods provide reasonable results under most circumstances (performing only marginally worse than total evidence), although the Kishino-Hasegawa test indicates that the difference in topology is often significantly different. We therefore argue that the published supertrees should be judged and used with about the same degree of confidence that a total evidence tree might be. Some caution might be prudent, however, particularly for the more weakly supported nodes.

Otherwise, MRP supertree construction behaved predictably, deviating slightly more from the total evidence and model trees as the source trees were fewer, larger, increasingly incongruent, or displayed decreasing amounts of overlap in their sets of terminal taxa. These represent more difficult reconstruction conditions where random noise is more likely to adversely affect the ability of MRP to approximate a total evidence approach. Except for when terminal taxon sets were not identical, however, congruence remained high, with >85% agreement between the supertree and total evidence solution.

The most important factor affecting supertree performance is, ironically, the most attractive feature of the method: the ability to combine trees with nonidentical taxon sets. At the most extreme deletion rate we examined (a 75% probability of a taxon being deleted from a given partition), supertrees

from all MRP methods showed <25% similarity to the model tree and were significantly different from the topology of the model tree >90% of the time. In virtually every replicate, the model tree was more parsimonious than the supertree with respect to the primary data (results not shown). Moreover, the otherwise positive effects of combining more source trees were completely negated, and proved to be detrimental, at the greatest probabilities of deletion. At best, reasonable accuracy was achieved only with source trees that were largely overlapping, although slightly less overlap ( $\leq 75\%$  between any pair; equals  $2 \times 50\%$  deletion probability) can be accommodated by combining more source trees.

Importantly, however, the fact that total evidence performed as poorly as MRP with nonidentical taxon sets suggests that the decrease in performance does not derive from any explicit shortcoming of MRP supertree construction. Instead, the cause may be the large proportion of missing data in such circumstances. An increase in the amount, and particularly the proportion, of missing data affects phylogenetic inference adversely by increasing the number of equally most-parsimonious solutions and decreasing the resolution of the consensus tree (Platnick et al., 1991; Wilkinson, 1995; Kitching et al., 1998; Wiens, 1998). The decrease in resolution is particularly important in the current context. All MRP methods and total evidence produced increasingly less-resolved solutions as the degree of overlap of the source trees decreased (Table 5). At the most extreme deletion probabilities, the solutions from either method were <30% resolved with respect to a fully bifurcating tree, and usually were <20%. In contrast, solutions derived from complete data sets were never <80% resolved.

The loss of resolution also explains the increasing similarity between MRP supertrees and the total evidence solution as two increasingly nonoverlapping source trees were combined. Because  $d_s$  treats polytomies as hard, two highly unresolved trees will be indicated to be very similar. Thus, the increasing similarity in this one case is the result of the loss of resolution being inferred as increasing similarity, thereby possibly obscuring actual topological differences. Instead, the CFI is more instructive here and displays

the expected trends. With two source trees, accuracy decreases linearly from 82.6% to 26.5% as the trees move from being completely overlapping to highly nonoverlapping. The respective numbers for 10 source trees are 92.8% and 15.6%. Again, weighted MRP slightly, but significantly, outperforms nonweighted MRP methods.

