

Identifying stable reference taxa for phylogenetic nomenclature

MARK WILKINSON

Accepted: 1 September 2005
doi:10.1111/j.1463-6409.2005.00213.x

Wilkinson, M. (2006). Identifying stable reference taxa for phylogenetic nomenclature. — *Zoologica Scripta*, 35, 109–112.

Mark Wilkinson, Department of Zoology, The Natural History Museum, London SW7 5BD, UK.
E-mail: marw@nhm.ac.uk

Introduction

Codes and customs of nomenclature aim to provide clarity and stability in the meaning, content and use of the names of taxa. In phylogenetic nomenclature, the meaning and content of a taxon name are contingent upon some phylogenetic hypothesis and subject to change (instability) should relevant views of phylogeny be revised. Many measures of support for phylogenetic hypotheses have been developed and, assuming that seemingly well-supported relationships are least likely to prove unstable, nomenclatural stability can be promoted by avoiding phylogenetic definitions that are contingent upon seemingly weakly supported relationships (e.g. Lee 2005). Measures such as bootstrap proportions, Bayesian posterior probabilities, and decay indices are routinely used to rank the support for clades (full splits) in a phylogeny and can be used to guide the pursuit of stability in phylogenetic nomenclature. Lee (2005) suggested that, in addition, *Leaf Stabilities* (LS), a family of measures developed by Joseph Thorley and myself (Thorley & Wilkinson 1999; Thorley 2000) be used to identify relatively unstable leaves (terminal taxa). This is useful because (p. 329) ‘choice of reference taxa can greatly influence the stability of phylogenetic taxonomies’ so that ‘such instability can be minimized by choosing phylogenetically stable reference taxa’. While applauding Lee’s (2005) concern for stability, and agreeing that leaf stability measures can sometimes be useful in distinguishing relatively stable and unstable taxa, my aim here is to draw attention to alternative approaches that may prove more useful.

Leaf stabilities

Measures of support are mostly given for full splits (components, clades) but can also be determined for less inclusive relationships. In unrooted trees the least inclusive phylo-

genetic hypothesis specifies the relationships among a quartet of leaves, e.g. ABCD, which can be resolved in one of three ways, e.g. AB/CD, AC/BD, AD/BC, in any unrooted phylogenetic tree. Bootstrap support for each of these alternative resolved quartets is given by the frequency of occurrence of the resolved quartet in the bootstrap-resampled phylogenies. Similarly, the decay index of a resolved quartet is the score of the optimal tree not including the resolved quartet minus the score of the optimal tree including the resolved quartet. In a rooted tree, the least informative phylogenetic hypothesis specifies the relationships a triplet of leaves, e.g. ABC, which can be resolved in one of three ways (AB)C, (AC)B, (BC)A in any rooted tree. It is useful to think of the root as a special leaf, R, and recognize that resolved triplets are a subset of the resolved quartets, those that include the root, i.e. (AB)C = AB/CR.

Leaf stability measures are based on the simple idea that: (1) measures of the support for a resolved quartet are indicative of the stability of the included leaves with respect to each other, and (2) an average of these stabilities/supports for all the quartets including a particular leaf is indicative of the stability of that leaf with respect to all the other leaves. As noted by Lee (2005), several different ‘simple (but seldom-used)’ leaf stability measures have been devised. All the measures are averages of different support indices for all quartets (or all triplets in rooted trees) that include a particular leaf. Thorley & Wilkinson (1999) used two bootstrap-based indices, the highest bootstrap support of any resolution of the quartet/triplet (LS_{MAX}), and the difference between the highest and the second highest bootstrap supports of the alternative resolutions of the quartet/triplet (LS_{DIF}). Lee (2005: 330) noted that ‘Thorley & Page (2000) also mention a third stability measure, the “entropy” of all possible resolutions, but do not

describe it in detail.' In information theory, entropy is a measure of randomness, that can be defined for an event as the negative sum, over all possible outcomes, of the product of the probability of each outcome times the log of the probability of that outcome (Shannon 1948). Entropy is maximal when all outcomes are equiprobable and zero when one of the outcomes has a probability of 1. For the entropic leaf stability (LS_{ENT}), defined by Thorley (2000), the probabilities of quartet resolutions (the outcomes) are given by their frequencies of occurrence in bootstrap trees (with any unresolved quartets counted as contributing equally to each possible resolution).

Each of these three LS measures is normalized to range between 0 (no support) to 1 (maximal support) and, although originally defined using bootstrap trees, analogous measures can be extracted from any set of trees (e.g. those produced in jackknife, Bayesian or quartet puzzling analyses) wherever the frequency of occurrence of a relationship gives a useful measure of the support for that relationship. Thorley (2000) also described a decay index leaf stability measure (LS_{DEC}), not mentioned by Lee (2005), which is simply the average of the highest decay indices of the quartets/triplets including a particular leaf and which, unlike the measures based on frequency of occurrence of a relationship in sets of trees, has an unbounded maximum. All these measures are implemented in RadCon (Thorley & Page 2000).

