Syst. Biol. 60(2):218–225, 2011

DOI:10.1093/sysbio/syq086
Advance Access publication on January 20, 2011

Consequences of Different Null Models on the Tree Shape Bias of Supertree Methods

ANNE KUPCZOK*

Center for Integrative Bioinformatics Vienna, Max F. Perutz Laboratories, University of Vienna, Medical University of Vienna, University of Veterinary Medicine Vienna, Dr. Bohr-Gasse 9, A-1030 Vienna, Austria;
**Correspondence to be sent to: Institute of Science and Technology, Austria Am Campus 1, A-3400 Klosterneuburg, Austria. E-mail: anne.kupczok@ist.ac.at.*

Received 30 September 2010; reviews returned 2 February 2010; accepted 15 September 2010
Associate Editor: Michael Charleston

Supertree methods are widely applied and give rise to new conclusions about phylogenies (e.g., Bininda-Emonds et al. 2007). Although several desiderata for supertree methods exist (Wilkinson, Thorley, et al. 2004), only few of them have been studied in greater detail, examples include shape bias (Wilkinson et al. 2005) or pareto properties (Wilkinson et al. 2007). Here I look more closely at two matrix representation methods, matrix representation with compatibility (MRC) and matrix representation with parsimony (MRP). Different null models of random data are studied and the resulting tree shapes are investigated. Thereby I consider unrooted trees and a bias in tree shape is determined by a tree balance measure. The measure for unrooted trees is a modification of a tree balance measure for rooted trees. I observe that depending on the underlying null model of random data, the methods may resolve conflict in favor of more balanced tree shapes. The analyses refer only to trees with the same taxon set, also known as the consensus setting (e.g., Wilkinson et al. 2007), but I will be able to draw conclusions on how to deal with missing data.

BACKGROUND

In this section, the relevant terms, in particular tree topologies and tree shapes, are introduced. A *(tree) topology* for n taxa is an unrooted leaf-labeled bifurcating tree with n leaves. The leaf labels are called taxa. A *split* $A|B$ is a bipartition of the taxon set into the two sets A and B. It corresponds to a branch in a tree where the taxa in A are on one side of the branch and the taxa in B are on the other side. A split is called k-split if $k = min\{|A|, |B|\}$. A bifurcating topology for n taxa has n 1-splits and $n-3$ k-splits with $k \geqslant 2$. From now on, only splits with $k \geqslant 2$

are considered, and they are called *inner splits* or simply splits. Overall, there exist $2^{n-1} - n - 1$ different inner splits for n taxa. Only compatible splits can be arranged in a tree, and two splits $A_1|B_1$ and $A_2|B_2$ are compatible if and only if at least one of the following sets is empty: $A_1 \cap A_2$, $A_1 \cap B_2$, $B_1 \cap A_2$ or $B_1 \cap B_2$. Due to this restriction, the number of possible trees in which a k-split can occur depends on k (Table 1).

A *(tree) shape* is obtained from a topology by ignoring the labels. Thus, a shape is an unlabeled bifurcating tree. Note that the term k-split can also be used in the context of a tree shape, for example, the shape S_5 (Table 2) has two 2-splits.

From six taxa on, more than one tree shape exists (Table 2). These tree shapes differ in their *balance*, which can be measured with tree balance statistics (e.g., Shao and Sokal 1990). Because these statistics are defined for rooted trees, the shapes here are rooted at their *centroid*. A centroid of a shape is a node where none of its incident edges leads to more than half of the terminal edges and if this holds for two nodes, the root is placed on the edge connecting these two nodes (see also Furnas 1984). If the root is placed on an edge, then a bifurcating rooted tree is obtained. But if the root is placed on a node, then there is a multifurcation at the root. In this case, I compute the mean over the tree balance statistics for all three possible resolutions. As tree balance statistic, the *Colless index* without any normalization is used (Colless 1982). It sums up the absolute differences in the number of leaves between the right and left subgraph for each node. Thus, higher Colless indices resemble higher imbalance. In table 2, the shape with the highest Colless index is also the most unbalanced one. However, the Colless index as applied here does not necessarily coincide with the intuitive understanding of the balance of an unrooted shape. First, for $n=6$, the Colless index cannot distinguish between the two shapes (Table 2), but in analogy to other n, I also call $S_{6,1}$ unbalanced. Second, for example, $S_{8,2}$ "looks" more balanced than $S_{8,3}$ although it has a higher Colless index. However, when comparing more complex shapes, a computable index is necessary to assess balance, for example, one cannot intuitively decide whether $S_{9,4}$ or $S_{9,6}$ is more balanced.

