

Discriminating Supported and Unsupported Relationships in Supertrees using Triplets

James A. Cotton*¹, Claire S. C. Slater^{1,2}, and Mark Wilkinson¹

September 29, 2005

¹ Department of Zoology, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK.

² Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EQ, UK.

*Author for Correspondence. Tel: +44 (0)20 7942 5306. Fax: +44 (0)20 7942 5054. james.cotton@nhm.ac.uk

Running Head: Points of View

Keywords: triplets, closure, closed set, support measure, supertree.

A supertree is a phylogeny formed by combining information from disparate phylogenetic trees. Supertree methods have been particularly used for constructing large phylogenies from previously published trees and there is a growing literature using supertree methods for phylogenetic inference in macro-evolutionary studies (e.g. Davies et al., 2004; Grotkopp et al., 2004; Salamin and Davies, 2004). This empirical supertree work has mostly used matrix representation with parsimony (standard MRP; Baum, 1992; Ragan, 1992) in which optimal supertrees are found by parsimony analysis of a matrix encoding the full splits of the input trees. A major concern with standard MRP (and some other) supertrees is that they can display relationships that seem to lack evidential support from the input trees, either individually or jointly (Bininda-Emonds and Bryant, 1998; Pisani and Wilkinson, 2002; Wilkinson et al., 2004b). This has prompted the development of measures of support that attempt to distinguish supported and unsupported relationships in supertrees (Bininda-Emonds, 2003; Wilkinson et al., 2005b).

A standard means of analysing the distribution of support across a phylogenetic hypothesis is to deconstruct a tree into the less complex relationships that the tree entails. For example, bootstrap proportions are typically reported for a set of full splits (clades on rooted trees) on the taxa of interest, and a number of other measures on trees are focused at identifying clade-based support (e.g. Bremer, 1994; Larget and Simon, 1999). In the supertree context, clades must be supported by input trees rather than by characters, and a supertree and the input trees generally have different leaf sets, so that a supertree clade may not be displayed by any input tree. This has left scope for ambiguity as to how to identify and quantify support in the supertree context (Bininda-Emonds, 2003; Wilkinson et al., 2005b). One solution is to seek some kind of soft, or reduced support, in which input clades that are compatible with or entailed by supertree clades are seen as providing some level of support for these supertree relationships.

Previous work has two important limitations. It focuses only on support (or lack of support) for supertree clades (components, full splits), ignoring support for less inclusive relationships like partial splits, triplets or nestings (Wilkinson, 1994). We show that a more sensitive measure of support, focusing on lower-level relationships, may give a different picture of which supertree relationships are supported and unsupported. In fact, a supertree clade can appear unsupported despite all the triplets it implies being supported. As noted by Wilkinson et al. (2005b), “input trees may jointly entail, and thus strictly support, novel relationships that are not strictly supported by any single input tree”. The second limitation of previous work is that it relies on pairwise comparisons between each input tree and the supertree and consequently does not fully account for support jointly entailed by combinations of input trees. As the primary use of supertree methods is to combine information from a set of input trees, being able to identify this kind of support seems particularly important. Some authors have claimed total-evidence-like properties of signal enhancement for supertree methods (Bininda-Emonds et al., 1999), but novel relationships displayed by a supertree (relationships not present on any of the input trees) are

worrying if they are not implied by combinations of the input trees (see Pisani and Wilkinson, 2002).

Focusing exclusively on a rooted supertree and rooted input trees, we present a method for examining the support for triplets in a rooted supertree which can be naturally extended to identify combined support for supertree relationships, based on inference rules for triplets. We show that considering this combined support can reveal support for additional supertree relationships and so better diagnose unsupported relationships.

Triplet Support

We define a triplet as a rooted binary tree with three leaves, such as the input trees (AB)C, (CD)B and (BC)D shown in figure 4. Following the notation of Bryant and Steel (1995), if R is a set of triplets, then the span of R , denoted $\langle R \rangle$, is the set of rooted trees on all the leaves of R that display all the triplets in R . A tree is defined as compatible with a triplet if the subtree induced by the leaves of the triplet is either unresolved or identical to the triplet, so that compatibility is defined such that polytomies represent ignorance rather than actual simultaneous divergence (are soft rather than hard in the sense of Maddison, 1989). $r(T)$ is the set of rooted triplets that are induced subtrees of rooted tree T , i.e. the set of triplets displayed by T .

