Chapter

Some Desiderata for Liberal Supertrees

MARK WILKINSON[1], JOSEPH L. THORLEY[2], DAVIDE PISANI[3,] FRANCOIS-JOSEPH LAPOINTE[4], JAMES O. MCINERNEY[5]

[1]*Department of Zoology, The Natural History Museum, London SW7 5BD, UK*

[2]*Fisheries Research Service, Freshwater Laboratory, Faskally, Pitlochry, Perthshire PH16 5LB, UK*

[3]*Department of Biology, The Pennsylvania State University. 208 Mueller Laboratory, University Park, 16802, PA, USA*

[4]*Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal (Québec), H3C 3J7, Canada.*

[5]*Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland.*

Abstract: Although a variety of supertree methods has been proposed, our understanding of these methods is limited. In turn, this limits the potential for biologists who seek to construct supertrees to make informed choices among the available methods. In this chapter we distinguish between supertree methods that offer a conservative synthesis of the relationships that are agreed upon or uncontradicted by all the input trees and liberal supertree methods which have the potential to resolve conflict. We list a series of potential desiderata (i.e. desirable properties) of liberal supertree methods, discuss their relevance to biologists and highlight where it is known that particular methods do or do not satisfy them. For biologists, the primary aim of liberal supertree construction is to produce accurate phylogenies and most of our desiderata relate to this prime objective. Secondary desiderata pertain to the practicality of supertree methods, particularly their speed.

Key words: Supertrees, Phylogeny, Consensus, Tree of Life, Axioms.

1.   INTRODUCTION

Although the field of supertree construction is young, tracing its origins to Gordon's (1986) seminal contribution, there is already a rich diversity of supertree methods and variants that have been developed or outlined, some of which are in increasingly common use. Unfortunately our understanding of these methods has not kept pace with their explosive development. The availability of multiple supertree methods means that those who seek to use them are confronted by methodological choices: which method(s) should they use? Understanding the properties of supertree methods must be key to rational choice.

Here we discuss some desiderata of supertree methods, the properties that we might like such methods to have. We have been inspired by the approach taken by some mathematicians to the characterisation and exploration of consensus methods in terms of consensus axioms (e.g. Barthélemy *et al.*, 1995; McMorris and Neumann, 1983), well-defined mathematical properties that they may or may not possess. Thus far, the important paper by Steel *et al.* (2000) is the only application of the axiomatic approach to supertrees. Some of the properties we discuss derive from the literature on consensus axioms, whereas others have never been discussed in that literature and originate from a biological rather than a mathematical perspective. At least some of the latter might be open to formal investigation as additional supertree (or consensus) axioms. However, we are biologists rather than mathematicians and our treatment is very informal. As biologists we are interested in particular properties inasmuch as they impact upon our ability to do biology. Thus we aim to clarify why the properties we discuss might be considered desirable for biologists, rather than taking their desirability to be self-evident or axiomatic.

2.   What is a Supertree?

A supertree is a tree that amalgamates, synthesises or otherwise represents the phylogenetic relationships included in a set of input trees. Under this loose definition, consensus trees are supertrees constructed in the special case of input trees with identical leaf sets. The diversity of consensus methods in evolutionary biology reflects in part a diversity of potential uses for them (Barrett *et al.*, 1991; Swofford, 1991; Wilkinson, 1994). For example, strict consensus methods are used to summarise unanimous agreement across a set of input trees thereby identifying those relationships that are 'strictly supported' (Nixon and Carpenter, 1996). In contrast, majority-rule consensus methods, which summarise those relationships occurring in a majority of the input trees, are used, for example, to represent

the results of bootstrapping (Felsenstein, 1985, Wilkinson, 1996), jackknifing, (Farris *et al.,* 1996), quartet puzzling (Strimmer and von Haeseler, 1996) and Bayesian analyses (Larget and Simon, 1999). The utility of consensus methods depends upon what we wish of the consensus summary and we might expect the same to be true of supertree methods.

We see four main possible uses for supertrees. For the most part, applications of supertree methods have sought to produce well-resolved large phylogenies from sets of smaller, typically conflicting, input trees. Here supertrees are meta-analytic syntheses of the input trees (Mann, 1990; Sanderson *et al.*, 1998; Bininda-Emonds *et al.,* 2002) that are intended to provide a phylogenetic framework for broad comparative studies. Resolution of input tree conflicts is hoped for, and thus liberal supertree methods are used. The degree to which resolution is achieved depends upon the input trees providing differential support for conflicting relationships (as assessed by the supertree method), and potentially also by the degree of effective overlap between input trees. Secondly, supertrees might also be used in quantitative studies of input tree congruence. For example, outliers or unstable may be identified, using one or more input tree - supertree distance measures (e.g. DasGupta *et al.*, 1997) and positional congruence scores (Estabrook *et al.*, 1985). Thirdly, supertrees can be used simply to explore and identify agreement and disagreement among sets of input trees. In this case the aim is to reveal conflict rather than resolve it, typically through the use of conservative supertree methods, with any resolution coming ultimately from additional data or new analyses sought or performed in the light of the supertree (Wilkinson *et al.*, 2001). Again, supertrees will be more or less resolved depending upon the extent of conflict, and the degree of effective overlap between input trees. The fourth potential use is in identifying where limited overlap between the leaf sets of input trees is an obstacle to their amalgamation, in order to guide further research aimed at providing effective overlap (Wilkinson *et al.*, 2001; Burleigh *et al.*, this volume). While all these uses are important, we focus here upon liberal supertrees that are capable in principle of providing well resolved meta-analytical syntheses in the face of conflicting input trees. Thus we do not discuss the more conservative strict or semi-strict supertree methods (Bryant, 2002; Goloboff and Pol, 2002) which may be particularly well suited to the latter two uses.

