Syst. Biol. 53(6):989–1001, 2004
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150490522737

The Information Content of Trees and Their Matrix Representations

MARK WILKINSON,[1] JAMES A. COTTON,[1] AND JOSEPH L. THORLEY[2]

[1]*Department of Zoology, Natural History Museum, London SW7 5BD, United Kingdom;*
E-mail: james.cotton@nhm.ac.uk (J.A.C.)
[2]*Fisheries Research Services, Freshwater Laboratory, Faskally, Pitlochry, Perthshire PH16 5LB, United Kingdom*

Any tree can be represented by a variety of matrices from which the tree can be reconstructed. Such matrix representations (encodings) of trees have been used in a number of contexts, including measuring tree similarity (Farris, 1973), cophylogeny (Brooks, 1981), and consensus (Nelson and Ladiges, 1994b; Phillips and Warnow, 1996), but they have become more prominent through the matrix representation with parsimony (MRP) approach to supertree construction (Baum, 1992; Ragan, 1992; Sanderson et al., 1998). Most published supertrees have utilized MRP (e.g., Jones et al., 2002; Kennedy and Page, 2002; Pisani et al., 2002; Purvis, 1995a), in which matrix representations of trees are combined into a composite matrix and analyzed with parsimony. In standard MRP, the elements of matrix representations are the relationships given by the full splits or bipartitions of leaves (terminal taxa) present in the trees. In the context of rooted trees, standard matrix representations have one binary (pseudo)character for each clade in which clade members are scored conventionally (and arbitrarily) as '1' and nonmembers and the root (MRP outgroup) as '0' (Table 1).

Purvis (1995b) developed an alternative matrix representation (Purvis coding) intended to compensate for apparent overweighting of larger trees in standard MRP. His method, which is applicable only to rooted trees, attempted to remove redundant information in standard matrix representations, but was criticized by Ronquist (1996) who argued that the information removed was not redundant. Williams and Humphries (2003) advocated the use of matrix representations based on three-item statements or triplets (Nelson and Ladiges, 1992, 1994b; Nelson and Platnick, 1991; Wilkinson et al., 2001), which are also applicable only to rooted trees, and they used such matrices, and an associated differential weighting, as a yardstick to compare and contrast standard and Purvis matrix representations. Matrix representations can also be based on pairwise pathlength distances between leaves (Lapointe and Cucumel, 1997; Lapointe et al., 2003) and nonbinary discrete characters (Semple and Steel, 2002).

We believe the literature on the relative merits of alternative binary matrix representations of trees to be confused and confusing. Here we use a quantitative measure, cladistic information content (CIC; Thorley et al., 1998), to compare the information content of trees and their matrix representations. Specifically, we revisit Ronquist's (1996) critique of Purvis coding. We demonstrate that the information removed by this method is redundant in the sense that Purvis (1995b) intended, but that it does not remove all such redundant information as Purvis supposed. Building on Ronquist (1996), we discuss the distinction between the representation of a tree and the representation of the data from which a tree has been inferred and its importance in choosing a matrix representation. This leads us to consider logical relations and dependencies among pseudocharacters from the same tree, and highlight fundamental inference rules that form the basis of the logical calculus of cladistic relationships. We show that fractional weighting as proposed by Nelson and Ladiges (1992), and the form of this used by Williams and Humphries (2003), does not account for all logical dependencies between triplets. This leads us to propose an alternative weighting scheme. Finally, we review the results and present a critical discussion of Williams and Humphries' (2003) comparison of alternative methods. We use the terms triplet and three-item

TABLE 1. Alternative matrix representations for the two trees in Figure 1. A, B, C, and D are the taxa, and R the root or MRP outgroup. The MR column indicates which matrix representations the particular characters belong to: s = standard, p = Purvis, t = triplet, and m1 and m2 = minimal matrix representations. Ch = character number, T = tree number, CIC = cladistic information content in bits.

Ch	T	MR	A	B	C	D	R	CIC
1	1	s, p, m1, m2, t	1	0	?	1	0	1.58
2	2	s	1	1	0	0	0	2.32
3	2	s, p	1	1	1	0	0	2.32
4	2	p, m1, m2, t	1	1	0	?	0	1.58
5	2	m1, t	1	?	1	0	0	1.58
6	2	m2, t	?	1	1	0	0	1.58
7	2	t	1	1	?	0	0	1.58

statement synonymously and interchangeably throughout. Triplet or three-item statement encodings of phylogenetic trees can be constructed using RadCon (Thorley and Page, 2000), which was also used to calculate CIC values. They are also implemented in the programs TAS and TASS (Nelson and Ladiges, 1994a, 1996). MFW, NLFW, and WHFW weights were calculated by hand, and weighed parsimony searches using these matrices can be performed using Paup* 4b10 (Swofford, 2003).

CLADISTIC INFORMATION CONTENT

Discussion of alternative matrix representations has often focused upon their information content but there has been no agreed measure. Thorley et al. (1998) and Thorley (2000) developed a family of measures of the phylogenetic information content of strict consensus trees that are also applicable to individual trees. They are based on the idea that information is conveyed by a consensus tree to the extent that it permits only a subset of all possible trees (so that, for example, a star tree conveys no information). The amount of information conveyed is determined by the number of trees permitted and by the probabilities of these trees. Information content of a tree is given by the negative log of the sum of the probabilities of the trees it permits (Thorley et al., 1998). In the case of cladistic information content (CIC), we are interested only in relationships specified by the tree topology with no regard to branch lengths or temporal sequences of branching. Other probability distributions for trees are reasonable (e.g., Harding, 1971) and lead to other measures of information content (for example, the probability distribution of trees under a Yule model leads to a measure called dendritic information content by Thorley, 2000), but for simplicity the CIC measure assumes that all fully resolved trees are equally probable (equally uncertain). In this case,

$$\mathrm{CIC} = -\log \frac{\text{number of permitted bifurcating cladograms}}{\text{number of possible bifurcating cladograms}}$$

The units of this measure depend on the base of the logarithm: here we use $\log_2$, which gives information in bits.

Consider the 15 possible rooted binary (fully resolved) trees on four leaves and let each have an equal probability, 1/15, of selection. If an observer is told the tree that was selected, then, $-\log_2 1/15 = 3.91$ bits of information are conveyed. If, however, an observer was instead told that the tree ((A,B),C,D) is the strict component consensus tree of a set of input trees that includes the selected tree, then, because this consensus tree permits three trees, $-\log_2 3/15 = 2.32$ bits of information are conveyed. Many authors have discussed the information content of trees and proposed measures thereof (reviewed in Thorley, 2000) but no other proposed measure is consistent with basic information theory (Hartley, 1928) or measures information content in standard units. CIC values are also readily interpretable (a CIC of 1 bit means a tree permits precisely half of the possible trees), and are very easy to calculate (Thorley and Page, 2000).

Characters are more than just trees (Rodrigo, 1996). However, taken in isolation, each character permits only those trees that it fits without homoplasy (and prohibits those that require extra steps), so that characters convey the same phylogenetic information as consensus trees and this can be measured using CIC. Note that CIC does not measure the weight attached to the information conveyed by a character: it is a measure of quantity not quality. Here we use CIC to measure the information content of pseudocharacters and compare this to the information content of the trees they represent. It should be noted that CIC considers only the taxa specified in a particular statement: the statement (ABCD) does not imply that A, B, C, and D form a clade with respect to all other taxa (which would have a very high CIC), but instead tells us nothing, and has a CIC of zero.

For our purposes, it is useful to distinguish two uses of the word 'independence'—which we will call statistical independence and absolute independence. For two events A and B, statistical independence implies that the occurrence (or otherwise) of A has no bearing on the probability of B occurring: i.e., if A and B are independent in this sense, $p(B) = p(B|A)$ and $p(A) = p(A|B)$. Absolute independence is the generally weaker condition where the truth or otherwise of one of A and B does not entail the truth or otherwise of the other. Where two propositions (relationships) are statistically independent, their information content (CIC) is additive. For example, the triplets A(B,C) and D(E,F) both permit equal size (315 trees) subsets of all 945 binary trees on the leaf set (A–F), and thus convey the same amount of information (1.585 bits). The two triplets together permit only 105 trees and so convey 3.17 bits—exactly the sum of the CICs for the two statistically and absolutely independent triplets. Where relationships are not statistically independent, CIC is not additive (see below).