The concept of support and conflict for triplets in a supertree is a very simple one. If the supertree triplet is displayed by the input tree, then it supports this supertree relationship. By focusing on the smallest unit of cladistic information, the ambiguous support that clade-based measures need to deal with is avoided, allowing a simple classification of supertree relationships at this level equivalent to that suggested by Wilkinson et al. (2005b). More formally, let $\{T_1 \dots T_n\}$ be an n -tuple of rooted input trees, and let S be a rooted supertree. For a particular input tree T_i , we can find the triplets supported by that tree as $r(S) \cap r(T_i)$, so that the set of triplets supported by the input trees is:

$$r(S)_{supp} = \bigcup_{i=1 \dots n} (r(S) \cap r(T_i))$$

Any triplet in this set is matched by a triplet in at least one input tree, and so receives at least some support, while triplets outside this set are not present in the input, and so are novel and potentially worrying. The size of this set is a measure of the extent to which relationships in a supertree are supported by at least one input tree, while $\frac{|r(S)_{supp}|}{|r(S)|}$ is the relative proportion of such supported relationships.

This triplet-based assessment of support can give us a very different perspective to clade-based measures. The (A,B) clade in the supertree shown on figure 1 is unsupported by any input tree clade, and is contradicted by the third input tree. Previous clade-based measures reflect this lack of support. However, all the 10 triplets in the supertree are present in the set $r(S)_{supp}$, and so the tree is fully supported in terms of triplets.

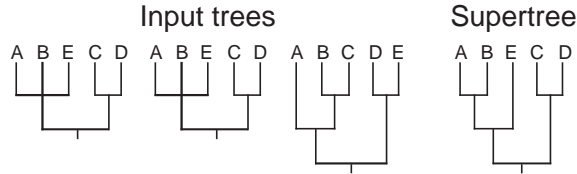


Figure 1: The supertree shown is the standard MRP supertree for the three input trees. The supertree clade (A,B) receives no support under either of the clade-based support measures that have been proposed (Bininda-Emonds, 2003; Wilkinson et al., 2005b) but is fully supported by the triplets in the three input trees.

A Closure Operation and Combined Support

A set of triplets are compatible if they can all be displayed on a single tree, and any set of compatible triplets implies one or more (perhaps non-binary) trees. For example, the triplets (AB)C and (AC)D together imply the tree (((AB)C)D), as this is the only tree that displays both triplets. This tree, however, implies a larger set of triplets – both (AB)D and (BC)D as well as the two original triplets – which are displayed on it. We can write that (AB)C+(AC)D→(AB)D and (AB)C+(AC)D→(BC)D, where the arrow stands for logical implication. These are two examples of *inference rules* in which sets of triplets logically imply another triplet. These rules, introduced by Dekker (1986) and further analysed by Bryant and Steel (1995) (see also Wilkinson et al., 2004a, for a simple introduction), lie at the heart of combining information from different phylogenetic trees. While both of these rules are *dyadic* in that they only involve inference from pairs of triplets, it is known that irreducible rules of any order exist (Bryant and Steel, 1995). Novel triplets not implied by smaller sets can potentially be inferred from sets of triplets of any size. By performing these inference rules exhaustively we could find all of the triplets jointly implied by a set of supertree triplets, called the closure of the set of triplets under the inference rules. This closure would thus identify all the relationships implied by combinations of triplets from different input trees, and so implied by combinations of the trees themselves. The power of these inference rules is illustrated by the fact that a phylogenetic tree on n taxa can be defined by just $n - 2$ of the $\frac{n(n-1)(n-2)}{6}$ triplets it implies (Steel, 1992). In all but the simplest cases, it would not be feasible to find the closure by enumerating and using the rules directly – indeed, most higher-order rules remain unknown, and there has been no systematic attempt to enumerate them. In fact, this set is just those triplets common to all the trees in the span of a set of triplets. Given R , a set of compatible triplets, the closure of R , is:

$$\overline{R} = \bigcap_{A \in \langle R \rangle} r(A)$$

Bryant and Steel (1995) present a polynomial-time algorithm for finding this closure. For a rooted supertree S the triplets in $r(S)_{supp}$ are a subset of $r(S)$, they are all displayed by S , and so are compatible, so the closure $\overline{r(S)_{supp}}$ exists and can be found using this algorithm. This set is then the set of triplets on supertree S that are supported by any of the input trees alone or in combination, and so identify which novel triplets in S are the sort of inferences we might welcome and those we might not.