3.   Some Liberal Supertree Methods

3.1 Matrix representations. Trees can be represented by a variety of corresponding matrices. A number of supertree methods combine matrix representations of input trees into a single matrix that can be analysed to

yield supertrees. Methods differ in the form of matrix representation employed and the kind of analysis. The average consensus procedure combines pairwise distance matrices and uses a least squares optimality criterion in searching for the best tree (Lapointe and Cucumel, 1997). We refer to this as a matrix representation with distances (MRD) method (Lapointe *et al*, submitted). Most practitioners have employed matrix representations that encode trees as 'pseudocharacter' data that is then analysed with parsimony, the matrix representation with parsimony (MRP) approach to supertree construction. In standard MRP (Baum, 1992; Ragan, 1992), one binary pseudocharacter encodes each internal branch on each input tree (component or cluster coding), and standard reversible (Fitch or Wagner) parsimony is used. Irreversible MRP (Bininda-Emonds and Bryant, 1998) differs only in its use of irreversible parsimony. Purvis MRP (Purvis, 1995a) uses reversible parsimony and differs from standard MRP in the matrix representation. Each matrix element splits the members of a clade from the members of its sister group (or of all possible sister groups in the case of polytomies) and the root, with any other leaves scored as missing. In triplet and quartet MRP (Thorley, 2000; Wilkinson *et al.*, 2001), one binary pseudocharacter encodes each resolved triplet or quartet respectively in each input tree, and standard reversible parsimony is used. Purvis (1995a) and Rodrigo (1996) suggested, and Pisani (2002) and Ross and Rodrigo (this volume) explored clique analysis as an alternative to parsimony, using component coding in their matrix representation with compatibility (MRC). In the special cases of triplet and quartet matrix representations, maximum parsimony and maximal cliques define the same optimal trees, so that MRC = MRP. A further matrix representation method involves recoding (flipping) individual entries in a component matrix representation, moving leaves into or out of clusters or from one 'side' of a split to another so as to render the matrix compatible. Optimal matrix representation with flipping (MRF) supertrees are those supported by the matrices requiring the fewest recodings (Chen *et al.*, 2003).

   3.2 MinCut. Aho *et al.* (1981) developed a fast algorithm for amalgamating a set of compatible trees. If the trees are compatible this method returns a single supertree that contains all the input trees. Where input trees conflict, the method yields no tree. The MinCut supertree method, developed by Semple and Steel (2000) modifies the Aho *et al.* method to deal with conflicting input trees. Essentially, this is done by breaking apart conflicting clusters in a certain minimal way that ensures a number of desirable properties for MinCut supertrees (for details see Page, 2003; Semple and Steel, 2000). MinCut has some Adams consensus like properties (Semple and Steel, 2000) and whether it is considered liberal or

conservative may depends on whether the clusters in the supertree are interpreted as nestings or components (see Wilkinson, 1994).

3.3 Quartet puzzling. Quartet puzzling (Strimmer and von Haeseler, 1996) is an heuristic method for building resolved, comprehensive trees from sets of quartets that may or may not conflict. It is therefore a liberal supertree method. However, as normally used, it draws upon the quartet trees inferred for all possible quartets for the full set of leaves under consideration, using these in a voting procedure to determine where to add leaves to a growing tree. This is a special case, and in the normal supertree context not all quartets will be included in the input trees. Pisani and Wilkinson (2002) indicated the potential for a quartet puzzling supertree method, but to be effective the voting procedure needs modification (Pentony *et al.*, in prep).

In quartet puzzling, tree construction is iterated with different addition sequences and random breaking of ties. The multiple trees produced are summarised with a majority rule consensus and the frequencies of relationships taken as an index of support (Strimmer and von Haeseler, 1996; Wilkinson *et al.* 2003). Each quartet puzzling iteration can be thought of as providing a fast and greedy heuristic approximation of the supertree that contains the largest number of input quartets. Thus the method is closely related to quartet MRP. With rooted trees, triplet puzzling (quartet puzzling but using only quartets in which one leaf is the root) would be analogously related to triplet MRC/MRP. We can envisage similar heuristics that choose a starting tree from the input trees and add taxa in one at a time according to inference and fusion rules (Bryant, 1997; Dekker, 1986; Wilkinson *et al.*, 2000) and greedy local optimizations that approximate objective functions based on a number of tree-to-tree distances.