INFORMATION AND REDUNDANCY

Purvis (1995b) presented an example of two conflicting three and four leaf input trees (Fig. 1) and corresponding alternative matrix representations (Table 1). Standard MRP resolves the conflict in favour of the relationships

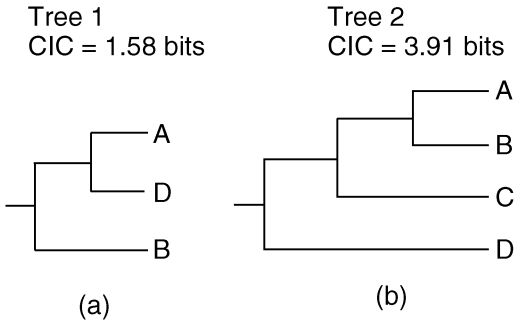

FIGURE 1. The two trees from the first example in Purvis (1995b).

posited in the larger tree. Purvis (1995b: 251) considered that this "choice cannot be justified on the basis of the evidence" because any placement of C in the smaller tree would lead to an unresolved consensus/MRP supertree. Purvis's position allows that trees might be differentially weighted, for example on the basis of the apparent strength of support, and that this might provide a basis for choosing between conflicting relationships (Purvis, 1995a). However, in the absence of such evidence, Purvis's argument is that trees should be given equal weight. Thus methods should not be biased towards relationships in larger trees in cases of conflict (i.e., they should be sizeless, Wilkinson et al., 2004).

Purvis diagnosed the cause of the unjustified result as the repetition of information in the standard matrix representation of the larger tree. The information that A and B are more closely related to each other than they are to D is repeated in both matrix pseudocharacters (Table 1, Ch2 and Ch3), as also noted by Nelson and Ladiges (1992; see below). Purvis coding was proposed to achieve equal weighting of input trees by removing redundant information. In Purvis's approach, each matrix element describes a sister-group relationship with other leaves coded as missing entries (Table 1). Compared to standard matrix representation, Purvis coding removes information by changing selected data points to missing entries (e.g., compare Ch2 and Ch4 in Table 1). However, as with the standard approach, any input tree can be reconstructed through parsimony analysis of its Purvis matrix representation.

Purvis's point is illustrated using CIC as a measure of the information content of the pseudocharacters in the alternative matrix representations (Table 2). With standard MRP, the sum of the information contents of the matrix elements representing the larger tree (Table 1, Ch2 and Ch3, 4.64 bits) is greater than the information content of the input tree (Fig. 1b, 3.91 bits), reflecting the redundancy and repetition of information diagnosed by Purvis. With Purvis's modification the sum of the pseudocharacters (Table 1, Ch3 and Ch4, 3.91 bits) is equal to the information content of the input tree in this simple case.

TABLE 2. Coding of Purvis's second tree (Fig. 3a). Legend is as in Table 1.

Ch	MR	A	B	C	D	E	F	G	H	R	CIC
1	s	1	1	0	0	0	0	0	0	0	3.7004
2	s	1	1	1	0	0	0	0	0	0	5.5749
3	s	0	0	0	0	1	1	0	0	0	3.7004
4	s, p	1	1	1	1	0	0	0	0	0	6.4229
5	s, p	0	0	0	0	1	1	1	1	0	6.4229
6	p	?	?	?	?	1	1	0	0	0	2.3219
7	p	1	1	1	0	?	?	?	?	0	2.3219
8	p	1	1	0	?	?	?	?	?	0	1.5850

Although it appears that Purvis's modification succeeds in removing redundant information in this case, it does not remove all redundant information. Consider the second pseudocharacter in the Purvis matrix (Table 1, Ch4). This three-taxon statement completely resolves the relationships of A, B, and C. It permits 5 of the 15 possible trees including D. The only information that is now nonredundant is that pertaining to the position of D. That D is the sister of A, B, and C is conveyed by the other pseudocharacter (Table 1, Ch3). However, this would convey the same information if either A or B were coded as missing entries, producing two alternative 'minimal' matrices (Table 1, Ch4 and Ch5 or Ch4 and Ch6), each comprising two three-taxon statements or triplets. The minimal matrices yield the original input tree when analyzed with parsimony, so the information removed is redundant in Purvis's sense. Note, however, that the sum of the CICs of the pseudocharacters in the minimal matrices (3.16 bits) is less than the information content of the input tree (3.91 bits). The explanation for this apparent paradox is that the matrix elements are not statistically independent. Considered in isolation, Ch4 and Ch5 of Table 1 both permit 5 out of 15 possible trees (1.58 bits) but when considered together, Ch5 permits just one of the five trees permitted by Ch4 and thus conveys more information (2.32 bits). Where the sum of the CICs for a matrix representation exceeds that of a tree, this demonstrates redundancy in the matrix representation but it provides only a lower bound for the amount of redundant information conveyed by the matrix representation as a whole.

COMPARING MATRIX REPRESENTATIONS

Ronquist (1996) challenged Purvis's (1995b) interpretation of the information removed by Purvis coding as redundant. He compared standard and Purvis matrix representations by measuring the (parsimony) lengths of all possible trees (Fig. 2). He defined 'phylogenetic information' qualitatively, coming to the reasonable conclusion that two matrices have different phylogenetic information content if the parsimony lengths of all trees under the two matrices are not identical (or do not differ by a constant value). He demonstrated that, compared to standard encoding, Purvis's modification reduces the difference between the mean tree length and the optimal tree length, and considered that this made it (p. 248) "less powerful in differentiating among alternative trees" (see Fig. 2). He concluded that "the information removed by Purvis's method cannot be described as redundant." However, the apparent disagreement lies in 'redundant' being used in quite different senses by these authors. Purvis's modification must in some way change

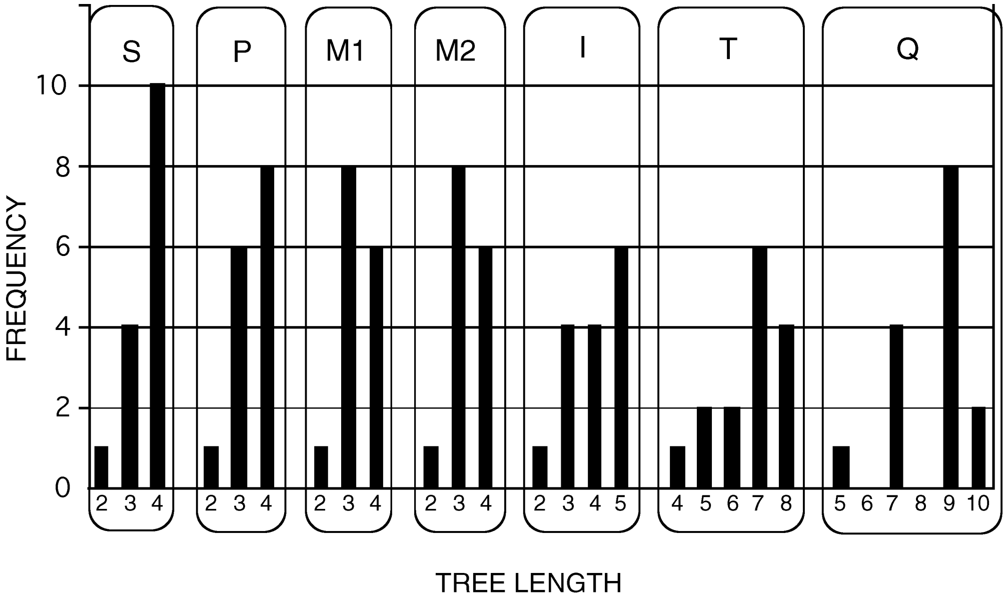

FIGURE 2. Tree lengths of potential supertrees based on different matrix representations of the larger tree (tree 2) from Figure 1 (see Table 1). S is the standard MRP matrix, P the Purvis modified matrix, M1 and M2 the two alternative minimal matrix representations. I are tree scores against the standard matrix under irreversible (Dollo) parsimony. T and Q are tree lengths on the triplet and quartet matrix representations respectively. Note that although M1 and M2 have the same distribution of tree lengths, different trees have minimal lengths against the two matrices.

the phylogenetic information content (sensu Ronquist) of the matrix if it is to have the desired affect of reducing the apparent bias of standard MRP toward relationships in larger trees (or to have any effect at all!). Ronquist's analysis demonstrates only that the information removed by Purvis's method is not without some consequence, which we knew must be the case. It does not demonstrate that it is nonredundant in the sense of being needed to reconstruct the tree, and does not undermine this justification of Purvis coding.