As well as identifying relationships supported (entailed) by combinations of input trees, the closure defined above has some helpful properties, all of which are results of Bryant and Steel (1995). For any rooted tree T , the set $r(T)$ is closed, and the intersection between any two closed sets is also closed. The triplets supported by a single input tree thus cannot imply any additional supported triplets and the triplets in $\overline{r(S)_{supp}} - r(S)_{supp}$ are all implied by combinations of input trees. As $r(S)_{supp} \subseteq r(S)$, we know that $\overline{r(S)_{supp}} \subseteq r(S)$, so that the closure will only contain triplets present in the supertree. Lastly, if $r(S)_{supp}$ defines a supertree – i.e. the supertree is the only tree that displays this set – then the span of these triplets will consist of a single tree, and the closure will include all of the triplets in the supertree so that the supertree is fully supported.

Fig. 2 shows an example of Gordon (1986), where there is no conflict between the two input trees (see e.g. Wilkinson et al., 2005b). As shown on figure 3, 187 out of 348 supertree triplets are displayed by one or other of the input trees, but all are supported when the information from the two input trees is combined using the closure described above. The 161 triplets included in the closed set but not supported by the two input trees individually (figure 3) particularly add support for the basal split between taxon 9 and the other taxa (figure 2), as might be expected from examining the input trees – input tree 1 shows that taxa 1..8 form a monophyletic group excluding taxon 9, while input tree 2 places the additional taxa 10..14 as part of this group. The interpretation of all relationships in the supertree as being supported is consistent with that based on one of the proposed clade-based measures (Wilkinson et al., 2005b).

We should note that, while we consider a closure operation on sets of triplets, trees can be broken down in a number of different ways (see e.g. Wilkinson, 1994) and similar closures can be defined analogously on sets of quartets (see Bryant and Steel, 1995) and on other partial splits (Meacham, 1983, n-taxon statements) as the sets of splits/quartets common to all members of a span. There is no convenient algorithm for finding either the full split closure or full quartet closure. While we have chosen to focus on rooted supertrees and rooted input trees, quartet-based support measures using the quartet closure operation would be more natural for the unrooted case. As most systematists are more used to thinking about full splits than the lower-order relationships like triplets and quartets, a closure operation on partial splits seems likely to be of interest. The only such closure for which a polynomial-time algorithm is available is a dyadic split closure introduced by Meacham (1983) and formalised by Semple and Steel (2001). There are close links between these different closure operations

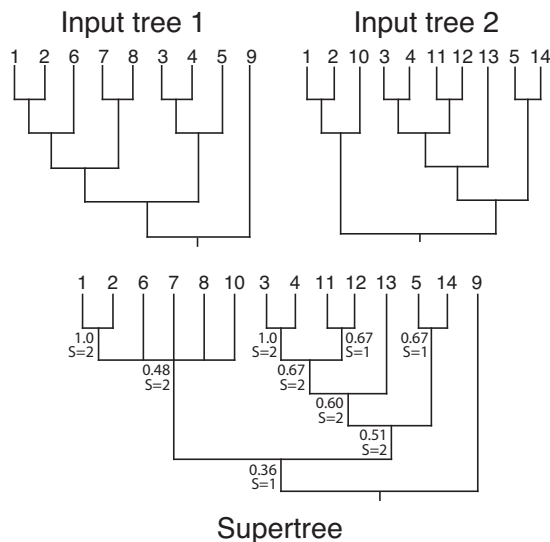


Figure 2: Trees from Gordon (1986). Upper values on nodes are the ratio of the number of triplets supporting the clade to the number supporting the clade if all triplets from the tree were present. Lower values are S , the number of input trees supporting each clade (as defined by Wilkinson et al., 2005b). The closure of this set contains all of the triplets on the supertree (i.e. each value would be 1.0 for the closure), showing that the supertree is fully supported by the input trees – in fact, the input trees are compatible and the supertree is the strict consensus of their span.