4. Accuracy

Many consensus axioms describing desirable mathematical properties of consensus methods have been discussed but mostly with little consideration of their relevance to what is desirable or important to biologists. In the specific context of the construction of liberal supertrees, we believe biologists are (or should be) primarily concerned with accuracy. By accuracy we mean correspondence with actual phylogenetic relationships ('accuracy with a capital A') rather than, for example, correspondence between the objective function of a method and heuristically selected supertrees. The ultimate aim must be to have accurate phylogenies that provide maximally useful phylogenetic frameworks for comparative biology (Lanyon, 1993). The ability of any method to construct accurate supertrees under a range of readily modeled analytical conditions can be assessed by simulation (e.g. Bininda-Emonds and Sanderson, 2001; Chen *et al.*, 2003; Lapointe and

Levasseur, this volume; Ross and Rodrigo, this voume). However, the ability of methods to produce accurate trees depends very much on properties of the data, and insights from simulations notwithstanding, for the most part we do not know how accurate real supertrees are. In the absence of assessment of accuracy, we can examine other properties as surrogates. For example, we might investigate whether supertrees include relationships that we might reasonably expect to be present, or conversely relationships we would not expect. Similarly, we can address whether the resolution of conflict is affected by properties of input trees other than those we might expect the resolution to be based upon, i.e. properties that are irrelevant to our understanding of the weight of support for particular relationships. The following is a far from exhaustive set of such properties.

4.1 Independence. Bryant (1997) gives formal definitions of two 'independence' consensus axioms that relate to the insensitivity of consensus methods to the addition or pruning of input tree leaves (but which might be characterized in terms of any well-defined operation on trees). The independence (of irrelevant alternatives) axiom considers two profiles of trees. If the two profiles can be rendered identical by pruning some set of leaves from the trees in each profile, then if we prune the same leaves from the consensus trees for each profile these consensus trees should also be identical. The second independence axiom states that given a set of input trees from which some particular leaves are pruned, the consensus or supertree of the pruned trees might be expected to be the same as the pruned consensus or supertree of the full input trees. A consensus method that satisfies the second axiom must also satisfy the first (Barthélemy *et al.*, 1995). It seems reasonable that extraneous information on the relationships of other (pruned) leaves should not impact upon the relationships inferred among other leaves. There has been little investigation of independence axioms in the context of supertrees. In this context, input trees can logically entail relationships in combination that are not present in any single input tree, so that pruning selected leaves from input trees could remove some entailed relationships and impact upon the supertree. In this case, the additional information is useful rather than irrelevant and failing to obey independence axioms would not necessarily be undesirable. The following three properties may be related to the general idea of independence. They are properties that biologists have or might find desirable but which have not been discussed much in the mathematical literature.

4.2 Sizeless. Suppose we wish each input tree to have equal weight. This might be reasonable if we had no basis for assigning differential weights. Purvis (1995a) provided an example showing a bias in standard MRP, in cases of conflict, towards relationships in larger trees, and Purvis coding was proposed to remedy the bias. Subsequently, Ronquist (1996) showed that

Purvis's coding method does not succeed in removing size bias, and suggested that this be done by weighting the pseudocharacters from each input tree inversely with respect to their number. Bininda-Emonds and Bryant (1998) showed the size bias was with respect to the sizes of conflicting subtrees rather than the sizes of the input trees per se. Consequently inverse weighting on tree size would not correct the size related bias. Sanderson *et al.* (1998) summarised that no method was known that always weighted trees equally. Of course, this is true only for liberal supertree methods and does not hold for more conservative strict and semi-strict supertree methods. Page (2003) used a simple example to show a size bias (toward larger trees) in MinCut supertrees that led him to propose a modification. The extent of size biases for different supertree methods is not well known. Because the addition and or pruning of leaves changes size then methods that are not sizeless will not obey independence axioms.

Size bias seems like a serious problem if we want to weight trees equally. Such equal weighting might be justified by the principle of indifference (Keynes, 1920) if there is no basis for differential weighting of trees. However, the principle of indifference might also be invoked to justify equal weighting of components or of triplets, and because larger (binary) trees include more components and more triplets then achieving equal weighting of these will entail unequal weighting of trees. Ronquist (1996) argued that the size bias of MRP methods was not unreasonable because larger trees contain more information. We are concerned with size biases in supertree methods only to the extent that these might promote inaccuracy. If large trees, were in general more accurate than smaller trees we would have reason to be unconcerned but we do not think this is generally the case. Our concern is really that whatever biases might exist, they should not be so severe as to prevent supertree methods from returning relationships that appear the best supported in terms of their frequency of replication in, or entailment by, the input trees and any additional information on their relative strength of support (see below).