Ronquist (1996: 249) is surely correct that "In the context of combining trees, the important point is not whether Purvis's method removes redundant information but whether it is desirable to modify the information content in the additive binary matrices in the way proposed by Purvis." Given that Purvis coding is (among others) sufficient to represent a tree, the question can be reversed so as not to presuppose priority for any particular matrix representation. Is it desirable to modify the information content of Purvis matrix representations in the manner achieved by standard matrix representations? 'What is the best matrix representation?' is an open question. Considering only a single input tree (or a set of fully compatible input trees), all matrix representations that allow the reconstruction of the input tree are equally good representations of this tree. The question is more important when matrix representations of conflicting input trees are combined—in this case, how the information in the input trees is represented can impact upon how the information is combined and how conflicts are resolved.

So far we have focused upon matrix representations of trees in isolation. A key point that arises from Ronquist's work is that matrix representations can be thought of as representations of the data from which the trees have been inferred. To investigate this, Ronquist compared tree length distributions for Purvis and standard matrix representations, with and without differential weighting based on bootstrap values and decay indices, to the tree length distribution of the original character data. It seems reasonable that the stronger the correlation between these tree length distributions, the better the matrix represents the original character data. Ronquists's insight provides a practical solution to choosing between alternative representations in cases where the tree length distribution supported by the original character data is known. In that case we seek the best representation of the fit of the data to the alternative trees. Given that MRP has been advocated as a surrogate to combined analysis of data in a supermatrix (Bininda-Emonds et al., 1999; but see Pisani and Wilkinson, 2002), any emulation of supermatrix analyses that it achieves might be expected to be enhanced by better representation of the character data. In the cases examined by Ronquist, weighted standard matrix representations were the best, but his examples were too limited to establish this as a general result. It is entirely possible that different data sets might be best represented by different kinds of pseudocharacter matrices. The superiority of a matrix representation method to represent the data, rather than the tree, is contingent upon the data and cannot be decided a priori.

Identifying the best matrix representation is a more difficult problem when we know the best tree(s) but have no information on tree length distributions for the original data. Here, Purvis's preference for equal weighting of trees may seem intuitively reasonable. However, as Ronquist (1996) demonstrated, Purvis coding does not generally lead to equal weighting of trees. This is also

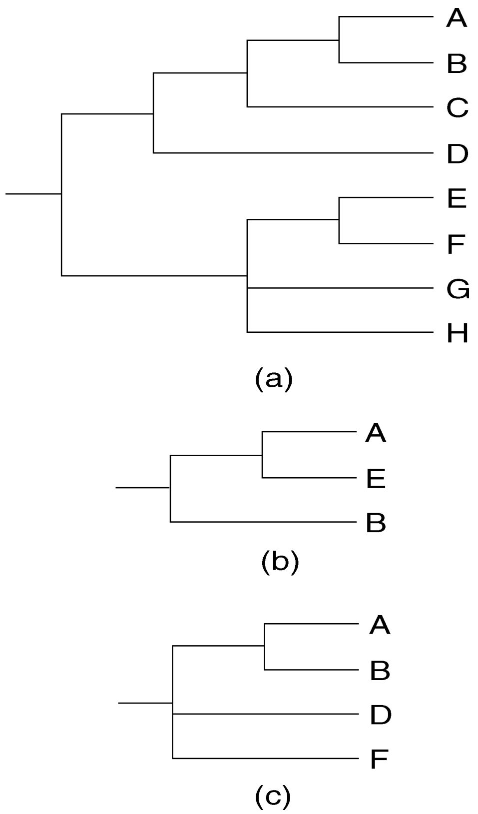

FIGURE 3. The three trees from Purvis's (1995b) second example.

clear from Purvis's (1995b) second example (Fig. 3), in which the only conflict is resolved in favor of the larger tree with both standard and Purvis MRP. Note that, in contrast to the simpler example considered earlier, the information content of the largest input tree (Fig. 3a) (CIC = 15.4591 bits) is not coincident with the sum of the information contents of its Purvis coding pseudocharacters (19.0746 bits), confirming that Purvis coding fails to remove all redundant information. As expected, standard coding (25.8215 bits) includes even more redundant information.

Contra Purvis (1995b), Ronquist (1996) argued that larger trees contain more information and are typically based upon more data (which may or may not be true in any given case) and thus should carry greater weight. Equal weighting of pseudocharacters gives greater weight to larger trees because they contribute more pseudocharacters. However, this is true of both standard and Purvis coding and thus provides no basis for choosing between them.

Equal weighting of trees and equal weighting of pseudocharacters seem to us to be equally reasonable (or unreasonable) default positions, corresponding as they do to two applications of the principle of indifference, a method which is renowned for giving rise to paradoxes (Keynes, 1921). In practice, equal weighting of trees does not seem to be readily achieved with parsimony analysis of matrix representations. Ronquist (1996) suggested weighting individual pseudocharacters inversely to the total number of pseudocharacters in the tree. Thus equal weighting of trees is achieved by differential weighting of the pseudocharacters of different sized input trees. However, Bininda-Emonds and Bryant (1998) showed that, in cases of conflict, the relative size (number of splits) of subtrees spanning the conflicting relationships, rather than tree size per se, may determine how conflict is resolved in MRP supertrees. Consequently, inverse weighting of pseudocharacters does not necessarily correspond to conflicts being resolved on the basis of "one tree one vote." Clique analysis of trees (or corresponding ordered multistate pseudocharacters) is better suited to equal weighting of trees (Purvis, 1995a; Rodrigo, 1996).

Ronquist (1996) also preferred standard MRP to Purvis coding because the former has a larger difference between mean and best tree lengths, and is therefore more powerful in differentiating among trees. We do not consider that this is necessarily a good thing when we are ignorant of the tree length distributions for the original data and how they correlate with those of the matrix representation. Triplet and quartet matrix representations provide even greater spread of tree lengths than standard MRP (Fig. 2), as does the use of irreversible parsimony with standard matrix representations, and would be preferable on this basis. However, increased power to differentiate among trees is correlated with increased redundancy of information in the matrix representation.

Thorley and Wilkinson (2003) suggested that supertree methods in general seek trees that minimise the sum of the distances between the supertree and the input trees. With MRP methods the distance between an input tree and a supertree is simply the number of steps (tree length) required to map the matrix representation of the input tree onto the supertree. Different matrix representations correspond to different tree-to-tree distance measures, and the distributions of tree lengths (Fig. 2) for different matrix representations illustrate this.

CLADISTIC RELATIONSHIPS

Rooted trees portray cladistic relationships. The simplest cladistic relationship is given by a three-item statement or triplet, a three-leaf tree that conveys the information that two leaves are more closely related (share a more recent common ancestor) than either is to a third. Triplets have a CIC of 1.58 bits. Larger trees can be conceptualized as containing a plurality of simpler hypotheses from which they are built and into which they can be decomposed (Wilkinson, 1994). For the moment, we refer to these as 'bricks' (Day and McMorrris, 2003), although, as shall become clear, the analogy this implies cannot be taken very far. As is illustrated by alternative matrix representations, we can decompose trees into different sorts of bricks.