– informally, inference rules on partitions can be reduced to those on quartets, and inference rules on rooted splits can be reduced to triplet rules. In fact, there is a particularly close link between the dyadic split closure and a simpler closure operation on quartets, called the semi-dyadic closure (see Huber et al., 2005; Semple and Steel, 2003, pp.128-130), for which a polynomial-time algorithm exists. This semi-dyadic closure is the closure under a subset of the inference rules discussed above (in fact, under a single dyadic rule on quartets), while the closure operation we discuss is the rooted analogue of the full closure operation on quartets. If we convert a set of triplets into quartets by adding a shared outgroup taxon, this closure on quartets (and the dyadic split closure) contains a proper subset of the triplets in $r(S)_{supp}$, and the triplet closure we use is thus more complete. If we restrict our attention to rooted trees and triplets, we can use the most complete closure for which a polynomial-time algorithm is available.

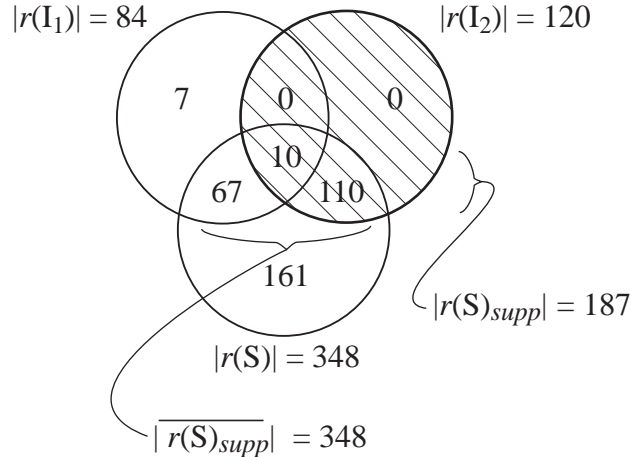


Figure 3: The example from Gordon (1986) showing triplets in the two input trees (I_1 and I_2 are input trees 1 and 2 respectively) and the supertree (S) shown on figure 2. Figures in the Venn diagram show the numbers of triplets in each category, and in the closure of the intersection sets with the supertree. In the absence of any conflict, the intersection of the set of triplets from each input tree with the supertree set defines the input tree, and so is closed.

Measures of Triplet Support

The sets $r(S)_{supp}$ and $\overline{r(S)_{supp}}$ distinguish supported and unsupported supertree triplets, most biologists are more used to interpreting measures of support for particular clades, which can be conveniently shown as values next to particular branches in a tree diagram. While we focus on support for supertree triplets, the relationship between triplet support and clade support is relatively straightforward. Each clade on the supertree is supported by those triplets it entails, while it is contradicted by any triplet that conflicts with a triplet it entails. A single triplet can thus contradict or support a number of supertree clades but cannot both support and contradict clades on a single tree. We can thus show triplet support for particular supertree clades as the fraction of triplets implied by a particular clade that are supported. However, larger clades will imply many more triplets than less-inclusive clades, so it may be harder to achieve a certain level of support for such clades, and these support values might be expected to decrease towards the root of a tree, as on figure 2.

While both the examples we have discussed show support for all the supertree triplets (figures 1 and 2) this is not always the case for MRP supertrees. Table 1 shows some examples of the number and proportion of supported triplets in some empirical MRP supertrees used as examples in previous work (Bininda-Emonds, 2003; Wilkinson et al., 2005b). It is notable that the closure operation identifies relatively few jointly entailed triplets for these examples, because they

	$ r(S) $	$ r(S)_{supp} $	$ \overline{r(S)_{supp}} $	$\frac{ r(S)_{supp} }{ r(S) }$	$\frac{ \overline{r(S)_{supp}} }{ r(S) }$	QS _{tree}	mean V
Lagomorpha	79458	77852	77899	0.979	0.980	-0.104	0.340
Mustelidae	11753	11525	11699	0.981	0.995	-0.143	0.521
Canidae	5510	5382	5439	0.977	0.987	-0.146	0.259
Viverridae	5983	5972	5979	0.998	0.999	-0.045	0.253
Carnivora	220	220	220	1.0	1.0	-0.029	0.199
Felidae	7139	6807	7092	0.953	0.993	-0.219	0.022

Table 1: Numbers and proportions of supported triplets for some empirical supertrees, Lagomorpha example is the unweighted analysis from Stoner et al. (2003), others are from Bininda-Emonds et al. (1999). QS_{tree} and mean V are component-based measures of support defined in Bininda-Emonds et al. (1999) and Wilkinson et al. (2005b) respectively.