An interesting approach to removing size biases would be to convert input trees with overlapping leaf sets to input trees with identical and full leaf sets by grafting leaves onto the input trees. There may be many ways of doing this for any given input tree, defining a span of candidate supertrees for each input tree (Bryant, 2002). Fast heuristics might be used to generate a single 'best' candidate from each input tree span that can then be amalgamated with (e.g. majority-rule) consensus. Semple and Steel (2002) have described a method for encoding a tree of any size with five multistate characters and a suggestion that has yet to be explored is that such representations might be used to avoid size biases (Bininda-Emonds *et al.*, 2002)

4.3 Shapeless. Tree shape or balance (Shao and Sokal, 1990) is a characteristic of input trees that might reasonably be considered irrelevant to their evidential significance. We might therefore desire supertree methods that, in cases of conflict, do not unduly favour relationships in asymmetric or in symmetric trees. Several supertree methods are biased with respect to tree shape. For example, in cases of conflict, standard and irreversible MRP and MRF are biased toward relationships in asymmetric trees and Purvis MRP is biased toward relationships in symmetric trees (Wilkinson *et al.*, 2001, in prep). These biases in the MRP methods appear to stem from the use of asymmetric distances or fit functions to define the optimal supertree.

Thorley and Wilkinson (2003) suggested that supertrees can be conceived of as trees that minimise the sum of the distances between the supertree and each input tree (see also Bryant, 2003). Hence methods can differ in the distance metric (objective function) and the typically heuristic method used to approximate optimal trees. The distance between the supertree and an input tree in MRP is given by the fit (parsimony steps) of the matrix representation of the input tree to the supertree. With standard, irreversible and Purvis MRP this distance is asymmetric, it is not equal to the fit of the matrix representation of the supertree (pruned of irrelevant leaves) to the input tree (Thorley and Wilkinson, 2003). In standard and irreversible MRP, symmetric trees have smaller distances to asymmetric trees than vice versa and the reverse is true of Purvis MRP (Wilkinson *et al.*, in prep). Shape bias of supertree methods has not been extensively investigated and it is not known to what extent failure to be shapeless matters in practice. However, we find it difficult to conceive of any justification for such bias and would prefer shapeless methods if they existed.

4.4 Positionless. Wilkinson *et al.* (2001) presented a simple example that suggested that some MRP methods tended to resolve conflicts in favour of more crownward (Purvis, triplet) or basal (irreversible) positions of leaves that contribute to the conflict. Bininda-Emonds and Bryant (1998) also noted the apparent basal bias of irreversible MRP. As with tree shape we find it difficult to conceive of justifications for such behaviour and would prefer supertree methods that have no such biases, an admittedly vaguely characterised property we term positionless (se also Cotton and Page, this volume). Very little is known about the extent to which existing supertree methods satisfy this potential desideratum, and further investigation would require a clearer conceptualization and quantification of the kind of positional relations referred to by biologists as more basal or more crownward.

4.5 Order invariance. We might expect that supertrees should be unaffected by the order in which input trees are processed (often termed neutrality) and, in the case of matrix representation methods, the order of

leaves in the matrix (often termed symmetry or equality or anonymity). Neutrality corresponds to property P1, and equality to P2 of Steel *et al.* (2000). MinCut has these properties (Semple and Steel, 2000). Heuristic methods may or may not be order invariant (e.g.. use of closest versus multiple random addition sequences in MRP), with greedy heuristics tending to sacrifice this desideratum for speed. Order invariance is desirable because we expect one accurate tree. However, the extent to which relationships in supertrees actually vary with input tree or leaf order can be determined and may provide useful information on which relationships are robustly supported by the input trees and those that are not.

4.6 Uniqueness. Methods that have the property of uniqueness always return a single supertree. Desiring a unique supertree might be seen as a natural consequence of desiring complete accuracy (on the assumption of only one true supertree). However, there may be good reason to prefer a method to return multiple trees (see Lapointe and Cucumel, 2002), such as when there are equally optimal solutions. MinCut is the only liberal supertree method that will always return a single tree. Uniqueness can be additionally imposed on other methods by conjoining them to consensus methods with this property (Steel *et al.*, 2000), resulting in unique consensus supertrees with properties determined by both the supertree and the consensus methods used. With quartet puzzling, use of the majority-rule consensus to summarise the individual supertrees produced by each iteration of the method is integral to that approach to supertree construction.

4.7 Plenary. A plenary supertree is one that includes all the leaves of the input trees. Desiring a plenary supertree is a natural consequence of desiring complete accuracy. All the supertree methods that we are considering are plenary, but those methods that return multiple trees can be rendered non-plenary through the use of non-plenary consensus methods, such as reduced consensus (Wilkinson and Thorley, 2003) and agreement subtrees (Finden and Gordon, 1985; Bryant, 1997). Non-plenary supertree methods may be most useful for identifying unstable leaves, localising conflict and identifying areas with ineffective overlap. The plenary axiom corresponds to property P4 of Steel *et al.* (2000)