Standard matrix representation decomposes trees into the partitions given by each internal branch, which we shall refer to as components. Each component specifies the members and nonmembers of a clade in the tree. We can conceive of each component as being produced by picking an internal branch on the tree and collapsing all other branches so as to give all and only the relationships specified by the selected internal branch. Trees can also be decomposed into subtrees, with each subtree produced by pruning leaves. Three-taxon statements or triplets are the smallest of such subtree 'bricks' and it is not surprising that matrix representations using three-taxon statements have been used in consensus and supertree construction (Nelson and Ladiges, 1994b; Wilkinson et al., 2001). With Purvis coding, the bricks are sister-group relationships and we can conceive of each brick being produced by selecting an internal node and one of its immediate descendent internal branches, collapsing all other branches above the node, and pruning all leaves that are not descended from that node.

Wilkinson (1994) pointed out that components and triplets both represent hypotheses with a common form, they both correspond to rooted trees with a single internal branch and both identify some set of leaves as more closely related to each other than any are to some other leaf or leaves. He characterized them as end members of a spectrum of such hypotheses varying in the number of leaves, which he collectively termed *n-taxon statements*. More euphoniously, Thorley and Wilkinson (2003) preferred the general term *split*, which applies to both rooted and unrooted trees, and distinguished rooted splits as necessary. Here reference to splits is always to rooted splits. Full splits include all of the taxa present on a tree, partial splits only a subset of the taxa. Components are thus full splits and represent the largest, most inclusive splits that are present in a given tree. Three-item statements or triplets are the smallest or least inclusive partial splits present in a given tree. Sister-group relationships, as represented in Purvis coding, can include full splits and partial splits of various sizes, including triplets.

Williams and Humphries (2003: 255) argued that "In contrast to component coding, in three item coding each node is considered to be a relation between branches" that "relates some branches more closely than other branches of the tree". Similarly, Williams (2004) emphasises that "Binary data represents *groups* (or sub-groups, Wilkinson, 1994a), three-item data represents *relationships* (Nelson and Platnick, 1991; Nelson and Ladiges, 1994b)." We find the suggestion that triplets represent relationships whereas components or splits do not to be unjustified and unhelpful. Although a component can be thought of as a group, and thus can be specified by listing all the leaves descended from an internal node in a rooted tree, it is not simply the set of listed leaves. In the context of a given tree, specifying the members of a group also specifies the nonmembers implicitly. A component is the hypothesis that the leaves in the group are all more closely related to each other than any is to any other leaves in the tree—it may differ in cardinality, but does not differ in form or meaning, from the hypothesis represented by a triplet (Wilkinson, 1994). Williams and Humphries contradictory view overlooks the fact that no information on relationships can be conveyed by specifying members of a group unless some nonmembers of the group are also specified or implied. In fact, Williams and Humphries (2003:255) convert components into the sets of three-item statements that they entail (see below), which surely depends upon interpreting them as hypotheses of relationships.

Cladistic Logic

Rooted splits of cardinality greater than three logically entail sets of partial splits of lower cardinality, including a set of triplets. Consider the simple tree (ABC)D with a single split. It entails the three triplets (AB)D, (AC)D, and (BC)D. Conversely, the set of triplets entails the larger split. However, as noted previously (e.g., Nelson and Platnick, 1980; Nelson and Ladiges, 1992), any two of the three triplets are sufficient to entail the third. This is because any tree that contains two of the triplets must contain the third. Thus, any two of the triplets are minimal matrices that suffice to specify the tree.

Dekker (1986) is an important foundational study of the logic of cladistic relationships that was developed as a contribution to historical linguistics. Considering unrooted trees, he showed that there are just two fundamental inferences that are possible from a pair of quartets (see Bryant and Steel, 1995). The rules applying to pairs of quartets are called *dyadic* rules—higher-order rules applying to triples, quartets, and larger sets of quartets also exist, but do not apply to sets of triplets from a single binary tree. Each of his dyadic inference rules specifies an additional quartet that must be present in any tree that contains the pair (Fig. 4a). In the context of rooted trees (Fig. 4a), Dekker's first dyadic rule translates into (1) (CD)B and (CE)B entail (DE)B. His second rule translates into two dyadic inference rules on triplets that reflect different rootings of the corresponding quartets: (2) (DC)B and (DE)C entail (CE)B, and (3) (AB)D and (DE)A entail (AB)E. The reader may have recognized that additional triplets can be inferred from the given pairs in rules 2 and 3. For example, (DC)B and (DE)C also entail (DE)B, by applying rule 2 followed by rule 1 (Fig. 4b). It may be more natural to think of these triplets as entailing both (DE)B and (CE)B, and it is certainly not incorrect to do so. Dekker sought an irreducible economy in his rules and their form reflects the fact that the fourth quartet can be inferred using rule 1 once the third has been inferred with rule 2 or 3. Meacham (1983) and Wilkinson et al. (2000) presented a set of fusion rules for N-taxon statements that express the results of dyadic inference rules in terms of higher cardinality splits and that govern equations such as (AB)DE and (AC)DE entail (ABC)DE.

These inference and fusion rules are very powerful. Any fully resolved tree with N leaves can be fully entailed by particular sets of just $N-2$ triplets (Steel, 1992), one for each internal branch, and equal to the number of full splits or components in the tree. Because all the other triplets are entailed by this small set of triplets, they are absolutely independent so that a fully resolved tree includes only $N-2$ absolutely independent triplets (or more accurately one or more sets of $N-2$ absolutely independent triplets). Thus in the minimal matrices considered above (Table 1), only two triplets suffice for parsimony to reconstruct a tree that contains four leaves and

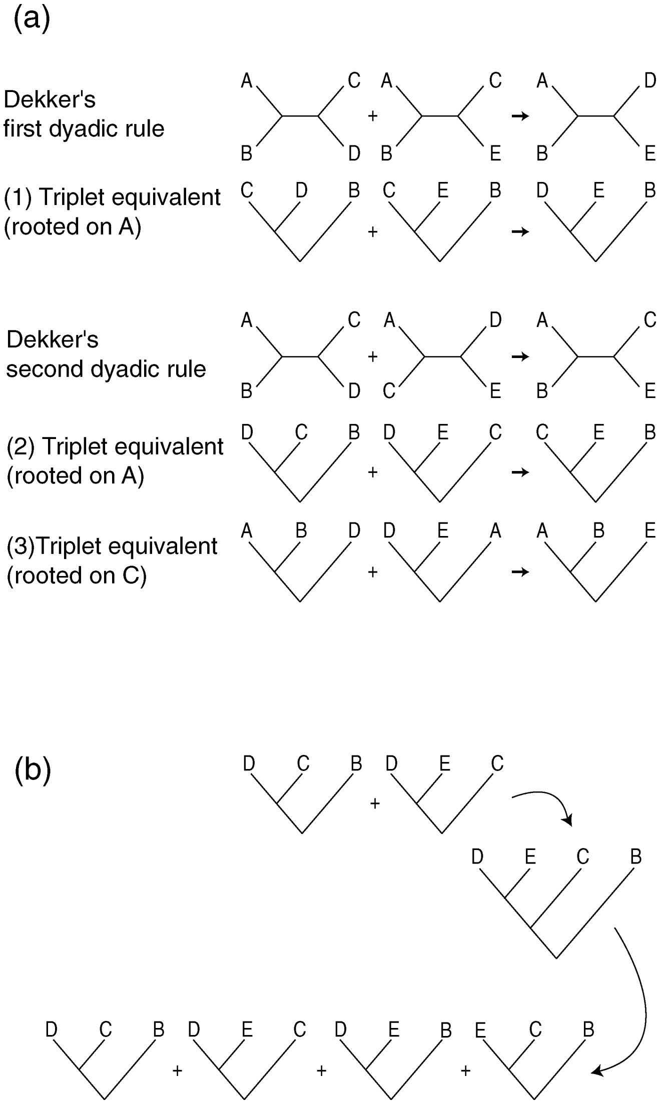

FIGURE 4. (a) Dekker's (1986) two dyadic inference rules for quartets and their rooted (triplet) equivalents. (b) These inference rules work because, for example, only a single rooted tree is compatible with the two triplets shown, and this tree implies two additional triplets. Any tree containing the two triplets (C, D) B and (D, E) C must also contain the triplets (D, E) B and (E, C) B.

a total of four triplets. We could include the two other triplets in the matrix representation, but this only adds information that is already present, albeit implicitly, in the minimal matrix.