Fortunately, the problems inherent in combining highly nonoverlapping source studies can be largely ameliorated by including one reasonably complete study in the analysis. When we constrained the relevant simulations such that taxa from the first partition had a deletion probability of 10%, accuracy with respect to the model tree improved dramatically, even when the remaining partitions were highly nonoverlapping (i.e., deletion probability of 75%). For example, when 10 source trees were combined in this manner, the standard MRP supertree was 95.8% resolved, showed 91.0% similarity to the model tree, and was significantly different from the model tree in only 50.0% of the replicates (results not shown). The two former values are not different statistically from the case where there was no deletion from any of the partitions, and the last value is about on a par with that case. Similar improvements occurred for the remaining MRP methods and total evidence, and when only two source trees were used. However, in the latter case, results were slightly but significantly different compared with when no taxon deletion occurred. Instead, performance was equivalent to when both partitions displayed a 10% taxon deletion probability. Therefore, a good general strategy when combining phylogenetic information may be to "seed" any supertree or total evidence analysis with a study that includes most of the desired taxa, even if this study is not that well resolved (e.g., a taxonomy). Similar results could also be achieved by imposing topological constraints identical to the seed tree. This procedure could also decrease search times by eliminating part of the tree search space. However, the procedure does not allow the seed tree topology to be contradicted, even if that topology conflicts with the majority of the source trees. In both cases, the accuracy of the supertree depends critically on the accuracy and amount of resolution in the seed tree. This is true even when the seed tree is included as a source tree in the MRP analysis, given that MRP tends to favor larger



source trees (Ronquist, 1996; but see Bininda-Emonds and Bryant, 1998).

In conclusion, MRP supertree construction appears to provide a reasonable approximation of a total evidence solution and performs comparably to the latter when estimating the model tree. Performance is particularly improved when the MRP matrix is weighted to reflect evidential support in the original studies. This overall conclusion is encouraging, given that many of the phylogenetic data accumulated to date are incompatible and thus cannot be analyzed within a conventional total evidence framework. We believe supertrees provide a tractable, accurate procedure to combine disparate data sources in the principle of total evidence to derive phylogenetic hypotheses based on the widest assortment of independent evidence possible.

#### ACKNOWLEDGMENTS

We thank Nick Isaac for the suggestion to examine how MRP performs when combining many smaller source trees and one large one. Martyn Kennedy, Rod Page, and Andy Purvis also provided valuable suggestions that improved the paper. O.R.P.B.-E. was supported by an NSERC postdoctoral fellowship.