4.8 Pareto. A number of important consensus axioms pertain to the extent to which relationships in the consensus are present in the input trees and vice versa. A consensus is Pareto with respect to a particular kind of relationships (clusters, nestings, triplets, etc.) if all such relationships that are present in every tree are present in the consensus. This is a very reasonable expectation if agreement is taken as strong surrogate for, and evidence of, accuracy. We therefore desire Pareto supertree methods. Most supertree methods do appear to be Pareto on one or more type of relationship. Thorley (2000) noted that the various MRP methods are Pareto on clusters (full

splits, components), Chen *et al.* (2003) and Semple and Steel (2000) showed that MRF and MinCut respectively share this property, and we suggest this is true of MRC methods also (see Pisani 2002). MinCut is also Pareto on nestings and upon triplets (Semple and Steel, 2000). We conjecture that MRC and MRP methods are Pareto with respect to the type of relationships encoded in the matrix representations (i.e. full or partial splits), but not Pareto on less inclusive relationships. Thus, standard MRP is not Pareto on triplets (Thorley, 2000; Bininda-Emonds *et al.*, 2002; Wilikinson *et al.*, in prep). Steel *et al.*'s (2000) properties P6 and P6' correspond to being Pareto on quartets and triplets respectively.

In the supertree context, input trees may have different leaf sets so that it may be impossible for any relationship to be present in all input trees. We might therefore expect that any relationships in one or more input trees that are uncontradicted by other input trees would be in the supertree. However, Steel *et al.* (2000) have shown that no supertree method can have this property for rooted trees (their P7) and we expect this to apply to unrooted trees also. This is because, although a given relationship X may be uncontradicted by any input tree, collections of input trees can entail relationships contradicting X. Thus we could further weaken the condition to expect any relationship in an input tree that is not contradicted by other input trees, singly or in combination, to be included in, or not contradicted by, the supertree. The semi-strict supertree method of Goloboff and Pol (2002) is a heuristic approach intended to satisfy this desideratum. It is unclear to what extent liberal, conflict busting, supertree methods do so.

4.9 Co-Pareto. A consensus is co-Pareto with respect to a particular kind of relationships, if every relationship of that kind that is present in the consensus tree is present in one or more input trees. Consensus methods that do not obey this axiom are problematic if we consider that relationships that do not occur in any input tree are unsupported. In general it is not reasonable to expect supertree methods to be co-Pareto because they may reasonably contain relationships that are entailed by the input trees in combination but not present in any of them singly. However, standard, irreversible and Purvis MRP are not co-Pareto on clusters or on triplets even in the special case (the consensus context) of input trees with identical leaf sets (Wilkinson *et al.*, in prep) where this requirement is reasonable.

If a method is co-Pareto it ensures that, any given relationship in the supertree is contained (displayed, included) in at least one input tree, and therefore that the relationship is compatible with at least one input tree. A weaker requirement is that each relationship is compatible with at least one input tree. A still weaker requirement is that no relationship in the supertree should be contradicted by all the input trees (in which case it also cannot be present in or entailed by any of them), or at least those input trees with leaf

sets that make contradiction a logical possibility. We know that standard, irreversible and Purvis MRP supertrees do not, and that MRC supertrees do satisfy this weakened co-Pareto axiom with respect to the kind of relationships encoded in the matrix representation and any relationships of lower cardinality (Wilkinson *et al.*, in prep.).

On the one hand, if a liberal supertree is conceived of as some sort of average or representation of central tendency of the input trees (Lapointe and Cucumel, 1997), then all contradicting relationships might be an acceptable compromise between conflicting relationships in the input trees. On the other hand, there would seem to be no good reason for a supertree to include any relationship that is contradicted by all the input trees because there is no obvious evidence for that relationship and clear counter evidence. We are concerned that relationships that contradict all the input trees are not likely to be accurate. Given that accuracy is our ultimate aim, we prefer supertrees that obey the weakened co-Pareto axiom. We consider methods that resolve conflict in favour of the best supported alternatives present or entailed by the input trees are more likely to be accurate.

4.10 Weightable. We may have reason to consider some input trees, or some relationships in some input trees as better supported than others. Indeed Purvis (1995b: 406) considered that "Because different kinds of source tree differ in their likelihood of being right, equal weighting of source trees cannot be defended (Barrett *et al.*, 1991)." An obvious and important desideratum for supertree methods is their capacity to use information on relative support for, or quality of, hypothesis, so that this information can play its part in resolving conflicts in the input trees. Weighting of input trees, or of particular relationships of input trees, can be achieved to some degree by all methods. The simple expedient of replicating input trees allows all liberal methods to use differential tree weights. MRD accommodates information on support provided by branch lengths in input trees (Lapointe and Cucumel, 1997). Matrix representation methods using discrete pseudocharacters are amenable to differential weighting of the type of relationships encoded in the matrix representation. Thus weighting schemes can be used that reflect measures of support for components or triplets, such as bootstrap proportions (Felsenstein, 1985; Wilkinson, 1996) or decay indices (Bremer, 1988; Donoghue *et al.*, 1992; Wilkinson *et al.*, 2000). Ronquist (1996) argued the virtues of differential weighting in the context of MRP and simulations suggest that weighting in this context can improve accuracy (Bininda-Emonds and Sanderson, 2001).