REDUNDANCY IN TRIPLET MRP

Nelson and Ladiges (1994b) first used parsimony analysis of triplet representations of trees to produce consensus trees for trees with identical leaf sets. Their three-item consensus was applied using either uniform or fractional weighting, the latter intended to remove the impact of logical dependencies among triplets. These methods are readily extended to cases where input trees do not have identical leaf sets, and Wilkinson et al.'s (2001) triplet MRP, as implemented in RadCon (Thorley and Page, 2000), is such an extension of uniformly weighted three-item consensus.

We have seen that the sum of the CICs of the pseudocharacters in both standard and Purvis matrix representation of a tree may exceed the CIC of the tree, indicating redundancy in the matrix representation. Representing a tree as the set of triplets can also lead to redundant information. For example, the sum of the CICs for the four triplets in the tree ((AB)C)D is 6.34 bits (Table 1), whereas the CIC for the tree is just 3.91 bits. Redundancy here is through the logical relations of different triplets, governed by Dekker-type inference rules, rather than through the repetition of individual triplets. The number of triplets in a fully resolved tree, $N \times (N-1) \times (N-2)/6$, increases polynomially with N, the number of leaves, whereas the number of absolutely independent triplets required to specify a tree grows linearly with N(Steel, 1992). This ensures that redundant information content of matrix representations that include all the triplets present in the tree increases polynomially with the number of leaves. For example, the largest tree in Figure 3 contains 54 triplets with a total CIC of 85.59 bits, far in excess of the total CICs for the standard and Purvis matrix representations (25.8215 and 19.0746 bits respectively) and of the tree itself (15.4591 bits). Thus triplet MRP might be expected to have a stronger bias toward larger trees than standard or Purvis MRP.

NELSON AND LADIGES FRACTIONAL WEIGHTING

Nelson and Ladiges (1992) suggested that logically dependent sets of three-item statements be downweighted so that their combined weight is equal to the minimum number needed to entail them all. A binary character for N leaves with D leaves scored for the derived character state entails $D \times (D-1) \times (N-D)/2$ triplets, of which $(D-1) \times (N-D)$ are needed to entail the rest and are therefore absolutely independent, and Nelson and Ladiges fractional weighting (NLFW) of each triplet is $2/D$ (Nelson and Ladiges, 1992). For given N, the number of triplets differs for different values of D so that when triplet encodings of different characters are combined and analyzed with uniform weights, they will be represented by different numbers of triplets and have different weights. The weighting can be substantial, for example, with $N = 20$ a character with $D = 2$ entails 18 triplets, all of them absolutely independent, whereas $D = 10$ entails 450 triplets of which 90 are absolutely independent. CICs for the two characters are 5.2095 bits and 18.4715 bits respectively, whereas the CICs for the triplet encodings sum to 28.44 and 711 bits respectively with uniform weighting. NLFW compensates for the logical dependencies and redundancy of information in uniformly weighted triplet encodings of binary character data by rendering their total weight equal to the number of absolutely independent triplets. It does not lead to an equal weighting of sets of triplets derived from different binary characters (with different numbers of taxa with the derived state), rather these have different overall weights that, to some extent, reflect the different information contents of different characters. Note that the logical dependencies among the triplets entailed by a single binary character, considered in isolation, are governed solely by the first dyadic inference rule on triplets described above.

Nelson and Ladiges's (1992) discussion of fractional weighting considered triplets derived from binary and multistate characters corresponding to cladograms. They recognized that trees with more than a single split correspond to multistate character state trees, and illustrated NLFW of multistate characters and trees with examples. They explained (p. 491) "the multistate character corresponding to the cladogram A(B(CD)), which specifies four three-taxon statements, each with an absolute value of 1: A(BC), A(BD), A(CD), B(CD). If the multistate character is rendered as two binary characters, corresponding to cladograms A(BCD) and AB(CD), then five three-taxon statements are specified: A(BCD) yields the series A(BC), A(BD), A(CD); AB(CD) yields the series A(CD), B(CD). The statement A(CD) occurs in both series. Uniform and fractional weights (both misleading) for the five statements are listed . . . and compared with their absolute values." By absolute values they mean "the ratio of independent statements to all possible statements."

Nelson and Ladiges's (1992) example illustrates the problem that Purvis highlighted and sought to address. For our purposes, more significant is the parenthetical comment that both the uniform and fractionally weighted three-item statements derived from the standard matrix representation are misleading. Nelson and Ladiges (1992) considered the absolute value of each three-item statement to be unity and the total absolute value of all three-item statements in the tree to be four. Three-item statements derived from the standard matrix representation are misleading because of redundant information in the standard matrix representation. With uniform weighting, A(CD) is overweighted through repetition, and the total value of the tree is correspondingly inflated. Although fractional weighting gives the correct total value for the tree in this case, A(CD) is still overweighted through repetition, and this is compensated by underweighting A(BC) and A(BD). According to Nelson and Ladiges (1992), in fully resolved pectinate trees all the triplets are independent and are thus unaffected by NLFW, four equally weighted triplets are considered to be the appropriate representation (with or without NLFW). However, recall that a minimum of just two triplets is needed to represent any fully resolved four-taxon tree: the four triplets are not absolutely independent.

Nelson and Ladiges (1992) further illustrated their fractional weighting (NLFW) in their Table 4, listing the number of triplets and the number of absolutely independent triplets for a a range of trees (Table 3). Trees with terminal polytomies include sets of logically dependent triplets that are downweighted with NLFW. They (p. 492) further explain that nonpectinate trees, i.e., those with internal nodes "that lead distally only to other nodes and not directly to terminal taxa . . . always specify fewer independent statements than total statements. Fractional weighting reduces the absolute value of statements of internal nodes. For cladogram 11, there is one internal node, ((AB)(CD)), with four statements, each with an absolute value of 3/4." Their table correspondingly indicates that the tree includes 10 triplets of which 9 are independent.

To clarify, E is the sister taxon of (ABCD) in this example and ((AB)(CD)) specifies a complete subtree including an ancestral node and its two descendent nodes. NLFW downweighting applies to the triplets that are entailed by the ancestral node but are not also entailed by the descendent nodes. Thus the ancestral node specifies the relationship (ABCD)E, which corresponds to six three-item statements (of which three are absolutely independent). Two of these—(AB)E and (CD)E—are also specified by the descendent nodes, leaving four—(AC)E, (BC)E, (AD)E, and (BD)E—that are entailed by the ancestral node alone. Any three of these allows inference of the fourth, and all four are weighted 3/4. Fractional weighting of three-item encodings of character data has previously been criticized (DeLaet and Smets, 1998) because it does not take into account homoplasy introducing

TABLE 3. Number of independent three-taxon statements for some five-taxon cladograms, including what Nelson and Ladiges (1992) considered independent triplets, the total number of triplets implied by the cladogram, and the correct number of absolutely independent triplets. Included are NLFW and MFW weights for each triplet, and the CIC of each cladogram. Modified from Table 4 of Nelson and Ladiges (1992).