have relatively large numbers of input trees that already support the majority of input tree triplets. For many other empirical examples, the closure will be too complex to calculate using existing algorithms. For example, in the data of Kennedy and Page (2002) on 121 seabird species, 114,928 triplets are supported by trees individually out of 284,331 triplets on the entire MRP strict consensus tree, and computing the closure of 114,928 triplets is currently impractical. Many published supertrees are considerably larger than this. Faster algorithms for finding the closure would help. The Bryant and Steel (1995) algorithm takes time $O(n^5)$, where n is the number of triplets. We can improve the running time somewhat by using the OneTree (Ng and Wormald, 1996) algorithm (or the algorithm of Aho et al., 1981, it is derived from) to reduce the number of triplets we need to look at (Thorley, 2000): as the OneTree tree will be a member of the span of the triplets we only need to check if triplets in the OneTree are part of the closure. Unfortunately, the OneTree is also the Adam’s consensus of the span (theorem 2.10 of Bryant, 2003), and can be very poorly resolved, so this does not improve the worst-case running time of Bryant and Steel’s algorithm. Better algorithms, based on ideas behind the OneTree algorithm, are probably possible for this problem, but no work has been done on this issue.

Inferring Triplet Supertrees

Like previous measures, $r(S)_{supp}$ and $\overline{r(S)_{supp}}$ are not designed for use as objective functions for choosing between alternative supertrees as they focus on identifying supported relationships rather than quantifying support. Importantly, our measure ignores triplets shared by multiple input trees, giving them the same weight as triplets that occur only singly. For example, there are two other trees also fully supported by the input tree triplets shown in figure 1, with both ((AB)C)(DE)) and ((A(BC)(DE)) having all 10 of the 10 triplets from

the input trees. The supertree shown in the figure is optimal under the MRP criterion because the six triplets relating A,B and E to the exclusion of C and D are repeated in the first two input trees. The measures discussed above also fail to penalise lack of resolution in a supertree, so that a tree with just a single branch (and so a single clade) can score as well as a fully resolved tree. These measures also make no attempt to quantify the degree of conflict between a supertree and input.

The act of decomposing input trees into sets of triplets does, of course, suggest a number of natural triplet-based criteria for choosing between supertrees, by finding the largest set of compatible triplets from the input tree triplet sets (counting triplets occurring in multiple input trees multiple times). A supertree could then be found using the (polynomial time) OneTree algorithm (Ng and Wormald, 1996) on this set. There are problems with implementing this idea. While checking the compatibility of a set of triplets is possible in polynomial time, identifying a maximal set of mutually compatible triplets is NP-complete (Bryant, 1997, p. 40, see Garey and Johnson, 1979 for a description of NP-completeness). Heuristic methods can be used to search through tree space for supertrees that display the largest compatible subset of a set of triplets. This would be identical to using either parsimony or compatibility to analyse a matrix encoding triplets, an approach suggested several times in the supertree literature (Nelson and Ladiges, 1994; Wilkinson et al., 2001) and recently investigated by Wilkinson et al. (2005a) as ‘triplet fit’. Such an approach appears to have a number of desirable properties when compared to some other MRP methods (Wilkinson et al., 2004b, 2005a). A different method would be to find the tree matching the largest (or some maximal weighted) closed set of input tree triplets, which will be considerably more computationally intense given the slow algorithm available to compute the closure, and is probably impractical.

The Meaning of “Supported” relationships

What do we mean when we say a particular relationship is “supported” by a particular tree? The set $r(S)_{supp}$ contains all of the triplets that are entailed by the supertree and at least one of the input trees, or by combinations of these triplets, and thus seems a natural set of “supported” relationships. There are, however, other supertree triplets that could be seen as supported by the input trees that are not in this set. Triplets from different input trees that conflict with the supertree can, in combination, imply triplets that are displayed, so these triplets together might be seen as supporting this particular supertree relationship, which is not a member of $r(S)_{supp}$ (figure 4). We suggest that this kind of support is ignored because it emerges from combinations of triplets that are rejected by the supertree under consideration. Something similar happens if we attempt to characterise conflict with the supertree using the closure operation – relationships in the supertree will be supported and conflicted by various combinations of input tree relationships, and every compatible subset of input tree triplets will have its own closure implying different additional relationships.