Examples

Lee (2005) used a simple hypothetical example to illustrate the use of LS measures to identify unstable taxa. All conflict in his data (Table 1) is associated with a single 'wildcard' leaf W that has a slightly suboptimal position, far from its placement in the most parsimonious tree (Fig. 1). Bootstrap-based leaf stabilities identify W as relatively unstable (Table 2) so

Table 1 Hypothetical data comprising 19 characters scored for 10 taxa (T). Lee's (2005) data comprise characters 1–13 scored for taxa O (the outgroup), A–F, and W and are successively augmented by the addition of taxon X and characters 14–16 and taxon Y and characters 17–19.

	1111	111	111
T	1234567890123	456	789
O	0000000000000	000	000
A	1100000000001	000	000
B	1111000000000	000	000
C	1111110000000	000	000
D	1111111100000	000	000
E	1111111111000	000	000
F	1111111111110	000	000
W	1100001111111	111	000
X	1100001111111	111	111
Y	1100001111111	111	111

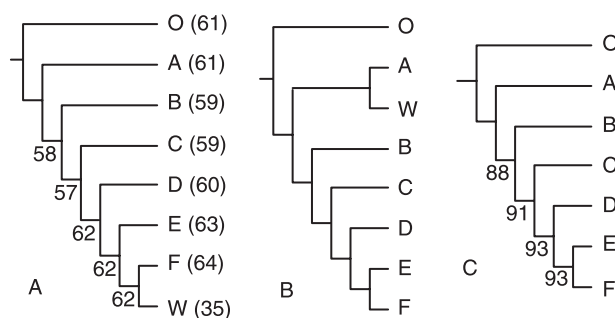


Fig. 1 A–C. —A. Most parsimonious tree for Lee's (2005) hypothetical data. —B. A slightly suboptimal tree. —C. An MRC tree constructed from the corresponding bootstrap analysis. Numbers adjacent to edges are bootstrap proportions and numbers in parentheses are leaf stabilities ($100 \times LS_{ENT}$).

Table 2 Leaf stabilities calculated with RadCon (Thorley & Page 2000) for Lee's (2005) hypothetical data and augmentations of them in Table 1, using 100 bootstrap replicates with trees treated as unrooted. Lowest ranking leaves are indicated in bold.

Lee (2005)			Lee (2005) + X			Lee (2005) + X + Y			
LS_{MAX}	LS_{DIF}	LS_{ENT}	LS_{MAX}	LS_{DIF}	LS_{ENT}	LS_{MAX}	LS_{DIF}	LS_{ENT}	
0.820	0.663	0.613	0.779	0.573	0.595	0.795	0.604	0.614	O
0.819	0.662	0.614	0.779	0.571	0.597	0.795	0.603	0.616	A
0.806	0.637	0.586	0.769	0.554	0.576	0.788	0.590	0.603	B
0.812	0.647	0.593	0.768	0.556	0.565	0.790	0.594	0.601	C
0.816	0.654	0.604	0.776	0.568	0.578	0.789	0.595	0.597	D
0.821	0.660	0.628	0.777	0.568	0.591	0.790	0.600	0.605	E
0.825	0.666	0.636	0.779	0.572	0.598	0.800	0.610	0.623	F
0.641	0.298	0.352	0.757	0.523	0.591	0.847	0.700	0.726	W
—	—	—	0.757	0.523	0.591	0.847	0.700	0.726	X
—	—	—	—	—	—	0.847	0.700	0.726	Y

that its potentially problematic use as a reference taxon can be avoided. Note that although Lee (2005: 330) describes his LS values 'as alternative measures of the stability of all triplets containing each taxon', the measures he presents are, like those given in Table 2, of quartet stabilities, corresponding to treating the input trees as unrooted, rather than the triplet stabilities that RadCon calculates for rooted trees. Had the input trees been treated as rooted on the outgroup O, as in Fig. 1, then the LS scores for O would necessarily be unity, reflecting that its phylogenetic position is assumed.

If we augment Lee's (2005) hypothetical data so as to include an additional leaf X, and sufficient characters to ensure that X and W are sister taxa (Table 1) then leaf stabilities change. Although X and W are both identified as the most unstable taxa using LS_{MAX} and LS_{DIF} , the result is far less clearcut, and it is C that is the most unstable using LS_{ENT} . This is because all quartets that include both W and X are resolved equivalently in all the bootstrap trees, improving their averages. If we add another taxon Y, scored as identical

to W and X then the pattern of relative instabilities revealed by the LS scores changes again, with W, X and Y now the most stable and either B or D the least stable as judged by the various LS measures.