Number from Nelson and Ladiges (1992)	Cladogram	Nelson and Ladiges (1992) 'Independent' triplets	All triplets	Independent triplets	NLFW	MFW	CIC (bits)
2	(AB)CDE	3	3	3	1	1	2.8074
3	(ABC)DE	4	6	4	4/6	4/6	3.5443
4	(ABCD)E	3	6	3	3/6	3/6	2.8074
5	(AB)C(DE)	6	6	4	1	4/6	5.1293
6	((AB)C)DE	7	7	3	1	3/7	5.1293
7	(AB)(CDE)	7	9	3	7/9	3/9	5.1293
8	((AB)CD)E	6	8	4	6/8	4/8	5.1293
9	((ABC)D)E	7	9	3	7/9	3/7	5.1293
10	(AB)(C(DE))	10	10	3	1	3/10	6.7124
11	((AB)(CD))E	9	10	3	9/10	3/10	6.7124
12	(((AB)C)D)E	10	10	3	1	3/10	6.7124

incongruence between characters, a problem that does not apply where the characters are matrix representations of trees. We find the details of NLFW of triplet encodings of trees difficult to justify. In particular, we see no reason for downweighting the selected triplets when trees are not asymmetric. It does not seem obvious that these triplets are any less well supported. Such downweighting can only serve to bias the results, in cases of conflict, towards relationships in asymmetric trees.

MINIMAL FRACTIONAL WEIGHTING (MFW)

Perhaps because of an initial focus on three-item statements as alternative representations of characters, no discussion of fractional weighting has fully discussed logical relations that can pertain among the triplets entailed by a tree. Discussion has mostly focussed on the subset of such dependencies that pertain among triplets entailed by a single component (or internal branch) of a tree (but see Nelson and Platnick, 1981). NLFW attempts to correct for only this latter subset of logical relations (governed by the first dyadic inference rule). It does not take into account logical dependencies that are governed by the other dyadic rules, and therefore seems to us to be incomplete. Extending Nelson and Ladiges's (1992) rationale for fractional weighting from characters to trees, the weight of a three-item statement should be the ratio of the minimal number of triplets needed to specify the tree, divided by the number of triplets in the tree. There is no need to consider splits separately. A fully resolved binary tree implies $N \times (N-1) \times (N-2)/6$ triplets, one triplet for each combination of three taxa, but is fully defined by $N-2$ triplets (one for each internal edge; Steel, 1992). Taking into account all possible inference rules, the total weight of a fully resolved tree is thus $N-2$, with each triplet having a weight $6/(N \times (N-1))$, irrespective of tree shape. We call this minimal fractional weighting (MFW) to distinguish it from NLFW. Calculating the number of absolutely independent triplets (and so the MFW weights) for nonbinary trees is more complex, depending on the degree and level of resolution in the tree.

There are important and obvious differences between NLFW and MFW. NLFW may give triplets from the same tree different weights, whereas MFW weights them uniformly. We consider differential weighting of triplets to reflect measures of support to be reasonable, but otherwise see no reason to weight the triplets from a single tree differentially. MFW is based on all relevant triplet inference rules, whereas NLFW is based only on the first dyadic rule. If our aim is to downweight in order to compensate for absolute dependence between triplets, then we see no reason not to take into account all the dependencies. MFW is consistent with Ronquist's view that larger trees convey more information and is not concerned with trying to give different sized trees equal weight. MFW counters the polynomial increase in redundancy in three-item matrices with the size (number of leaves) of the input trees, and thus reduces the potential bias towards larger trees. Considering fully resolved trees for simplicity, tree weight $(N-2)$ increases linearly with the number of leaves (N) with MFW. With MFW, individual triplets from larger trees have less weight than those from smaller trees, as the weight of an individual triplet is inversely proportional to the number of triplets in the tree. In contrast, with NLFW individual triplets from a larger tree may have higher or lower weights than triplets from a smaller tree, depending on tree shape.

WILLIAMS AND HUMPHRIES

Williams and Humphries (2003) noted that the disagreement between Purvis (1995b) and Ronquist (1996) highlighted "a need to explore possible differences between" alternative matrix representations. They raised the question (p. 256) "of whether component coding, Purvis coding, and multistate coding are equivalent simply because they yield the same correct cladogram." Although the answer is context specific, in the usual context of combined matrix representations of conflicting input trees, the alternatives are clearly not equivalent. They represent the same information (the relationships displayed by the input trees), but do so differently and with different degrees of redundancy. This impacts upon how information from different trees is combined and how conflict is, or is not, resolved.

Williams and Humphries explored this issue by converting alternative standard and Purvis matrix representations into corresponding three-item matrix representations, both with and without fractional weighting. This translation into a common language enabled them to compare the alternatives, and pinpoint and quantify differences between them. They focused on a simple pectinate, five-leaf example equivalent to cladogram 12 of Nelson and Ladiges (1992), that includes some 10 distinct three-item statements. The alternative matrix representations considered by Williams and Humphries, and their translations into matrices of three taxon statements, are given in Table 4, together with their CICs. The standard matrix representation is translated into 15 three-taxon statements including some duplication of statements entailed by different components. In contrast, multistate and Purvis codings both translate to the 10 unique

TABLE 4. Three possible binary codings of the cladogram A(B(C(DE))). Includes the CIC of each character. The CIC of the cladogram is 6.7142 bits, the sum of CICs for the three characters in the standard component coding is 9.1581 and the sum of CICs for the three characters in the modified Purvis coding is 6.7124, illustrating the redundancy in the standard component coding. Modified from Table 1 of Williams and Humphries (2003).

	Component coding			Purvis coding			MS
	1	2	3	1	2	3	1
A	0	0	0	0	?	?	1
B	1	0	0	1	0	?	2
C	1	1	0	1	1	0	3
D	1	1	1	1	1	1	4
E	1	1	1	1	1	1	4
CIC (bits)	2.8074	3.5433	2.8074	2.8074	2.3219	1.5850	6.7124

TABLE 5. Comparison of numbers of three-item statements derived from the cladogram A(B(C(DE))). Component = which component(s) from Table 4 the triplet is implied by. UW = uniform weighting, FW = fractional weighting (WHFW). MFW = our minimal fractional weighting. Modified from Williams and Humphries (2003).

	Standard coding			Purvis coding				
Three-item statement	Component	UW	FW	Component	UW	FW	Multistate	MFW
A(BC)	1	1	1/2	1	1	1/2	1	3/10
A(BD)	1	1	1/2	1	1	1/2	1	3/10
A(BE)	1	1	1/2	1	1	1/2	1	3/10
A(CD)	1,2	2	7/6	1	1	1/2	1	3/10
A(CE)	1,2	2	7/6	1	1	1/2	1	3/10
A(DE)	1,2,3	3	13/6	1	1	1/2	1	3/10
B(CD)	2	1	2/3	2	1	2/3	1	3/10
B(CE)	2	1	2/3	2	1	2/3	1	3/10
B(DE)	2,3	2	5/3	2	1	2/3	1	3/10
C(DE)	3	1	1	3	1	1	1	3/10
Total	15 statements	15	10	10 statements	10	6	10	3

three-taxon statements that are present in the tree with no repetition and William's and Humphries (2003) remark (p. 257) that Purvis coding treats the cladogram "as if it really were a multistate character."

The fractional weights calculated by Williams and Humphries for their translated matrices are shown in Table 5. Fractional weights are calculated independently for each set of triplets entailed by each standard or Purvis coding pseudocharacter. Where triplets are replicated by different pseudocharacters, their weights are summed. Williams and Humphries fractional weighting (WHFW) differs from NLFW in the two-stage decomposition of trees into components (or Purvis pseudocharacters), and then into the fractionally weighted triplets entailed by the pseudocharacters. Following Nelson and Ladiges (1992), the three-item matrix representation of a five-leaf pectinate tree includes ten uniformly weighted triplets, and NLFW has no effect, as in Williams and Humphries' multistate character representation of the tree (Table 3, tree 12). Using WHFW, the translated Purvis coding includes the same 10 unique three-taxon statements but with a combined weight of six. With translated standard coding, individual weights range from 1/2 to 13/6, and the weights now sum to 10. In contrast, MFW would assign each of the 10 triplets a weight of 0.3 because only three triplets are required to entail the tree.