4.11 Assessable.   In phylogenetic analysis support is a much used surrogate for accuracy. We have more faith in well-supported relationships and we endeavour to provide indices of support for relationships in phylogenetic trees. That the support for relationships in phylogenetic

supertrees should be analogously assessable is another obvious desideratum. The quartet puzzling approach to supertree construction provides support indices directly from the variance in output trees arising from random variation in the choice of starting tree, order of addition of leaves and the breaking of ties. MRD yields supertrees in which branch lengths reflect relative support (Lapointe *et al.*, 1994; Lapointe and Cucumel, 1997; Wilkinson *et al.*, 2003).

All supertree methods can be investigated with one or more methods designed to yield indices of support and this is to be encouraged. For example, some authors have reported decay indices (Bremer support) for clades in MRP supertrees (e.g. Bininda-Emonds *et al.*, 1999; Liu *et al.*, 2001; Pisani *et al.*, 2002), although the utility of this particular support index has been questioned (Pisani *et al.*, 2002). Methods such as the bootstrap and jackknife could be used with any supertree method but a question arises as to what should be resampled. It seems natural that we would resample input trees, as suggested by Lapointe and Cucumel (2003) for assessing consensus trees. It is possible also to resample pseudocharacters in matrix representations. The latter approach was used by Purvis (1995b) who nonetheless noted that pseudocharacters derived from a single input tree are not independent. An additional potential problem with bootstrapping pseudocharacters rather than trees is that those input trees that yield more pseudocharacters (e.g. because of their large size) will tend to contribute disproportionately to the resampled data.

There are many techniques that are used to evaluate phylogenetic hypotheses inferred from primary data. We can expect that analogues of some of these will be used increasingly as the field of supertree construction matures. With methods that employ objective functions, we envisage the development and use of randomization tests of the null hypothesis that the fit of the input trees to the supertree is no better than expected by chance alone, i.e. from randomly permuted trees (Creevey *et al.*, in prep). Rejecting the null hypothesis would be a minimum requirement for supertrees to be taken seriously. Randomisation tests might also be used to identify significant outliers within sets of input trees Lapointe and Cucumel, 2003; see also Daubin *et al.*, 2002). Additional assessment of supertrees might be attained using multiple supertree methods, particularly if we have no good basis for choosing between the alternative methods, on the basis that disagreement is suggestive of weakly supported inferences. Purvis and Webster (1999) compared standard and Purvis MRP and found that the methods tend to agree, but that they disagree more as conflict in the input trees increases. Several workers have explored a range of weighting regimes to explore the robustness of real supertrees (e.g. Bininda-Emonds *et al.* 1999; Lapointe and Kirsch, 2001; Liu *et al.*, 2001; Purvis, 1995b). Alternative weighting

schemes that reverse known or suspected biases might be particularly useful when methods that are known to be biased are used.

5.   Practicality

To be at all useful, a supertree method must be practical. Generally, methods are used only when the major steps are implemented in software. There is considerable variation in the ease of implementing any of the methods at the present time, and we expect progress to be sufficiently rapid as to ensure that any discussion will be out of date. Thus we do not discuss the implementation of methods here save to repeat a previous warning that applicability in practice should not be confused with acceptability in principle (Wilkinson *et al.* 2001).

5.1 Speed. It has been stressed that supertree methods avoid difficulties of combining different data types (Bininda-Emonds *et al.*, 2002; Sanderson *et al.*, 1998), giving them a clear advantage over the alternative pathway to large trees, namely the phylogenetic analysis of combined data. They can also offer advantages in speed. It may be faster to assemble sets of input trees than to combine data, in which case this initial speed advantage is shared by all supertree methods. MinCut is a polynomial time algorithm (property P5 of Steel *et al.*, 2000) that fulfils our desire for speedy analyses. In contrast, with all matrix representation methods, computational complexity increases exponentially with the number of leaves. This necessitates the use of heuristics and even the MinCut method can be conceived of as a heuristic for finding supertrees that minimise the sum of triplet distances to the input trees, and thus closely related to triplet MRP/MRC. Note that methods that rely on heuristics to approximate the best supertrees under some objective function need not satisfy desiderata satisfied by the exact method (Steel *et al.*, 2000). Given that most matrix representation methods use the very same approaches and programmes as are used in analyses of combined data, they would appear to offer no clear benefit in terms of speed over combined analyses of data.

5.2 Generality. A reasonable desideratum of any method is that it not be restricted to special cases, particularly if those special cases are not often encountered in practice. This is not to say that a less general method would not satisfy other desiderata that make it the method of choice in a specific context. A number of supertree methods have been developed for the special case of input trees that do not conflict and cannot be applied to the more usual case of conflicting input trees (e.g. Aho *et al.*, 1981; Gordon, 1986; Steel, 1992; Thorley and Wilkinson, 2003).

Steel *et al.* (2000) have shown that some combinations of desirable properties of supertree methods can only be satisfied in the special case of

rooted trees (or of trees sharing some other leaf in common). Despite major theoretical limitations, if some or all of the potential input trees are unrooted we would still like to have methods capable of exploiting this potential. Purvis MRP cannot be applied to unrooted trees because the matrix representations encode sister-group relationships and these are only defined in rooted trees. Similarly restricted to rooted trees are methods where the objective function can be interpreted as the sum of the triplet distances between the input trees and the supertree. This includes the MinCut method which, like the Adams consensus, is defined only for rooted trees.