There is a relationship between the kinds of splits in the shape and their balance: Unbalanced shapes are the ones with exactly two 2-splits (shapes $S_{6,1}$, $S_{7,1}$, $S_{8,1}$ and $S_{9,1}$). On the other hand, the balanced shapes maximize the number of 2-splits.

I study two null models containing no phylogenetic information. The first is the well-known distribution that each bifurcating tree for a particular number of taxa is equally likely (proportional to distinguishable arrangements, *PDA*; Rosen 1978). See Table 2 (lines "total") for the possible numbers of trees for $n=5,\ldots,9$. *Perfect PDA* denotes the data set which contains each tree exactly once. Second, I introduce the model that each split is equally likely (proportional to distinguishable splits, *PDS*). Analogously, *perfect PDS* denotes the data set that contains each possible split exactly once. It is possible to relate each split with a multifurcating tree with only one inner branch and the PDS corresponds to an equal distribution of those multifurcating trees. Note that there is no distribution of bifurcating trees that corresponds to the PDS model (Steel and Pickett 2006). In particular, the PDS and PDA models are distinct because some splits occur in more trees than others (Table 1). For a given k-split, the number of trees containing this split is $\mathcal{T}_{k+1} \times \mathcal{T}_{n-k+1}$ where $\mathcal{T}_i$ is the number of unrooted trees with i taxa ($\mathcal{T}_i = (2i-5)!! = 1 \times 3 \times \ldots \times (2i-5)$). For example, for $n=6$, there are 15 different 2-splits and each one is present in 15 different bifurcating trees, but there are 10 different 3-splits, and each one is present in only nine trees. And for $n=9$, the number of trees containing a particular 2-split is more than six times higher than the number of trees containing a particular 4-split. Thus, this gap increases with n. As a consequence, two trees generated randomly under the PDA model tend to have no k-split with $k \geqslant 3$ in common, if the number of taxa goes toward infinity (Steel and Penny 1993). Note that the PDA model also implies a nonuniform shape distribution. In particular, unbalanced shapes are most likely (Table 2).

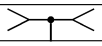
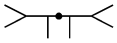
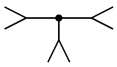
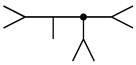
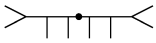
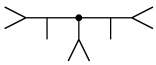
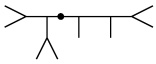
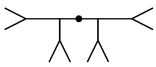
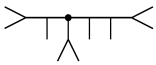
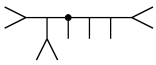
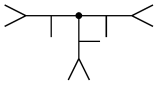
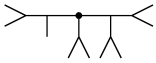
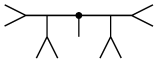
TABLE 1. Number of different splits for each k and the number of different trees containing a particular k-split

n	k	Number of splits	Number of trees per split
5	2	10	3
6	2	15	15
	3	10	9
	Total	25	
7	2	21	105
	3	35	45
	Total	56	
8	2	28	945
	3	56	315
	4	35	225
	Total	119	
9	2	36	10,395
	3	84	2835
	4	126	1575
	Total	246	

METHODS

I study the two related supertree methods MRC (Rodrigo 1996; Ross and Rodrigo 2004) and MRP (Baum 1992; Ragan 1992; Baum and Ragan 2004). Both code the input trees first as a binary matrix, the *matrix representation* (MR), and subsequently compute a supertree by optimizing an objective function on this matrix. In the MR, each column is a binary coding of an input tree split $A|B$, with `1`s corresponding to the taxa in A and `0`s corresponding to the taxa in B. Because only unrooted trees are considered, the direction, that is, whether A or B is coded `0`, is not important. The number of columns in an MR is denoted as length l. This is the total number of splits in the input trees. MRC and MRP are split-based methods because they first extract the split information of the input trees and subsequently compute a supertree from this split information.