The examples show that leaf stabilities can identify particularly unstable taxa in some cases, but that they do not always give a clear picture of relative stabilities of groups of leaves, and that they are sensitive to taxon sampling. Consequently, the rankings of taxa provided by leaf stabilities need not correspond closely with the utility or otherwise as referent taxa in phylogenetic nomenclature. For example, LS might suggest that W, X and Y are the best reference taxa despite their instability with respect to the remaining taxa.

Alternative approaches

Lee (2005: 330) advocated using LS measures to differentiate the most and least stable leaves because ‘the former tend to form the robust “backbone subtree”, while the latter taxa tend to be the “wildcard” taxa which move around this subtree.’ Similarly (p. 329), he claims that ‘using this approach enables one to apply names only to stable clades in the best supported “backbone” subtree within a phylogeny.’ However, if the aim is to identify the best-supported ‘backbone’ relationships in the tree, then other methods may be better suited than the heuristic use of LS.

Both the majority-rule reduced consensus (MRC) method (Wilkinson 1996; Wilkinson & Thorley 2003) and double-decay analysis (DDA) (Wilkinson *et al.* 2000) have been designed to help phylogeneticists identify well-supported relationships that may be otherwise obscured by the relative instability of a subset of leaves. In DDA, the decay indices of all triplets are calculated and summarized in a partition table of support values for all full and partial splits. Trees containing only relationships with some minimal value of support can be built up from this comprehensive assessment of the support for each triplet. The MRC method was developed to provide a more sensitive alternative to the majority-rule component consensus (Margush & McMorris 1981) that is usually employed to summarize sets of bootstrap, jackknife, Bayesian or quartet puzzling trees. Essentially, this method produces a partition table of the support for full and partial splits, and majority-rule consensus trees are constructed for each subset of leaves defined by one or more nonredundant (full or partial) split.

I used my implementation of MRC (Wilkinson 2001) to investigate the support in Lee’s (2005) example data and also in my augmented versions of his data. Table 3 summarizes support nongraphically using the partition tables generated by each analysis. Considering Lee’s (2005) data, the partition table reveals that there is much stronger support for partial splits that do not include W (rows 5, 6, 7 and 8) than for the corresponding full splits that include W (rows 11, 10, 14 and

Table 3 Partition tables providing summaries of support (%) for full and partial splits from Lee’s (2005) hypothetical data and augmentations of them using 100 bootstrap replicates. These were analysed using the REDBOOT program of REDCON 3.0 (Wilkinson 2001), with a ‘threshold for collapse’ of (5%) and trees treated as rooted. Inferences regarding the stability of particular leaves are based on the differences between the support for full splits that include those leaves and partial splits that do not. For example, comparing the partial splits 5, 6, 7 and 8 with the corresponding full splits 11, 10, 14 and 13 reveals much greater support for the partial splits which exclude the relatively unstable taxa W, X and Y.

	Lee (2005)		Lee (2005) + X		Lee (2005) + X + Y	
	OABCDEFW	%	OABCDEFWX	%	OABCDEFWXY	%
1	.*****	100.00	.*****	100.00	.*****	100.00
2	..?****?	99.33	..?****??	100.00	..?****???	99.33
3	...?***?	99.18	...?***??	99.33	...?***???	99.20
4	..?****?	98.36	..?****??	99.33	..?****???	99.20
5**?	93.05**??	92.53**???	87.14
6**?	92.63**??	91.73**???	93.75
7**?	90.83**??	90.53**???	91.01
8**?	88.53**??	94.00**???	92.53
9	..?*****	68.46	..?*****	63.92	..?*****	65.92
10	...****	62.07	...****	60.00	...****	63.08
11**	61.83**	60.00**	57.92
12**	61.83**	58.83**	63.25
13**	58.32**	58.58**	59.62
14**	57.27**	53.33**	56.62
15	—	—**	100.00**	100.00
16	—	—	—	—**	98.67

13); this comparison allows us to identify W as the most unstable leaf, in agreement with LS. Fig. 1A is the majority rule component consensus and Fig. 1C one of four other MRC trees that reveal the (otherwise obscured) support for partial splits that do not include W and support the same conclusion as to the relative stability of W. Note that there are also substantial, but less impressive, increments in support when comparing partial splits that exclude one of B, C or D (rows 2, 3, 4 and 9) to corresponding splits in which they are included (rows 8, 6, 7 and 13 respectively) and it is this instability that is reflected in the remaining MRC trees (not shown). With the augmented data, we obtain essentially similar results. Inspection of either the partition table (Table 2) or selected MRC trees (Fig. 2) reveals that W, X and Y are stable with respect to each other but relatively unstable with respect to the remaining taxa. This is a much clearer picture than that provided by LS, and one that, unlike LS, supports the intuition that W, X and Y would make poor reference taxa for a phylogenetic nomenclature, at least with respect to clades that include also B, C, or D.