Williams and Humphries (2003) noted with approval that fractional weighting applied to the three-item translation of the standard MRP matrix yielded a combined weight equal to the number of three-item statements in the tree (p. 257): "for the total array of 15 three-item statements, application of fractional weighting reduces the relative value yielding a final weight of 10 and reflecting the total weight of the original 10 contributions. The precision rests in accounting for all the relevant information." However, although this may hold for fully pectinate trees, counterexamples provided by trees of different shapes show clearly that there is no general equivalence in the number of triplets and their combined fractional weights under WHFW. For example, the tree in Figure 3a includes 54 triplets, but the combined weight using WHFW of triplets derived from standard matrix representation is only 46. If the claim that WHFW of triplet translations of standard matrix representations accounts for all the relevant information rests on the precise agreement between number of triplets and their combined weight, then the claim cannot be generally true. We are also unclear as to what is meant by 'accounting for all of the information' in this context. As we have seen, Nelson and Ladiges (1992) recognized, in a similar case, that WHFW of triplets derived with repetition from standard matrix representations was misleading, and we agree with their assessment.

Considering the translated Purvis matrix representations, Williams and Humphries (2003: 257) note that "Of the 10 statements derived from the three truncated binary components, some are logically dependent and require correction." With WHFW, their total weight is now just six. They consider "This reduced total weight is deficient, in part because of the missing values added for each component but also because of a critical loss of informative statements." Logically, any deficiency in weight must be totally attributable to the added missing values because this is the only way in which Purvis coding differs from the standard matrix representation. More importantly, we can see no loss of informative statements, given that all triplets are present and have non-zero weight (Table 4). Perhaps 'loss of informative statements' refers to those that are replicated when triplet matrices are constructed from standard matrix representations.

Williams and Humphries (2003: 257) continued: "At first sight, one might conclude that Purvis's method of coding is more accurate than its three item equivalent because it corrects for redundancy without any weighting. However it remains inaccurate because the recovered information is significantly below that needed when compared with either the three separate components or the multistate character." Here the strong claim that Purvis coding is somehow inaccurate and contains less information than it in some sense needs, rests on the fact that the triplet-translated and fractionally weighted Purvis coding has less total weight than the corresponding translated standard matrix representation. Williams and Humphries do not explain why this serves as a benchmark, or discuss the misleading differential weighting that distinguishes the latter from the former. Rather they seem to consider Purvis coding deficient simply because the three-item translation, with the WHFW that they consider to be required, has lower total weight than their proposed method. On second look, we suggest that Purvis and multistate coding translate into three-item representations with uniform weights that are less misleading than the translation from standard matrix representation, with or without WHFW.

Williams and Humphries (2003) go on to consider which of the various coding schemes are 'most efficient': "For consensus methods to be efficient, the components are best treated as independent, a view that corresponds to that of Ronquist (1996). But use of binary components does indeed overweight, a view that corresponds to that of Purvis (1995b). Three-item coding, corrected for redundancy, may satisfy both Ronquist and Purvis." To be precise, they suggest that their WHFW three-item matrix translations of standard matrix representations may resolve the disagreement of Purvis (1995b) and Ronquist (1996) and provide the most efficient representation of a tree.

We do not find this suggestion convincing. Purvis's coding was intended to reduce the redundancy in standard matrix representations, whereas, as we have seen, the information content and redundancy of triplet matrices typically far exceeds that of the corresponding standard or Purvis matrix representation. WHFW compounds the problem with its two-stage process of encoding a tree into a standard matrix representation, and then translating this into a triplet matrix representation. As Nelson and Ladiges (1992) recognized, this introduces redundancy through repetition. Even when total weight does equal the number of triplets (as in Williams and Humphries' example), the apparent compensation typically involves downweighting some triplets that are not replicated, yielding differential weights that we consider hard to justify. As with NFLW, WHFW does not take into account logical dependencies among triplets in a tree that are governed by the second or third dyadic inference rules. If should be noted that Williams and Humphries use of 'efficiency' to mean something akin to making efficient use of the data as evidenced by greater resolving power is markedly different to previous use of the term in consensus theory (Bryant, 1997; Wilkinson and Thorley, 2001). In any case, we think if important to distinguish between resolution that is and that is not worth having, and that resolution per se is a poor criterion to judge alternative methods.

CONCLUDING REMARKS

Standard MRP is currently the most frequently used method of supertree construction. However, various matrix representations of trees and associated weighting schemes are possible, confronting workers with methodological choices, and discussions have begun to address which are the best methods. We believe this question remains unanswered, and that the answer may be context-specific. For example, when the underlying data are available, the best matrix representation might reasonably be taken to be that which best reflects the original data. This can be assessed by, for example, comparing the tree length distributions of the data and their matrix representation (Ronquist, 1996). In contrast, in the absence of comparison with the original data, there seem to be few compelling reasons to prefer any of the alternative matrix representations. The standard component-based approach appears to have become the arbitrary default, but we do not think other approaches that encode relationships in different ways, such as triplets or sister groups, should be dismissed.

Although Purvis coding does not produce equal weighting of input trees as intended, compared to standard coding it includes less redundant, repeated information and reduces the magnitude of the differential in the weights of differently sized trees. If standard MRP has a bias toward larger trees (Bininda-Emonds and Bryant, 1998; Purvis, 1995a; Ronquist, 1996), then Purvis coding would be expected to reduce the bias. In contrast, because there are many more triplets than there are components or sister-group relationships, triplet MRP methods run the risk of exacerbating any size bias. Based on the logical dependencies between triplets as encapsulated in the rooted analogues of Dekker's (1986) inference rules, minimum fractional weighting offers a suitable compensatory downweighting. In contrast, NLFW and WHFW take into account only a subset of the logical dependencies among triplets.

In the absence of good reasons, either theoretical or derived from simulations or empirical studies, for preferring one matrix representation (or tree-to-tree distance; Thorley and Wilkinson, 2003) over another, we believe it is sensible to explore the extent to which methodological choices matter in practice. Confronted with a range of poorly understood options, pluralism may be a better strategy than using an arbitrary default, not least because of the insights that can be obtained on the strength of the results (Wilkinson and Benton, 1996). We stress that we do not advocate our minimal matrices. Alternative minimal matrices have the same tree-length distributions (Fig. 2), but the lengths of some individual trees differ so that, in the context of supertree construction from multiple input trees, different minimal matrices could arbitrarily support different optimal supertrees.

Williams and Humphries (2003) sought to use triplets as a comparative yardstick. Translating standard and Purvis matrix representations into a common language is a sensible strategy, but we do not concur with many of the conclusions that Williams and Humphries based upon it. The number of triplets in a tree has been used in a number of measures intended to measure the information content of trees, but this is problematic because the information conveyed by the triplets is not independent (Thorley, 2000). CIC provides a better measure of the information content of a tree and its use here reveals the redundancy in triplet, Purvis, and standard matrix representations when the pseudocharacters are treated as independent.

Triplets have a special place in the pantheon of cladistic relationships because of their irreducibility. They are the smallest informative statements of cladistic relationships. All other relationships, trees, components, sister-group relationships, and even nestings (Adams, 1986), can be reduced to combinations of triplets that they logically entail (Wilkinson, 1994). We might therefore be tempted to view triplets as an atomic unit of cladistic relationships. Perhaps this perspective has led some to consider matrix representations of triplets to be in some sense more fundamental than other matrix representations, and preferable to them a priori (e.g., Williams and Humphries, 2003). However, as we have seen, there are extensive logical dependencies between the triplets (and between splits more generally) that are present in a tree, and this leads to extensive redundancy of the cladistic information in triplet matrix representations. It is difficult to identify any kind of logically independent 'bricks' in trees that can be represented in a matrix without some redundancy of cladistic information through repetition or other logical dependence. This limitation of the 'brick' analogy suggest to us that triplets are more analogous to quarks than to atoms, irreducible but not independent. The existence of alternative matrix representations demonstrates that trees can be decomposed in different ways into constituent units. Rather than consider any of these units as preferable a priori, we should accept that they are all simply proxies for the tree that must serve a practical purpose. Future work must be directed at determining how useful the alternatives are in practice.