TABLE 2. Shape probabilities and critical numbers under the uniform tree model (PDA)

	n	Shape	Number of trees	Probability	Colless index	Critical numbers (i_c)			
						MRC (%)		MRP (%)	
5	S_5		15	1	2.3	1	(6.7)	1	(6.7)
6	$S_{6,1}$		90	0.857	2	6	(5.7)	6	(5.7)
	$S_{6,2}$		**15**	**0.143**	**2**	**1**	**(0.95)**	**1**	**(0.95)**
	Total		105						
7	$S_{7,1}$		630	0.667	5.7	60	(6.3)	60	(6.3)
	$S_{7,2}$		**315**	**0.333**	**4**	**1**	**(0.11)**	**1**	**(0.11)**
	Total		945						
8	$S_{8,1}$		5040	0.485	6	720	(6.9)	900	(8.7)
	$S_{8,2}$		2520	0.242	5.3	270	(2.6)	540	(5.2)
	$S_{8,3}$		2520	0.242	3	630	(6.1)	990	(9.5)
	$S_{8,4}$		**315**	**0.030**	**0**	**1**	**(0.01)**	**1**	**(0.01)**
	Total		10395						
9	$S_{9,1}$		45360	0.336	11	8820	(6.5)	12600	(9.3)
	$S_{9,2}$		45360	0.336	8.3	7560	(5.6)	13860	(10.3)
	$S_{9,3}$		22680	0.168	8	8820	(6.5)	13860	(10.3)
	$S_{9,4}$		7560	0.056	6	3150	(2.3)	7560	(5.6)
	$S_{9,5}$		**11340**	**0.084**	**5.3**	**1**	**(0.0007)**	**1**	**(0.0007)**
	$S_{9,6}$		2835	0.021	5	1260	(0.93)	1260	(0.93)
	Total		135135						

Note: Bold shapes are the optimal shapes with MRC and MRP under the PDA model. The dot marks the centroid of each shape. For each n, the shapes are ordered by decreasing Colless index, that is, by increasing balance. The critical number is the number of trees necessary to be added to the perfect PDA model until the true tree is in the supertree set. The fraction in parentheses is the number of trees added divided by the number of trees in the perfect PDA model.

With MRC the optimizing function is maximizing the number of compatible columns, that is, the corresponding splits can be arranged in a tree without conflict. In the consensus setting studied here, the MRC tree is equivalent to the tree with the maximal number of splits in the data, that is, to the asymmetric median tree (Phillips and Warnow 1996). Equivalently, the method can also be defined as *minimizing* the number of columns not present in the supertree, a notation that is used from now on. The number of columns not present in the supertree is called compatibility length (CL).

MRP uses parsimony to reconstruct a tree from the matrix representation. For a candidate supertree, the

sum of parsimony lengths (PLs) over the columns along that tree is computed. The optimizing function is minimizing the sum of PLs. The sum over the whole MR is called PL of the tree.

In this note, all optimal trees are found by exhaustive search. Note that there can be multiple supertrees with the same minimal CL and PL, respectively, thus I talk about the "supertree set." Optimal MRC trees were found with the python script `mrc.py` that evaluates all possible topologies for a particular number of taxa (available from http://www.cibiv.at/software/mrc/). Optimal MRP trees were computed with the branch-and-bound option in PAUP* (Swofford 2002).

RESULTS

Perfect Distributions

First, I perfectly model the null distributions. Under the perfect PDS model, the MR contains exactly one split of each kind (e.g., 25 columns for $n = 6$, Table 1). Then the CLs are equal for all trees. Because each tree contains $n - 3$ inner splits, the CL of each tree is $l-n+3$. The PLs of all trees are also equal if each possible split is given as input (Steel 1993). Thus, for MRC and MRP, the supertree set contains every tree if the perfect PDS model is used as input.

Under the perfect PDA model, the MR is built by coding each tree once. Then the CLs are not equal for all trees. Instead, the optimal supertrees for MRC are all of one shape. This shape is called the *optimal shape* for MRC. Because the PDA model is label invariant, the supertree set contains all trees of the optimal shape. The analog holds for MRP. For six to nine taxa the optimal shape is the same for MRC and MRP (bold shapes in Table 2). For example, for $n=6$ the supertree set contains 15 trees, all being of shape $S_{6,2}$. These optimal shapes tend to be balanced, that is, they have a low Colless index. Note that MRC and MRP do not always result in the same optimal shapes (Fig. 1).

In the following, I investigate MRC and MRP with deviations from the perfect models. Thereby, the focus is on two special shapes for each n: the *optimal shape* that was found under the perfect PDA model and the *unbalanced shape* that has exactly two 2-splits.