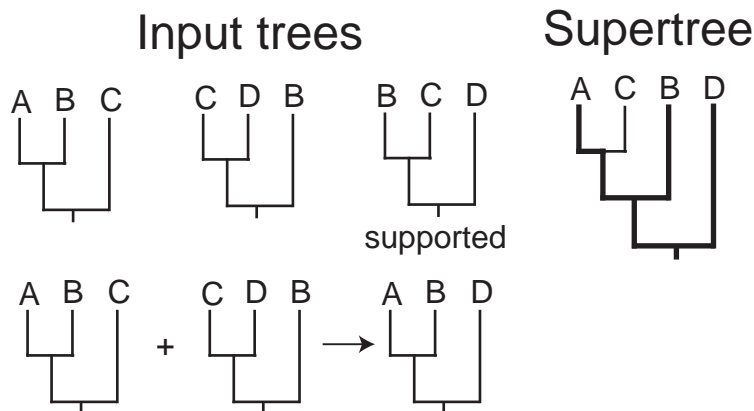


Figure 4: Only one triplet from the supertree is supported by the input trees. The other (conflicting) triplets together imply (and so, to one view, support) another triplet from the supertree. These two input trees can thus be seen as together supporting part of the supertree despite both conflicting with it.

The example of figure 1 raises more fundamental questions about what we mean by relationships being supported. The split between AB and CDE in this supertree is not supported by any input tree *clades* but is fully supported by input tree *triplets*. In the sense of Wilkinson et al. (2005b), this would be considered an unsupported group, as it is contradicted by one of the input trees, and seen as a problem of MRP analysis, but actually MRP is doing something rather desirable here. By breaking up components as it maps them onto the optimal tree, MRP is finding a tree that matches well on the triplet level. Parsimony on the component matrix is, in this example, acting like a compatibility (or parsimony) analysis of triplets, but this property will not hold in general. Compatibility on the component matrix doesn't have this property, simply identifying the three resolutions of the duplicated input tree as the different input trees share no clades.

Our analysis of the support for supertree relationships is based on decomposing both input trees and supertree into triplets, allowing us to discriminate between support for supertree relationships that comes from combinations of input trees and support from any individual input tree. We show that this triplet-based view itself can give a very different picture of the support for a rooted supertree and that the additional combined support can be particularly important. Together, these additional levels of analysis bring into focus different concepts of support and conflict for phylogenetic hypotheses, and show that more work is needed before we have a good understanding of how different supertree methods perform or of the properties we should expect of such methods. Our example at least shows that there is no single correct view of support for supertrees, just as there is no single correct method for inferring supertrees in the first place.

Acknowledgements

This work was funded by BBSRC grant 40/G18385. We thank Claire Pickthall for proofreading the manuscript and Rod Page, Mike Steel and two anonymous reviewers for comments that improved the finished paper. Tripfuse, a software package to investigate triplet inference rules, which implements the methods described here, is available from <http://taxonomy.zoology.gla.ac.uk/~jcotton/software.html>.

References

- Aho, A. V., Y. Sagiv, T. G. Szymanski, and J. Ullman. 1981. Inferring a tree from lowest common ancestors with an application to the optimization of conditional expressions. *SIAM Journal of Computing* 10:405–421.
- Baum, B. R. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3–10.
- Bininda-Emonds, O. R. P. 2003. Novel versus unsupported clades: assessing the qualitative support for clades in MRP supertrees. *Systematic Biology* 52:839–848.
- Bininda-Emonds, O. R. P. and H. N. Bryant. 1998. Properties of matrix representation with parsimony analysis. *Systematic Biology* 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). *Biological Reviews of the Cambridge Philosophical Society* 74:143–175.
- Bremer, K. 1994. Branch support and tree stability. *Cladistics* 10:295–304.
- Bryant, D. 1997. *Building Trees, Hunting for Trees, and Comparing Trees*. Ph.D. thesis University of Canterbury.
- Bryant, D. 2003. A classification of consensus methods for phylogenetics. Pages 163–185 *in* *Bioconsensus* (M. Janowitz, F.-J. Lapointe, F. McMorris, B. Mirkin, and F. Roberts, eds.). American Mathematical Society, Providence, RI.
- Bryant, D. and M. Steel. 1995. Extension operations on sets of leaf-labelled trees. *Advances in Applied Mathematics* 16:425–453.
- Davies, T. J., T. G. Barraclough, M. W. Chase, P. S. Soltis, D. E. Soltis, and V. Savolainen. 2004. Darwin’s abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences USA* 101:1904–1909.