ACKNOWLEDGMENTS

We thank Ian Corfe, Chris Creevey, David Gower, Gary Nelson, and Davide Pisani for their comments on the manuscript and David Williams and Gary Nelson for help in implementing WHFW and NLFW. This work was supported by BBSRC grant 40/G18385.

REFERENCES

Adams, E. N. 1986. *N*-trees as nestings: Complexity, similarity and consensus. J. Classif. 3:299–317.

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., and H. N. Bryant. 1998. Properties of matrix representation with parsimony analyses. Syst. Biol. 47:497–508.

Bininda-Emonds, O. R. P., J. L. Gittleman, and A. Purvis. 1999. Building large trees by combining phylogenetic information: A complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. Cam. Phil. Soc. 74:143–175.

Brooks, D. R. 1981. Hennig's parasitological method: A proposed solution. Syst. Zool. 30:229–249.

Bryant, D. 1997. Building trees, hunting for trees and comparing trees. Unpublished PhD Thesis. Department of Mathematics, University of Canterbury, New Zealand.

Bryant, D., and M. Steel. 1995. Extension operations on sets of leaf-labelled trees. Adv. Appl. Math. 16:425–453.

Day, W. H. E., and F. R. McMorrris. 2003. Axiomatic consensus theory in group ghoice and biomathematics. SIAM, Philadelphia.

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

DeLaet, J., and E. Smets. 1998. On the TTSC-FTSC formulation of standard parsimony. Cladistics 14:239–248.

Farris, J. S. 1973. On comparing the shapes of taxonomic trees. Syst. Zool. 22:193–200.

Harding, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. Adv. Appl. Prob. 3:44–77.

Hartley, R. V. L. 1928. Transmission of information. Bell Sys. Tech. J. 7:535–563.

Jones, K. E., A. Purvis, A. MacLarnon, O. R. P. Bininda-Emonds, and N. B. Simmons. 2002. A phylogenetic supertree of the bats (Mammalia:Chiroptera). Biol. Rev. Cam. Phil. Soc. 77:223–259.

Kennedy, M., and R. D. M. Page. 2002. Seabird supertrees: Combining partial estimates of procellariiform phylogeny. Auk 119:88–108.

Keynes, J. M. 1921. A treatise on probability. Macmillan, London.

Lapointe, F.-J., and G. Cucumel. 1997. The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. Syst. Biol. 46:306–312.

Lapointe, F.-J., M. Wilkinson, and D. Bryant. 2003. Matrix representation with Parsimony or with Distances: Two sides of the Same Coin? Syst. Biol. 52:865–868.

Meacham, C. A. 1983. Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. Pages 304–314 *in* Numerical taxonomy, NATO ASI Series, vol. G1, (J. Felsenstein, ed.). Springer-Verlag, Berlin.

Nelson, G., and P. Y. Ladiges. 1992. Information-content and fractional weight of 3-item statements. Syst. Biol. 41:490–494.

Nelson, G., and P. Y. Ladiges. 1994a. TAS—Three Area Statements. MS-DOS Program for Cladistic Biogeography, version 3.3. New York and Melbourne.

Nelson, G., and P. Y. Ladiges. 1994b. 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.

Nelson, G., and P. Y. Ladiges. 1996. TASS—Three Area Subtrees v2.1. MS-DOS Program. New York and Melbourne.

Nelson, G., and N. I. Platnick. 1980. Multiple branching in cladograms: Two interpretations. Syst. Zool. 29:86–91.

Nelson, G., and N. I. Platnick. 1981. Systematics and biogeography: Cladistics and vicariance. Columbia University Press, New York.

Nelson, G., and N. I. Platnick. 1991. 3-Taxon statements: A more precise use of parsimony. Cladistics 7:351–366.

Phillips, C., and T. J. Warnow. 1996. The assymetric median tree—A new model for building consensus trees. Discrete App. Math. 71:311–335.

Pisani, D., and M. Wilkinson. 2002. Matrix representation with parsimony, taxonomic congruence, and total evidence. Syst. Biol. 51:151–155.

Pisani, D., A. M. Yates, M. C. Langer, and M. J. Benton. 2002. A genus-level supertree of the Dinosauria. Proc. R. Soc. Lond. 269:915–921.

Purvis, A. 1995a. A composite estimate of primate phylogeny. Phil. Trans. R. Soc. Lond. 348:405–421.

Purvis, A. 1995b. A modification to Baum and Ragan's method for combining phylogenetic trees. Syst. Biol. 44:251–255.

Ragan, M. A. 1992. Phylogenetic inference based on matrix representation of trees. Mol. Phylogen. Evol. 1:53–58.

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

Ronquist, F. 1996. Matrix representation of trees, redundancy, and weighting. Syst. Biol. 45:247–253.

Sanderson, M. J., A. Purvis, and C. Henze. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol. Evol. 13:105–109.

Semple, C., and M. Steel. 2002. Tree reconstruction from multi-state characters. Adv. Appl. Math. 28:169–184.

Steel, M. 1992. The complexity of reconstructing trees from qualitative characters and subtrees. J. Classif. 9:91–116.

Swofford, D. L. 2003. PAUP*, Phylogenetic Analysis using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thorley, J. L. 2000. Cladistic information, Leaf stability and Supertree construction. Unpublished PhD Thesis. Department of Biology, University of Bristol, Bristol, UK.

Thorley, J. L., and R. D. M. Page. 2000. RadCon: Phylogenetic tree comparison and consensus. Bioinformatics 16:486–487.

Thorley, J. L., and M. Wilkinson. 2003. A view of supertree methods. Pages 185–194 *in* Bioconsensus (M. F. Janowitz, F. J. Lapoine, F. R. McMorris, and F. S. Roberts, eds.). American Mathematical Society, Providence, Rhode Island.

Thorley, J. L., M. Wilkinson, and M. Charleston. 1998. The information content of consensus trees. Pages 91–98 *in* Advances in data science and classification (A. Rizzi, M. Vichi, and H. H. Bock, eds.). Springer-Verlag, Berlin.

Wilkinson, M. 1994. Common cladistic information and its consensus representation—reduced Adams and reduced cladistic consensus trees and profiles. Syst. Biol. 43:343–368.

Wilkinson, M., and M. J. Benton. 1996. Sphenodontid phylogeny and the problems of multiple trees. Phil. Trans. R. Soc. Lond. Ser. B 351:1–16.

Wilkinson, M., and J. L. Thorley. 2001. Efficiency of strict consensus trees. Syst. Biol. 50:610–613.

Wilkinson, M., J. L. Thorley, D. T. J. Littlewood, and R. A. Bray. 2001. Towards a phylogenetic supertree of Platyhelminthes? Pages 292–301 *in* Interrelationships of the Platyhelminthes (D. T. J. Littlewood, and R. A. Bray, eds.). Taylor and Francis, London.

Wilkinson, M., J. L. Thorley, D. Pisani, F.-J. Lapointe, and J. O. McInerney. 2004. 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.

Wilkinson, M., J. L. Thorley, and P. Upchurch. 2000. A chain is no stronger than its weakest link: Double decay analysis of phylogenetic hypotheses. Syst. Biol. 49:754–776.

Williams, D. M. 2004. Supertrees, components and three-item data. Pages 389–408 *in* Phylogenetic supertrees: Combining information to reveal the tree of life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Williams, D. M., and C. J. Humphries. 2003. Component coding, three-item coding, and consensus methods. Syst. Biol. 52:255–259.

First submitted 20 December 2003; reviews returned 29 March 2004; final acceptance 13 August 2004

Associate Editor: Mike Steel