Resampling Randomly from the Distributions

Here, I resample a particular number of trees t from each distribution. For the PDS model, instead of resampling t trees, $t\times(n-3)$ splits are randomly drawn with replacement. I evaluate the fraction of optimal shapes among the resulting supertrees. If the supertree set for one data set contains p trees of which q exhibit the optimal shape, then this data set shows a fraction of q/p of the optimal shape. The results in Figure 2 are obtained by averaging this fraction over 100 randomly generated data sets.

Under the PDA model, a certain fraction of the optimal shape is expected (Table 2). These fractions are marked dashed in Figure 2. With the PDS model, the optimal shape occurs a bit less frequently in the supertree sets than expected. In contrast, if the input trees are generated under the PDA model, the supertrees show the optimal shapes more often than expected. The observed fraction of optimal shapes grows with the number of input trees. Furthermore, the bias is less strong for MRP, in particular for odd numbers of taxa.

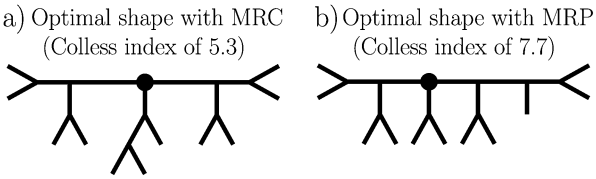


FIGURE 1. Optimal shapes for $n=11$. Note that MRC and MRP can result in different optimal shapes.

Perfect Distributions with Phylogenetic Information

Next, I ask how the methods behave in the presence of "little" phylogenetic information disturbed by noise. First, noise is modeled by the perfect PDA model. Note that the supertree methods show a shape bias if the perfect PDA model is used as input (Table 2). The little phylogenetic information is modeled by adding the "true" tree i times to the perfect PDA model. The *critical number* i_c is the smallest i such that the true tree is in the supertree set. I observe that i_c depends strongly on the shape of the true tree (Table 2). It ranges from 1 tree for the optimal shape up to 10%. In general, the critical numbers are lower for MRC than for MRP. If $i_c=1$, the true tree is the only supertree for all i. Otherwise, the supertree set does not contain the true tree for $i < i_c$, the supertree set contains the true tree for $i = i_c$, and the true tree is the only supertree for $i > i_c$.

For example, for $n=6$ and $S_{6,1}$, the true tree is T_1 (Fig. 3a). If this tree is added once to the pefect PDA model, the supertree for MRC and MRP, respectively, is T_2 (Fig. 3b). Thus, the supertree is balanced and includes the two 2-splits of the true tree ($\{\mathtt{t1},\mathtt{t2}\}|\{\mathtt{t3},\mathtt{t4},\mathtt{t5},\mathtt{t6}\}$ and $\{\mathtt{t5},\mathtt{t6}\}|\{\mathtt{t1},\mathtt{t2},\mathtt{t3},\mathtt{t4}\}$). For MRC, this result is explained by the CLs: The balanced tree T_2 has a length of $315-3\times15+1=271$ (315 splits from the PDA model, of which the three 2-splits from T_2 occur in 15 trees each and one 3-split from T_1 increases the length by one; see also Table 1). Analogously, the unbalanced tree T_1 has a length of $315-2\times15-9=276$ (the two 2-splits from T_1 occur in 15 trees each and the one 3-split from T_1 occurs in 9 trees). For this example, i_c is computed by $315-3\times15+i_c=315-2\times15-9$, thus $i_c=6$ and from $i=7$ on the true tree is the only supertree. Below the critical number the nonuniform split distribution, which prefers 2-splits, outvotes the 3-split.

If one "true" tree is added to the perfect PDS model, then the supertree equals the true tree for MRC and MRP and all shapes presented in Table 2.

