

Branch Lengths and Support

MARK WILKINSON,¹ FRANÇOIS-JOSEPH LAPOINTE,² AND DAVID J. GOWER¹

¹Department of Zoology, Natural History Museum, London SW7 5BD, U.K.

²Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec H3C 3J7, Canada

Although technical definitions exist for various support metrics, the notion of support per se has received little explicit attention. Thus, despite its widespread use in phylogenetics, “support” is absent from the glossaries and/or indices of several recent texts (e.g., Kitching et al., 1998; Page and Holmes, 1998; Schuh, 2001). Farris et al. (2001) recently argued that interpreting branch lengths of trees as indicative of support for corresponding groups is both common and unfortunate. To support their claim, they presented two contrived examples in which “long branches do not indicate support and in fact create a highly misleading impression if so interpreted” (Farris et al., 2001:298). Here, we argue that Farris et al.’s claim presupposes a particular view of support and that their conclusions do not hold under an alternative and reasonable perspective that they did not discuss.

Farris et al.’s first example is very simple. Consider four species, A–D, and data comprising 500 characters, “of which half split the terminals AB/CD and the rest AC/BD” (Farris et al., 2001:298). There are two most-parsimonious trees, AB/CD and AC/BD. Farris et al. pointed out that the consensus of these trees is unresolved and that the decay indices (=Bremer support) of the single internal branch of each tree are therefore 0 despite these branches having a parsimony length of 250 steps. Farris et al. offered no explicit definition of support, but we assume that they view relationships as supported only if they are present in all the optimal trees and therefore in the consensus. In the context of parsimony analysis, this view corresponds to Nixon and Carpenter’s (1996) notion of strictly supported relationships, under which it is not possible for data to support alternative incompatible phylogenetic hypotheses (see also Farris, 2001). However, support need not be construed in this way.

Our preferred interpretation of Farris et al.’s first example is that the data support two alternative hypotheses equally. To say that half the characters split the terminals AB/CD and the rest AC/BD is to say that half the characters support AB/CD and half support AC/BD. This seems to be a natural way to talk about the relationship between characters and hypotheses, without which the statement that characters “split the terminals” (upon which their example depends) requires some further explication. The alternative view that data may contain multiple signals that support a complex mix of incompatible hypotheses is not uncommon (e.g., Henny and Penny, 1993; Lento et al., 1995; Wägele, 1996; Frohlich and Estabrook, 2000; Pisani and Wilkinson, 2002). Farris et al.’s “support” also appears

to be a differential or relative concept. In their example, there is no differential support provided by the complete data set for AB/CD over AC/BD or vice versa, so they contend that neither hypothesis is supported. We agree that there is no differential support for one hypothesis over the other, but this is precisely because both hypotheses are equally (and nonnegatively) supported and equally contradicted.

A problem with Farris et al.’s view becomes clear as soon as the third possible resolved tree, AD/BC, is also considered. In Farris et al.’s view, none of the hypothesis AB/CD, AC/BD, or AD/BC (Fig. 1a–c) is supported (none appears in the consensus tree, Fig. 1d), but this as a misleading and therefore unfortunate view of the relationship between the data and the hypotheses. Only for AD/BC is there no support, in the sense that there are no supporting characters. Further, there is extensive evidence against this hypothesis (the decay index of which is –500). Thus, using Farris et al.’s preferred measure, the data provide considerable evidence against AD/BC (or equivalently they support the hypothesis “not AD/BC”). This interpretation makes sense only if the data support the equivalent composite hypothesis, AB/CD or AC/BD, which does not seem coherent under Farris et al.’s conception of support.

By focusing only on relationships in the consensus, Farris et al.’s view of support promotes a simplistic representation of potentially complex patterns of support and conflict that phylogeneticists might be better encouraged to explore than ignore. Split decomposition graphs, or

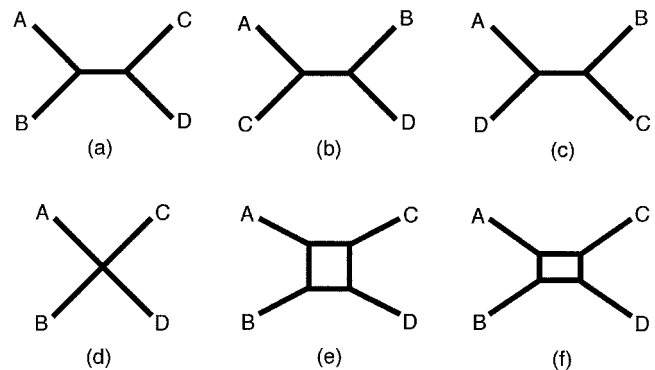


FIGURE 1. The three distinguishable unrooted trees (a, b, and c) for four species, A–D, and the strict component consensus of any two of those trees or of all three (d). The split decomposition graphs (splitsgraphs) of trees a and b (e and f) are presented for different numbers of characters supporting the competing hypotheses.