Discussion

I agree with Lee (2005) that if one is engaged in phylogenetic nomenclature, choosing reference taxa so as to minimize

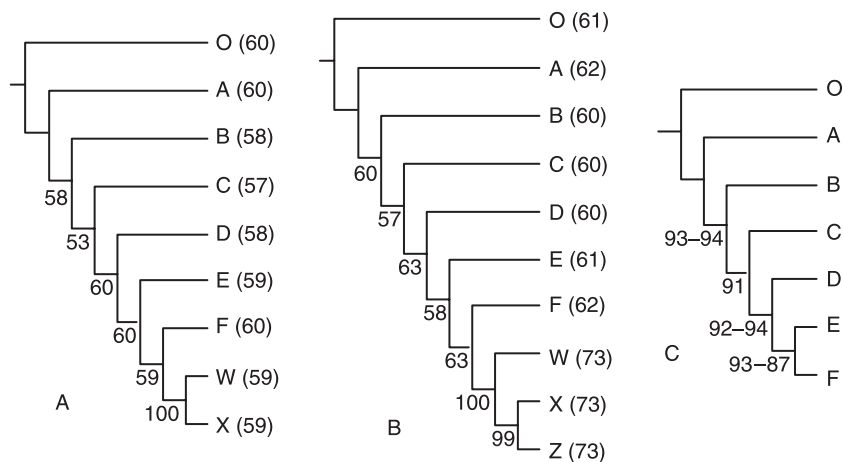


Fig. 2 A–C. Most parsimonious tree for Lee's (2005) hypothetical data augmented with one (A) or two (B) additional leaves and (C) an MRC tree constructed from the corresponding bootstrap analyses. Numbers adjacent to edges are bootstrap proportions, with multiple numbers on (C) corresponding to the analyses with one or two additional leaves; numbers in parentheses are leaf stabilities ($100 \times LS_{ENT}$).

nomenclatural instability is sensible, and that taxa that are of least certain or stable phylogenetic relationships are poorly suited to serve as reference taxa for phylogenetic nomenclature. Leaf stabilities provide a potentially useful measure for identifying unstable taxa but, as averages of triplet or quartet stabilities, they are sensitive to taxon sampling and can provide a very incomplete and potentially misleading picture of support if they are not interpreted carefully. In my view, the MRC method used to summarize sets of, say, bootstrap trees and the analogous DDA provide more promising tools for those interested in discriminating between well- and less well-supported relationships and between relatively stable and unstable leaves. Recent descriptions of both methods and discussion of some of their limitations can be found in Wilkinson (2003). There is considerable scope for further work on the problem of efficiently discovering and summarizing support and stability from sets of trees and a pressing need for better implementation of available methods.

Acknowledgements

This work was supported by BBSRC grant 40/G18385.

References

Lee, M. S. Y. (2005). Choosing reference taxa in phylogenetic nomenclature. *Zoologica Scripta*, 34, 313–328.

- Margush, T. & McMorris, F. R. (1981). Consensus n-trees. *Bulletin of Mathematical Biology*, 43, 239–244.
- Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423 and 623–656.
- Thorley, J. L. (2000). Cladistic information, leaf stability and super-tree construction. PhD Thesis, School of Biological Sciences, University of Bristol, UK. Available via <http://web.onetel.com/~joethorley/thesis.html>.
- Thorley, J. L. & Page, R. D. M. (2000). RadCon: phylogenetic tree comparison and consensus. *Bioinformatics*, 16, 486–487.
- Thorley, J. L. & Wilkinson, M. (1999). Testing the phylogenetic stability of early tetrapods. *Journal of Theoretical Biology*, 200, 343–344.
- Wilkinson, M. (1996). Majority-rule reduced consensus methods and their use in bootstrapping. *Molecular Biology and Evolution*, 13, 437–444.
- Wilkinson, M. (2001). *REDCON 3.0*, software and documentation. Department of Zoology, The Natural History Museum, London. Available via <http://www.nhm.ac.uk/research-curation/projects/software/mwphylogeny.html>.
- Wilkinson, M. (2003). Missing data and multiple trees: stability and support. *Journal of Vertebrate Paleontology*, 23, 311–323.
- Wilkinson, M. & Thorley, J. L. (2003). Reduced consensus. In M. F. Janovitz, F.-J. Lapointe, F. R. McMorris, B. Mirkin & F. S. Roberts (Eds) *Bioconsensus* (pp. 195–204). DIMCAS Series in Discrete Mathematics and Theoretical Computer Science volume 61. The American Mathematical Society.
- Wilkinson, M., Thorley, J. L. & Upchurch, P. M. (2000). A chain is no stronger than its weakest link: Double decay analyses of phylogenetic hypotheses. *Systematic Biology*, 49, 754–776.