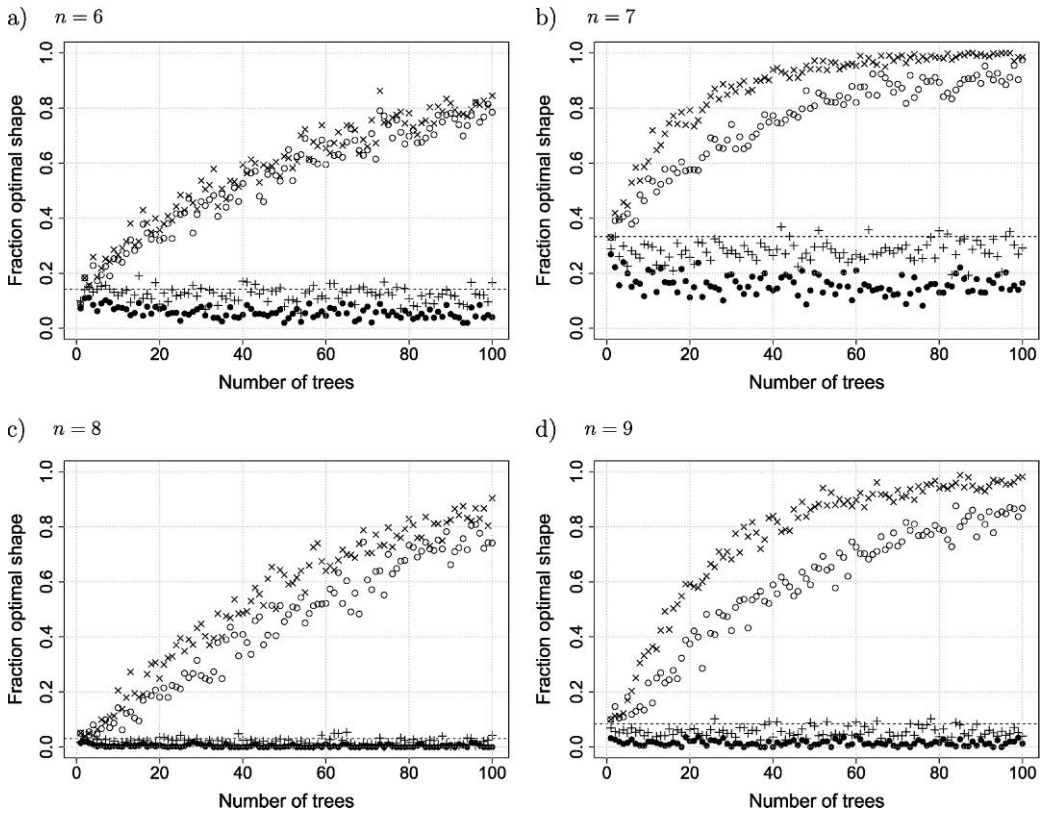


FIGURE 2. Resulting supertree shapes if random input trees are given (100 repetitions for each number of trees). The dashed line is the expected fraction of the optimal shape under the PDA model (Table 2).

Resampling Randomly from the Distributions and Adding Phylogenetic Information

The previous section showed that 10% of phylogenetic information is sufficient such that the supertree equals the true tree in a perfect setting. However, this behavior is disturbed by noise. To show this, I add the "true" tree with a fraction of 10% to the random trees used for the analysis displayed in Figure 2. Therefore, only the data sets with 10, 20, ..., 100 trees are taken and the true tree is added 1, 2, ..., 10 times. The true tree is either one topology showing the optimal shape or one topology showing the unbalanced shape. The accuracy for one data set is $1/p$, if p supertrees are found and the true tree is among them, and 0 otherwise. The accuracies are averaged over 100 data sets.

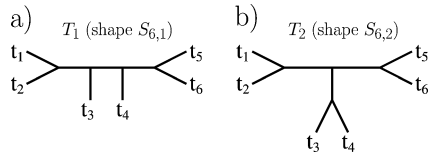


FIGURE 3. Example trees.

Under the PDS model (Fig. 4, left column), there is no observable difference in accuracy between the true trees showing different shapes. For the PDA model (Fig. 4, right column), however, the tree with the optimal shape is reconstructed correctly with higher probability than the tree with the unbalanced shape. Furthermore, the accuracies increase with the number of true trees. Apparently, 10 true trees out of 110 trees provide more information compared with 1 true tree out of 11 trees. The accuracies are higher for MRC than for MRP. This is consistent with the results for the perfect setting that MRP needs more trees until the true tree is in the supertree set (Table 2).

CONCLUSIONS

Wilkinson et al. (2005) report a shape bias toward an unbalanced tree for MRP which at first view contradicts my result that the optimal shapes for both MRP and MRC are more balanced. However, their setting is different from mine. They investigate two arbitrarily chosen input trees of different shapes and observe an asymmetry in the PLs: the PL of a balanced input tree on an unbalanced supertree is shorter than vice versa.