- Dekker, M. 1986. Reconstruction Methods for Derivation Trees. Master's thesis Department of Mathematics and Computer Science, Vrije Universiteit, Amsterdam.
- Garey, M. R. and D. S. Johnson. 1979. Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman.
- Gordon, A. 1986. Consensus supertrees: the synthesis of rooted trees containing overlapping sets of labeled leaves. *Journal of Classification* 3:335–348.
- Grotkopp, E., M. Rejmánek, M. J. Sanderson, and T. L. Rost. 2004. Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution* 58:1705–1729.
- Huber, K. T., V. Moulton, C. Semple, and M. Steel. 2005. Recovering a phylogenetic tree using pairwise closure operations. *Applied Mathematics Letters* 18:361–366.
- Kennedy, M. and R. D. M. Page. 2002. Seabird supertrees: combining partial estimates of procellariiform phylogeny. *Auk* 119:88–108.
- Larget, B. and D. L. Simon. 1999. Markov Chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. *Molecular Biology and Evolution* 16:750–759.
- Maddison, W. P. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5:365–377.
- Meacham, C. A. 1983. Theoretical and computational considerations of the compatibility of quantitative taxonomic characters. Pages 304–314 *in* Numerical Taxonomy (J. Felsenstein, ed.). Springer-Verlag, Berlin.
- Nelson, G. and P. Ladiges. 1994. Three-item consensus: Empirical test of fractional weighting. Pages 193–209 *in* Models in Phylogeny Reconstruction (R. W. Scotland, D. J. Seibert, and D. M. Williams, eds.). Clarendon, Oxford.
- Ng, M. P. and N. C. Wormald. 1996. Reconstruction of rooted trees from subtrees. *Discrete Applied Mathematics* 69:19–31.
- Pisani, D. and M. Wilkinson. 2002. Matrix Representation with Parsimony, taxonomic congruence and total evidence. *Systematic Biology* 51:151–155.
- Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetics and Evolution* 1:53–58.
- Salamín, N. and T. J. Davies. 2004. Using supertrees to investigate species richness in grasses and flowering plants. Pages 461–487 *in* Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

- Semple, C. and M. Steel. 2001. Tree reconstruction via a closure operation on partial splits. Pages 126–134 *in* Proceedings of Journées Ouvertes: Biologie, Informatique et Mathématiques (O. Gascuel and M.-F. Sagot, eds.) Lecture Notes in Computer Science. Springer-Verlag, Berlin.
- Semple, C. and M. Steel. 2003. Phylogenetics. Oxford University Press, Oxford.
- Steel, M. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. *Journal of Classification* 9:91–116.
- Stoner, C. J., O. R. P. Bininda-Emonds, and T. M. Caro. 2003. The adaptive significance of coloration in Lagomorphs. *Biological Journal of the Linnean Society* 79:309–328.
- Thorley, J. L. 2000. Cladistic information, Leaf stability and Supertree construction. Ph.D. thesis University of Bristol.
- 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., J. A. Cotton, C. Creevey, O. Eulenstein, S. R. Harris, F.-J. Lapointe, C. Levasseur, J. O. McInerney, D. Pisani, and J. L. Thorley. 2005a. The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Systematic Biology* 54:419–432.
- Wilkinson, M., J. A. Cotton, and J. L. Thorley. 2004a. The information content of trees and their matrix representations. *Systematic Biology* 53:989–1001.
- Wilkinson, M., D. Pisani, J. A. Cotton, and I. Corfe. 2005b. Measuring support and finding unsupported relationships in supertrees. *Systematic Biology* In Press.
- Wilkinson, M., J. L. Thorley, D. T. J. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree of the platyhelminthes? Pages 292–301 *in* Interrelationships of the Platyhelminthes (D. T. J. Littlewood and R. A. Bray, eds.). Chapman-Hall, London.
- Wilkinson, M., J. L. Thorley, D. Pisani, F.-J. Lapointe, and J. O. McInerney. 2004b. Some desiderata for liberal supertrees. Pages 227–246 *in* Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.