splitsgraphs (Bandelt and Dress, 1992; Dress et al., 1996; Huson, 1998) provide one means of investigating these patterns. A tree is returned by the split decomposition method only when all characters in the data support a single tree. Farris et al.'s data are represented by a splitsgraph (Fig. 1e) showing equal support for AB/CD and AC/BD. The split AD/BC is not included because it has no support. Only a splitsgraph of data that support the three possible topologies equally would be unresolved and would be identical to the consensus tree (Fig. 1d).

Consider a slight modification to the example so that now 300 characters support the split AB/CD and 200 support the split AC/BD. Under parsimony, the single best hypothesis is AB/CD because it has the best differential support. A splitsgraph of the same data (Fig. 1f) shows that the split AB/CD is better supported but that AC/BD is also supported and that BC/AD is unsupported. Branch lengths are used to indicate these relative levels of support, and there is a clear relationship between branch lengths and support in this case. Note that any trees fit into a split decomposition graph (Bandelt and Dress, 1992).

For Farris et al., AB/CD and AC/BD are unsupported by their hypothetical data despite having long branches. They contended that "This difficulty also affects significance tests that are based on branch length" (Farris et al., 2001:298) and pointed out that the internal branches of AB/CD and AC/BD are both highly significant using Rzhetsky and Nei's (1992) confidence probability test. Farris et al. (2001:298) consider that this reflects a "fundamental flaw in tests based on branch length" because these are based on the assumption that the tree is given. In their examples "the data leave doubt as to which tree is correct" so that "tests of magnitude are not adequate to demonstrate support" (Farris et al., 2001:298). We caution that this conclusion might hold for Farris et al.'s notion of support, but not for alternatives. With our interpretation, it is perfectly legitimate to identify alternative and incompatible internal edges as significantly well supported. Farris et al.'s example simply serves to demonstrate the power of the confidence probability test to identify the strong but conflicting support for relationships that they designed into (but seemingly consider to be absent from) their hypothetical data set taken as a whole.

Farris et al. (2001:298) concluded that "researchers who wish to assess support would be well advised to use methods that are more reliable in that application," and they listed consensus trees, decay indices, and character resampling as such methods. In their second example, Farris et al. (2001:298) considered hypothetical data that yielded two equally optimal trees (Fig. 2a, 2b) under maximum likelihood (ML). They claimed that "the consensus of these trees is unresolved, so that there is no support for any grouping." This is mistaken. Both trees include the split ABEF/CDGH, which is therefore present in the strict component consensus (Fig. 2c). Additionally, if the trees are considered rooted (they are presented as such but rooting per se was not discussed), another relationship (that F is more closely related to

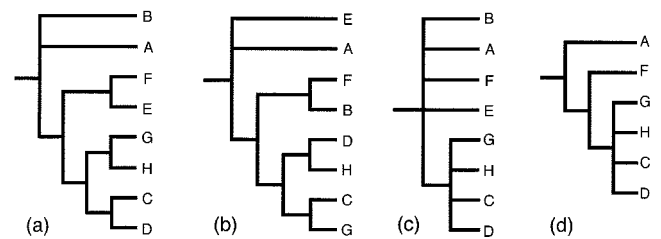


FIGURE 2. Two trees (a) and (b) modified from Farris et al. (2001) and their strict reduced consensus profile (c) and (d). Tree c is also the strict component consensus tree.

CDGH than it is to A) is also in both trees (Fig. 2d). If consensus methods are used to investigate support (sensu Farris et al.), then we recommend using sensitive methods such as greatest agreement subtrees (Gordon, 1980) and reduced consensus (Wilkinson, 1994).

Alternatively or additionally, consensus methods can be used to combine trees using branch length information (Lapointe, 1998). Both ML trees obtained by Farris et al. have internal branches of the same length, about 0.16, with the exception of the shared split ABEF/CDGH. The average consensus procedure (Lapointe and Cucumel, 1997) returns two consensus solutions. These trees have exactly the same topologies as the original ML trees, but whereas the common split ABEF/CDGH has the shortest internal branch in the ML trees (0.00006), it has the longest internal branch (0.10187) in the average consensus trees. Here, the consensus method provides an assessment of support that is indicated by relative branch lengths and that reflects the congruence of the ML trees.