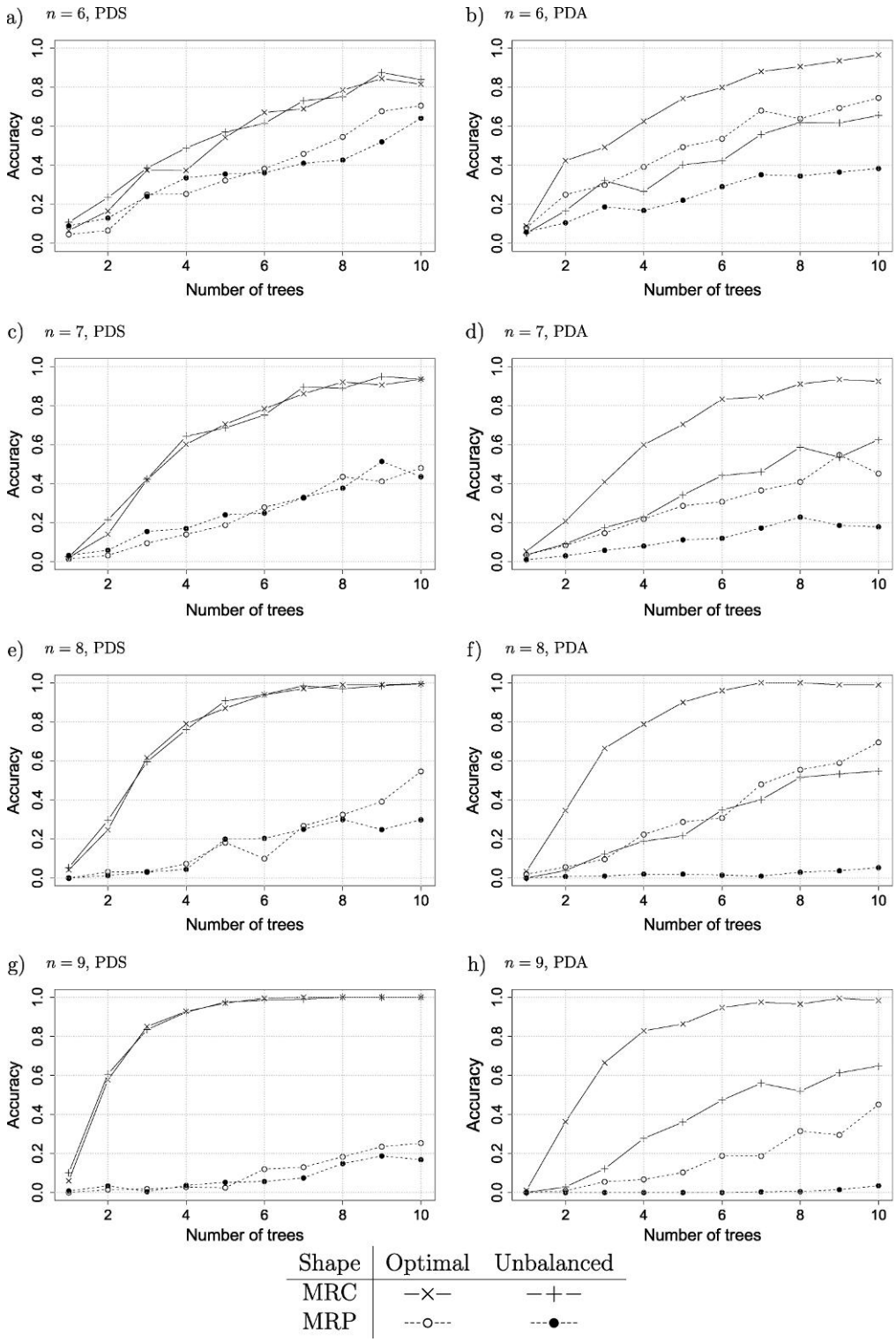


FIGURE 4. Adding information to random trees. The random trees are taken from Figure 2 and a fraction of 10% of the true tree is added (x-axis is labeled with the number of added true trees). Accuracy is the fraction the true tree was found by the supertree methods (average over 100 repetitions). Optimal shapes are $S_{6,2}$, $S_{7,2}$, $S_{8,4}$ and $S_{9,5}$. Unbalances shapes are $S_{6,1}$, $S_{7,1}$, $S_{8,1}$ and $S_{9,1}$ (see Table 2).