#### REFERENCES

- 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.
- BAUM, B. R., AND M. A. RAGAN. 1993. Reply to A.G. Rodrigo's "A comment on Baum's method for combining phylogenetic trees." *Taxon* 42:637–640.
- BININDA-EMONDS, O. R. P., AND H. N. BRYANT. 1998. Properties of matrix representation with parsimony analyses. *Syst. Biol.* 47:497–508.
- BININDA-EMONDS, O. R. P., J. L. GITTLEMAN, AND A. PURVIS. 1999. Building large trees by combining phylogenetic information: A complete phylogeny of the extant Carnivora (Mammalia). *Biol. Rev.* 74:143–175.
- BROOKS, D. R. 1981. Hennig's parasitological method: A proposed solution. *Syst. Zool.* 30:229–249.
- COLLESS, D. H. 1980. Congruence between morphometric and allozyme data for *Menidia* species: A reappraisal. *Syst. Zool.* 29:288–299.
- COLLESS, D. H. 1981. Predictivity and stability in classifications: Some comments on recent studies. *Syst. Zool.* 30:325–331.
- DAY, W. H. E., AND F. R. MCMORRIS. 1985. A formalization of consensus index measures. *Bull. Math. Biol.* 47:215–229.
- DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* 17:144–163.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49:652–670.
- GUSFIELD, D. 1991. Efficient algorithms for inferring evolutionary trees. *Networks* 21:19–28.
- GUSFIELD, D. 1997. Algorithms on strings, trees, and sequences: Computer science and computational biology. Cambridge Univ. Press, Cambridge.
- HILLIS, D. M., J. P. HUELSENBECK, AND C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. *Science* 264:671–677.
- KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* 29:170–179.
- KITCHING, I. J., P. L. FOREY, C. J. HUMPHRIES, AND D. M. WILLIAMS. 1998. *Cladistics: The theory and practice of parsimony analysis*, 2nd edition. Oxford Univ. Press, Oxford.
- KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25.
- LIU, F.-G. R., M. M. MIYAMOTO, N. P. FREIRE, P. Q. ONG, M. R. TENNANT, T. S. YOUNG, AND K. F. GUGEL. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science* 291:1786–1789.
- MADDISON, D. R., D. L. SWOFFORD, AND W. P. MADDISON. 1997. NEXUS: An extensible file format for systematic information. *Syst. Biol.* 46:590–621.
- MADDISON, W. P. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5:365–377.
- MARGUSH, T., AND F. R. MCMORRIS. 1981. Consensus *n*-trees. *Bull. Math. Biol.* 43:239–244.
- MICKEVICH, M. F. 1978. Taxonomic congruence. *Syst. Zool.* 27:143–158.
- PLATNICK, N. I., C. E. GRISWOLD, AND J. A. CODDINGTON. 1991. On missing entries in cladistic analysis. *Cladistics* 7:337–343.
- POINCARÉ, H. 1901. Second complément à l'analyse situs. *Proc. Lond. Math. Soc.* 32:277–308.
- PONSTEIN, J. 1966. Matrices in graph and network theory. Van Gorcum, Assen, The Netherlands.
- PURVIS, A. 1995a. A composite estimate of primate phylogeny. *Philos. Trans. R. Soc. London. B* 348:405–421.
- PURVIS, A. 1995b. A modification to Baum and Ragan's method for combining phylogenetic trees. *Syst. Biol.* 44:251–255.
- PURVIS, A., AND A. J. WEBSTER. 1999. Phylogenetically independent comparisons and primate phylogeny. Pages 44–70 in *Comparative primate socioecology* (P. C. Lee, ed.). Cambridge Univ. Press, Cambridge.
- RAGAN, M. A. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1:53–58.
- RAMBAUT, A., AND N. C. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- ROBINSON, D. F., AND L. R. FOULDS. 1981. Comparison of phylogenetic trees. *Math. Biosci.* 53:131–147.
- RODRIGO, A. G. 1993. A comment on Baum's method for combining phylogenetic trees. *Taxon* 42:631–636.
- RODRIGO, A. G. 1996. On combining cladograms. *Taxon* 45:267–274.
- RONQUIST, F. 1996. Matrix representation of trees, redundancy, and weighting. *Syst. Biol.* 45:247–253.

- ROSS, S. M. 1996. Stochastic processes, 2nd edition. Wiley, New York.
- SANDERSON, M. J., A. PURVIS, AND C. HENZE. 1998. Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol. Evol.* 13:105–109.
- STEEL, M. A., AND D. PENNY. 1993. Distributions of tree comparison metrics—some new results. *Syst. Biol.* 42:126–141.
- STOTHARD, J. R. 2000. Future trypanosomatid phylogenies: Refined homologies, supertrees and networks. *Mem. Inst. Oswaldo Cruz* 95:523–526.
- SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, Oxford.
- SWOFFORD, D. L. 1999. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.
- WIENS, J. J. 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? *Syst. Biol.* 47:625–640.
- WILKINSON, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Syst. Biol.* 44:501–514.
- WILKINSON, M., J. L. THORLEY, D. T. J. LITTLEWOOD, AND R. A. BRAY. 2001. Towards a phylogenetic supertree of Platyhelminthes? Pages 292–301 *in* Interrelationships of the Platyhelminthes (D. T. J. Littlewood and R. A. Bray, eds.). Taylor and Francis, London.
- WOJCIECHOWSKI, M. F., M. J. SANDERSON, K. P. STEELE, AND A. LISTON. 2000. Molecular phylogeny of the “temperate herbaceous tribes” of papilionoid legumes: A supertree approach. Pages 277–298 *in* Advances in Legume Systematics (P. Herendeen and A. Bruneau, eds.). Royal Botanic Gardens, Kew, U.K. (WWW Internet address for full tree: <http://loco.ucdavis.edu/marty/HGSupertree.html>)

Received 23 August 2000; accepted 14 November 2000

Associate Editor: R. Page