Traditional decay analysis is the determination of the decay indices of all relationships present in the strict component consensus of most-parsimonious trees. It provides an indication of the relative support for clades provided by the data as a whole, in accordance with Farris et al.'s notion of support. However, there is much that traditional decay analysis does not reveal. For example, it ignores the relative support for suboptimal hypotheses that have negative decay indices of various magnitudes. Several recent extensions to traditional decay analysis have been developed to reveal hidden support from subsets of the data (Gatesy et al., 1999) and to reveal support for less inclusive relationships than clades (Wilkinson, 1997; Wilkinson et al., 2000). DeBry (2001) explored the importance of branch lengths in interpreting the significance of their associated decay indices, concluding that that raw decay indices are not comparable among trees or even within trees unless branch lengths are taken into account.

Character resampling methods such as bootstrapping and jackknifing both yield sets of trees that can be summarized with a majority-rule consensus. A more comprehensive summary is a partition table, which can include additional relationships that are present in a minority of the trees from the resampled data. The relationships in the majority-rule consensus are better supported than those that are only in the partition table, by virtue of

their greater frequency. Farris et al. did not discuss these methods, but to be consistent, they must view the latter relationships as unsupported (not in the consensus). In contrast, our interpretation is that relationships in the partition table but not in the consensus are weakly supported, and only relationships that do not occur at all in the partition table appear to be totally unsupported.

One reason for preferring our view is that information in the partition table but not in the consensus tree can be useful precisely because it allows further assessment of differential support. For example, suppose a particular clade has a bootstrap proportion (BP) of 60%. Now consider two alternatives: (1) there is a single conflicting clade with a marginally lower BP of 40%, and (2) there are 40 alternative conflicting clades each with much lower BPs of 1%. Although the BPs of the particular clade are the same in both cases, the clade is less well supported in the first case. This assessment is based upon the distribution of support among competing hypotheses, and a bootstrap difference measure based on this logic has been used in assessing the stability of individual leaves (Thorley and Wilkinson, 1999). Such considerations make no sense under Farris et al.'s more restrictive view of support. Farris (2001) distinguished between well-supported branches (those that occur in the parsimony jackknife majority-rule consensus) and unsupported branches and described groups that appear in a successive support weighting tree but not in the parsimony jackknife tree as having support that "though definitely positive, is relatively weak" (Farris, 2001:392). Avoiding such apparent contradictions is another reason for preferring our view.

Farris et al. presented nothing approaching an explicit definition of support, and we have inferred their meaning from their usage. We offer the following clarification of our usage in the hope that it will promote critical discussion and improvement. Data provide evidence for (i.e., support for) a hypothesis to the extent that the hypothesis provides a better explanation of the evidence than some alternative hypothesis. What constitutes an explanation is an important issue in philosophy of science (Salmon, 1984; Pitt, 1988) but need not concern us unduly here. Similarly, we need not get bogged down in consideration of what constitutes better explanation in any general sense (Lipton, 1991). In phylogenetics, we consider quality of explanation to be measured by some function of the fit of data to trees under some objective function or to be informed by sets of optimal trees produced by, for example, bootstrapping, jackknifing, and quartet puzzling. Our definition is loose enough to accommodate the fact that various functions and measures of fit can be used to define specific measures of support.

As with Farris et al.'s conception, our definition is a relative or differential one. It is applicable to the entire data set treated as a whole, to subsets of the data (by taxa or characters), and to individual characters. When we say that a character (or a data matrix) supports a hypothesis, we imply that it fits that hypothesis better than it fits one or more incompatible hypotheses. In typical usage, the incompatible hypotheses are not specified and

usually include all other hypotheses, but this is not always the case. For example, Wilkinson and Nussbaum (1996) discussed indirect support, in which characters support one hypothesis over some but not all incompatible alternatives.

We see no reason to adopt a notion of support that restricts use of the word to features of optimal trees or their consensus trees and that is seemingly intended to rule out any relationship between branch lengths and support by definitional fiat. Instead, we prefer a more general concept that encourages a broader view and more detailed examination of support, under which the relationship between support and branch lengths remains open to empirical and theoretical investigation.

ACKNOWLEDGMENTS

We thank Mike Steel and Daniel Huson for constructive comments.