This can be explained by an inherent feature of the parsimony score: The PL of a coded k-split does not exceed k on any tree. More balanced trees contain more splits with small k (Table 2), that is, where the numbers of taxa are nonuniformly distributed. They necessarily have lower maximal PLs. Splits with higher k do not only

have higher maximal PLs, but in addition, the PL distribution on random trees is shifted toward higher values (Maddison and Slatkin 1991). This explains that coding the balanced tree in the MR and evaluating it on the unbalanced tree is favorable compared with coding the unbalanced tree and evaluating it on the balanced tree.

For more than two random input trees an optimal shape is preferred which is more balanced (Fig. 2). This holds for MRC as well as MRP although CLs are not asymmetric. Furthermore, it even holds if unbalanced and balanced shapes are not uniformly distributed among the input shapes, but when unbalanced input tree shapes are favored (e.g., for $n=6$, the input trees are drawn from a distribution that contains the unbalanced shape in 86%; see also Table 2). This shape bias is positively misleading, that is, it grows with the number of input trees. However, it grows more slowly for MRP. An explanation may be that the bias toward unbalanced shapes due to the asymmetric PLs (Wilkinson et al. 2005) acts as counterbalance.

This shape bias is only observed for the PDA model (all trees are equally likely), not for the PDS model (all splits are equally likely). The two null models do also behave differently if little phylogenetic information is added, that is, where one or more "true" trees are added to the perfect distributions. Only under the PDA model, the supertree set may not contain the true tree for some tree shapes. Instead, up to 10% of the true tree are needed such that it is contained in the supertree set (Table 2). When adding one true tree with an unbalanced tree shape, the supertree is not this tree but a more balanced tree. This balanced supertree preserves only some splits present in the true tree. The shape bias is also present if not the complete list of trees but random trees are given, and the true tree is added with a fraction of 10% (Fig. 4). In this case, a tree with the optimal shape is reconstructed correctly with higher probability than a tree with the unbalanced shape. Furthermore, there is a difference in the accuracy for different n. For MRC, the accuracy is increasing with n. That may be explained by the fact the the probability of a particular 2-split is $1/(2n-5)$ and thus decreases with n (see also McKenzie and Steel 2000). Thus, the information added by multiple trees is more specific for higher n. This effect is not present in MRP. It is probably counterbalanced by the fact that higher n can also have splits with higher k, which in turn can have higher PLs.

I note that the shape bias presented here is a consequence of the complex space of trees. Under the PDA model, 2-splits are more likely than other splits (Table 1) and no distribution on bifurcationg trees ensures that all splits are equally likely (Steel and Pickett 2006). This results in a bias toward balanced trees with any split-based supertree reconstruction method. The same relation causes the problem that uniform priors on tree topologies do imply nonuniform priors on splits in Bayesian analysis, thus posterior probabilities may be biased (Pickett and Randle 2005). Supertree methods that are not based on the encodings of full splits may be less affected by the presented bias. Different matrix encodings are possible (e.g., Wilkinson, Cotton, et al. 2004) for the compatibility or parsimony analyses, or a completely different class of methods could be used (e.g., Semple and Steel 2000, and variants thereof). Note that the latter methods are based on rooted input trees. An investigation of the tree shape bias in the rooted setting would also be of interest.

In general, I do not assume that the shape bias has an impact on usual supertree reconstruction. Here, problems occur if 90–100% are random data or for small numbers of trees. This is not expected to be the case in any analysis. However, in sparse data sets, only few input trees may carry the information for some splits. If they are highly conflicting, the shape bias may play a role at least locally. Note that an implicit bias in a phylogeny reconstruction method also has little influence as soon as data is present (Vinh et al. 2010).

Furthermore, I cannot make any conclusions about whether MRC or MRP should be preferred because they show the bias to a different extent depending on the model. If only random trees are given as input, the bias is less strong in MRP (Fig. 2). On the other hand, if little information is added, MRC can better find the true tree (Fig. 4).