Quartet methods and MRD are more general in being applicable in principle to both rooted and unrooted trees. Path-length distance matrices usually represent unrooted trees but are equally applicable to rooted trees and are invariant with respect to different rootings. In contrast, ultrametric distance matrices are always associated with rooted trees. Although most supertree construction has been done using rooted input trees and standard MRP, this method is also more general. In practice, matrix representations can be constructed for unrooted trees and combined with each other, alone or with matrix representations of rooted input trees. Unlike rooted trees, there are multiple equivalent matrix representations because, in the absence of a root with a fixed (but arbitrary) pseudocharacter state code, the assignment of pseudocharacter states to the subsets defined by the split is arbitrary and can be reversed with no loss of meaning. Use of any one matrix is arbitrary. This is unimportant in the case of standard (and quartet) MRP, but with irreversible parsimony the results will depend on the arbitrary choice of matrix representation, and we consider this undesirable in principle.

6.   Discussion

The field of supertree construction is still young and would benefit from further discussion and clarification of what is expected of good liberal supertree methods and the extent to which these expectations can be satisfied. This would be a useful prelude to the identification or development of good supertree methods that can be used in practice. Here we have discussed a few properties that might be considered supertree desiderata and highlighted some examples of methods that we know or conjecture do not satisfy these desiderata. We remain largely uncertain to what extent any failure to display these properties is important in practice, something that may be addressed through empirical or simulation studies. In addition, the important work of Steel *et al.* (2000) notwithstanding, the tasks of determining thoroughly which existing methods have these or other properties, of determining the compatibility of different desiderata, and of designing novel supertree methods with particular desirable properties await.

ACKNOWLEDGEMENTS

REFERENCES

Aho, A. V., Sagiv, Y., Szymanski, T. G. and Ullman, J. D. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of relational expressions. SIAM Journal of Computing 10: 405-421.

Barrett, M., Donoghue, M. J. and Sober, E. 1991. Against consensus. Systematic Zoology 40:486-493.

Barthélemy, J.-P., McMorris, F. R. and Powers, R. C. 1995. Stability conditions for consensus functions defined on n-trees. Mathematical Computer Modeling 22: 79-87.

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. and Bryant, H. N. 1998. Properties of matrix representation with parsimony analyses. Systematic Biology 47:497-508.

Bininda-Emonds, O. R. P., Gittleman, J. L. and Purvis, A. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biological Review 74: 143-175.

Bininda-Emonds, O. R. P., Gittleman, J. L. and Steel, M. A. 2002. The (super) tree of life: procedures, problems and prospects. Annual Review of Ecology and Systematics 33:265-289.

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

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795-803.

Bryant, D. 1997. Building trees, hunting for trees and comparing trees. Ph.D.Thesis. University of Canterbury, NZ.

Bryant, D. 2002. Strict consensus supertrees. Technical Report, School of Computer Science, McGill University, Canada.

Bryant, D. 2003. A classification of consensus methods for phylogenies. In M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts (eds), Bioconsensus. DIMACS series in discrete mathematics and theoretical computer science, pp. 163-183. American Mathematical Society, Providence.

Chen, D., Diao, L., Eulenstein, O., Fenández-Baca, D. and Sanderson, M. J. 2003. Flipping: A supertree construction method. In M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts (eds), Bioconsensus. DIMACS series in discrete mathematics and theoretical computer science, pp. 135-160. American Mathematical Society, Providence.

DasGupta, B., He, X., Jiang, T., Li, M., Tromp, J. and Zhang, L. 1997. On distances between phylogenetic trees. Proceedings of the 8th Annual ACM SIAM Symposium on Discrete Algorithms, pp. 427-436**.**

Daubin, V., Gouy, M. and Perriere, G. 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Research 12:1080-90.

Dekker, M. C. H. 1986. Reconstruction methods for derivation trees. Master's Thesis, Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam.

Donoghue, M. J., Olmstead, R. G., Smith, J. F. and Palmer, J. D. 1992. Phylogenetic relationships of Dipsacales based on rbcL sequences. Annals of the Missouri Botanical Gardens 79:333-345.

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

Farris, J. S., Albert, V. A., Kallersjo, M., Lipscomb, D. and Kluge, A. G. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99-124.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Finden, C. R. and Gordon, A. D. 1985. Obtaining common pruned trees. Journal of Classification 2: 225-276.

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

Gordon, A. D. 1986. Consensus supertrees: the synthesis of rooted trees containing overlapping sets of labeled leaves. Journal of Classification 3: 31-39.

Keynes, J. M. 1920. A treatise on probability. MacMillan.

Lanyon, S. M. 1993. Phylogenetic frameworks: towards a firmer foundation for the comparative approach. Biological Journal of the Linnean Society 49:45-61.