REFERENCES

- BANDELT, H. J., AND A. W. DRESS. 1992. Split decomposition: A new useful approach to phylogenetic analysis of distance data. *Mol. Phylogenet. Evol.* 1:242–252.
- DEBRY, R. W. 2001. Improving interpretation of the decay index for DNA sequence data. *Syst. Biol.* 50:742–752.
- DRESS, A., D. HUSON, AND V. MOULTON. 1996. Analyzing and visualizing sequence and distance data using SPLITSTREE. *Disc. Appl. Math.* 71:95–109.
- FARRIS, J. S. 2001. Support weighting. *Cladistics* 17:389–394.
- FARRIS, J. S., M. KÄLLERSJÖ, AND J. E. DE LAET. 2001. Branch lengths do not indicate support—Even in maximum likelihood. *Cladistics* 17:298–299.
- FROHLICH, M. W., AND G. F. ESTABROOK. 2000. Wilkinson support calculated with exact probabilities: An example using *Floricaula/LEAFY* amino acid sequences that compares three hypotheses involving gene gain/loss in seed plants. *Mol. Biol. Evol.* 17:1914–1925.
- GATESY, J., P. O'GRADY, AND R. H. BAKER. 1999. Corroboration among data sets in simultaneous analysis: Hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* 15:271–313.
- GORDON, A. D. 1980. On the assessment and comparison of classifications. Pages 149–160 in *Analyse de données et informatique* (R. Tomassone, ed.). INRIA, Le Chesnay, France.
- HENDY, M. D., AND D. PENNY. 1993. Spectral analysis of phylogenetic data. *J. Classif.* 10:5–24.
- HUSON, D. H. 1998. SplitsTree: A program for analyzing and visualizing evolutionary data. *Bioinformatics* 14:68–73.
- KITCHIN, I. J., P. FOREY, C. J. HUMPHRIES, AND D. M. WILLIAMS. 1998. *Cladistics: The theory and practice of parsimony analysis*. Oxford Univ. Press, Oxford, U.K.
- LAPOINTE, F.-J. 1998. For consensus (with branch lengths). Pages 73–80 in *Advances in data science and classification* (A. Rizzi, M. Vichi, and H.-H. Bock, eds.). Springer-Verlag, Berlin.
- 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.
- LENTO, G. M., R. E. HICKSON, G. K. CHAMBERS, AND D. PENNY. 1995. Use of spectral analysis to test hypotheses on origin of pinnipeds. *Mol. Biol. Evol.* 12:28–52.
- LIPTON, P. 1991. *Inference to the best explanation*. Routledge, London.
- NIXON, K. C., AND J. M. CARPENTER. 1996. On consensus, collapsibility and clade concordance. *Cladistics* 12:305–321.
- PAGE, R. D. M., AND E. C. HOLMES. 1998. *Molecular evolution: A phylogenetic approach*. Blackwell Science, Oxford, U.K.
- PISANI, D. P., AND M. WILKINSON. 2002. Matrix representation with parsimony, taxonomic congruence and total evidence. *Syst. Biol.* 51:151–155.

- PITT, J. C. 1988. Theories of explanation. Oxford Univ. Press, New York.
- RZHETSKY, A., AND M. NEI. 1987. A simple method for estimating and testing minimum evolution trees. *Mol. Biol. Evol.* 9:945–967.
- SALMON, W. 1984. Scientific explanation and the causal structure of the world. Princeton Univ. Press, Princeton, New Jersey.
- SCHUH, R. 2001. Biological systematics: Principles and applications. Cornell Univ. Press, Ithaca, New York.
- THORLEY, J. L., AND M. WILKINSON. 1999. Testing the phylogenetic stability of early tetrapods. *J. Theoret. Biol.* 200:343–344.
- WÄGELE, J.-W. 1996. First principles of phylogenetic systematics, a basis for numerical methods used for morphological and molecular characters. *Vie Milieu* 46:125–138.
- 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. 1997. Characters, congruence and quality: A study of neuroanatomical and traditional data in caecilian phylogeny. *Biol. Rev.* 72:423–470.
- WILKINSON, M., AND R. A. NUSSBAUM. 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). *Copeia* 1996:550–562.
- WILKINSON, M., J. L. THORLEY, AND P. M. UPCHURCH. 2000. A chain is no stronger than its weakest link: Double decay analyses of phylogenetic hypotheses. *Syst. Biol.* 49:754–776.

First submitted 12 April 2002; reviews returned 6 July 2002;

final acceptance 26 August 2002

Associate Editor: Mike Steel