These findings allow implications about the design of supertree methods that explicitly model missing data. I conclude that modeling missing data by generating all possible trees may introduce a bias toward more balanced tree shapes when applying split-based supertree methods. To date, supertree methods usually do not model missing data explicitly (but see the definition of majority-rule(+) supertrees in Cotton and Wilkinson 2007). Instead, for MRC, treating missing taxa as gaps is equivalent to generating all possible binary characters by replacing the gaps with `0`s and `1`s (Rodrigo 1996). This corresponds to the PDS model that is not affected by the positively misleading tree shape bias.

Funding

Financial support to Arndt von Haeseler from the Wiener Wissenschafts-, Forschungs- and Technologiefonds (WWTF) is greatly appreciated.

Acknowledgments

The author likes to thank Mareike Fischer for helpful discussions and Arndt von Haeseler for valuable discussions and comments on the manuscript. The comments of the referees helped to substantially improve the manuscript.

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., Ragan M.A. 2004. The MRP method. In: Bininda-Emonds O. R. P., editor. Phylogenetic supertrees: combining information to reveal the tree of life. Dordrecht (The Netherlands): Kluwer Academic, p. 17–34.

Bininda-Emonds O.R.P., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. 2007. The delayed rise of present-day mammals. Nature. 446: 507–512.
Colless D.H. 1982. Phylogenetics: the theory and practice of phylogenetic systematics II [book review]. Syst. Zool. 31:100–104.
Cotton J.A., Wilkinson M. 2007. Majority-rule supertrees. Syst. Biol. 56:445–452.
Furnas G.W. 1984. The generation of random, binary unordered trees. J. Classif. 1:187–233.
Maddison W.P., Slatkin M. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. Evolution. 45:1184–1197.
McKenzie A., Steel. M. 2000. Distributions of cherries for two models of trees. Math. Biosci. 164:81–92.
Phillips C.A., Warnow T.J. 1996. The asymmetric median tree—a new model for building consensus trees. Discr. Appl. Math. 71: 311–335.
Pickett, K.M., Randle C.P. 2005. Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Mol. Phylogenet. Evol. 34:203–211.
Ragan M.A. 1992. Phylogenetic inference based on matrix representation of trees. Mol. Phylogenet. Evol. 1:53–58.
Rodrigo A.G. 1996. On combining cladograms. Taxon. 45:267–274.
Rosen D.E. 1978. Vicariant patterns and historical explanation in biogeography. Syst. Zool. 27:159–188.
Ross H.A., Rodrigo A.G. 2004. An assessment of matrix representation with compatibility in supertree reconstruction. In: Bininda-Emonds, O.R.P., editor. Phylogenetic supertrees: combining information to reveal the tree of life . Dordrecht (The Netherlands): Kluwer Academic, p. 35–63.
Semple C., Steel M. 2000. A supertree method for rooted trees. Discr. Appl. Math. 105:147–158.
Shao K.-T., Sokal R.R. 1990. Tree balance. Syst. Zool. 39:266–267.
Steel M. 1993. Distributions on bicoloured binary trees arising from the principle of parsimony. Discr. Appl. Math. 41:245–261.
Steel M., Pickett K.M. 2006. On the impossibility of uniform priors on clades. Mol. Phylogenet. Evol. 39:585–586.
Steel M.A., Penny. D. 1993. Distributions of tree comparison metrics—some new results. Syst. Biol. 42:126–141.
Swofford D.L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
Vinh L.S., Fuehrer A., von Haeseler A. Forthcoming 2010. Random tree-puzzle leads to the Yule-Harding distribution. Mol. Biol. Evol. doi:10.1093/molbev/msq212.
Wilkinson M., Cotton A., Creevey C., Eulenstein O., Harris S.R., Lapointe J.F., Levasseur C., McInerney J.O., Pisani D., Thorley J.L. 2005. The shape of supertrees to come: tree shape related properties of fourteen supertree methods. Syst. Biol. 54: 419–431.
Wilkinson M., Cotton J.A., Lapointe F.-J., Pisani D. 2007. Properties of supertree methods in the consensus setting. Syst Biol 56: 330–337.
Wilkinson M., Cotton J.A., Thorley J.L. 2004. The information content of trees and their matrix representations. Syst. Biol. 53: 989–1001.
Wilkinson M., Thorley J.L., Pisani D., Lapointe F.-J., McInerney J.O. 2004. Some desiderata for liberal supertrees. chap. 10. In: Bininda-Emonds, O.R.P. editor. Phylogenetic supertrees: combining information to reveal the tree of life. Dordrecht, The Netherlands: Kluwer Academic.