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

Lapointe, F.-J. and Cucumel, G. 2002. Multiple consensus trees. Pages 359-364 *in* Classification, Clustering and Data Analysis: Recent Advances and Applications (K. Jajuga, A. Sokolowski, and H.-H. Bock , eds), Springer-Verlag, Berlin.

Lapointe, F.-J. and Cucumel, G. 2003. How good can a consensus get? Assessing the reliability of consensus trees in phylogenetic studies. In M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts (eds), Bioconsensus. DIMACS series in discrete mathematics and theoretical computer science, pp. 221-230. American Mathematical Society, Providence.

Lapointe, F.-J. and Kirsch, J. A. W. 2001. Construction and verification of a large phylogeny of marsupials. Australian Mammalogy 3:9-22.

Lapointe, F.-J., Kirsch, J. A. W. and Bleiweiss, R. 1994. Jackknifing of weighted trees: validation of phylogenies reconstructed from distance matrices. Molecular Phylogenetics and Evolution 3:256-267.

Larget, B. and Simon, D. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution 16:750-759.

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

Mann, C. 1990. Meta-analysis in the breech. Science 249: 476-479.

McMorris, F. R. and Neumann, D. 1983. Consensus functions defined on trees. Mathematical Social Sciences 4:131-136.

Nixon, K. C. and Carpenter, J. M. 1996. On consensus, collapsability and clade concordance. Cladistics 12:305-201.

Page, R. D. M. Modified mincut supertrees. In R. Guigó and D. Gusfield (eds), Workshop in Algorithms and Bioinformatics (WABI 2002), pp. 537-551. Lecture Notes in Computer Science 2452.

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

Pisani, D. and Wilkinson, M. 2002. MRP, taxonomic congruence and total evidence. Systematic Biology 51:151-155.

Pisani, D., Yates, A. M., Langer, M. C. & Benton, M. J. 2002. A genus-level supertree of the Dinosauria. Proceedings of the Royal Society of London, Series B. 269:915-921.

Purvis, A. 1995a. A modification to Baum and Ragan's method for combining phylogenetic trees. Systematic Biology 44: 251-255.

Purvis, A. 1995b A composite estimate of primate phylogeny. Philosophical Transactions of the Royal Society of London Series B. 348:405-421.

Purvis, A. and Webster, A. J. 1999. Phylogenetically independent comparisons and primate phylogeny. In P. C. Lee (ed.) Comparative primate socioecology, pp. 44-70. Cambridge University Press, Cambridge.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. Molecular Phylogenetics and Evolution 1:53-58.

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

Ronquist, F. 1996. Matrix representation of trees, redundancy, and weighting. Systematic Biology 45:247-253.

Sanderson, M. J., Purvis, A. and Henze, C. 1998 Phylogenetic supertrees: assembling the trees of life. Trends in Ecology and Evolution 13:105-109.

Semple, C. and Steel, M. 2000. A supertree method for rooted trees. Discrete Applied Mathematics 105:147-158.

Semple, C. and M. Steel. 2002. Tree reconstruction from multistate characters. Advances in Applied Mathematics, 28:169-184.

Shao, K. and Sokal, R. R. 1990. Tree balance. Systematic Zoology 39: 266-276.

Steel, M., Dress, A. W. M. and Böcker, S. 2000. Simple but fundamental limitations on supertree and consensus tree methods. Systematic Biology 49:363-368.

Strimmer, K. and von Haeseler, A. 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Molecular Biology and Evolution 13:964-969.

Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295-333 *in* Phylogenetic analyses of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford University Press, New York.

Thorley, J. L. 2000. Cladistic information. leaf stability and supertree construction. Ph.D. Thesis, School of Biological Sciences, University of Bristol, UK.

Thorley, J. L., and Wilkinson, M. 2003. A view of supertree methods. In M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts (eds), Bioconsensus. DIMACS series in discrete mathematics and theoretical computer science, pp. 1885-189. American Mathematical Society, Providence.

Wilkinson, M. 1994. Common cladistic information and its consensus representation: reduced Adams and reduced cladistic consensus trees and profiles. Systematic Biology 43: 343-368.

Wilkinson, M. 1996. Majority-rule reduced consensus methods and their use in bootstrapping. Molecular Biology and Evolution 13:437-444.

Wilkinson, M., Lapointe, F.-J. and Gower, D. J. 2003. Branch lengths and support. Systematic Biology 52:127-130.

Wilkinson, M., and Thorley, J.L. 2003. Reduced consensus. In M. Janowitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin, and F. S. Roberts (eds), Bioconsensus. DIMACS series in discrete mathematics and theoretical computer science, pp. 195-203. American Mathematical Society, Providence.

Wilkinson, M., Thorley, J. L., Littlewood, D. T. J. and Bray, R. A. 2001. Towards a Phylogenetic Supertree for the Platyhelminthes? In D. T. J. Littlewood R. A. Bray (eds.), Interrelationships of the Platyhelminthes, pp. 292-301. Chapman-Hall, London.

Willkinson, M., Thorley, J. L., and Upchurch, P. M. 2000. A chain is no stronger than its weakest link: Double decay analyses of phylogenetic hypotheses. Systematic Biology 49: 754-776.