Sampling Confidence Envelopes of Phylogenetic Trees for Combinability Testing: A Reply to Rodrigo

Francois Lutzoni; F. Keith Barker

Systematic Biology, Volume 48, Issue 3 (Sep., 1999), 596-603.

Stable URL:
http://links.jstor.org/sici?sici=1063-5157%28199909%2948%3A3%3C596%3ASCEOPT%3E2.0.CO%3B2-5

of randomly distributed fossil horizons. Paleobiology 20:459–469.

MAXWELL, W. D., AND M. J. BENTON. 1990. Historical tests of the absolute completeness of the fossil record of tetrapods. Paleobiology 16:322–335.

NORELL, M. A., AND M. J. NOVACEK. 1992a. Congruence between superpositional and phylogenetic patterns: Comparing cladistic patterns with fossil records. Cladistics 8:319–337.

NORELL, M. A., AND M. J. NOVACEK. 1992b. The fossil record and evolution: Comparing cladistic and paleontologic evidence for vertebrate history. Science 255:1690–1693.

NORELL, M. A., AND M. J. NOVACEK. 1997. The ghost dance: A cladistic critique of stratigraphic approaches to paleobiology and phylogeny. J. Vertebr. Paleontol. Suppl. 17:67A.

PAUL, C. R. C. 1982. The adequacy of the fossil record. Pages 75–117 *in* Problems of phylogenetic reconstruction (K. A. Joysey and A. E. Friday, eds.). Academic Press, London.

PAUL, C. R. C. 1990. Completeness of the fossil record. Pages 293–303 *in* Palaeobiology, a synthesis (D. E. G. Briggs and P. R. Crowther, eds.). Blackwell, Oxford, England.

RAUP, D. M. 1972. Taxonomic diversity during the Phanerozoic. Science 177:1065–1071.

RIEPPEL, O. 1997. Falsificationist versus verificationist approaches to history. J. Vertebr. Paleontol. Suppl. 17:71A.

SIDDALL, M. E. 1996. Stratigraphic consistency and the shape of things. Syst. Biol. 45:111–115.

SIDDALL, M. E. 1997. Stratigraphic indices and tree balance: A reply to Hitchin and Benton. Syst. Biol. 46:569–573.

SMITH, A. B. 1994. Systematics and the fossil record. Blackwell, Oxford, U.K.

WAGNER, P. J. 1998. Phylogenetic analyses and the quality of the fossil record. Pages 165–187 *in* The adequacy of the fossil record (S. K. Donovan and C. R. C. Paul, eds.). Wiley, Chichester, England

WILLS, M. A. 1999. The gap excess ratio, randomization tests, and the goodness of fit of trees to stratigraphy. Syst. Biol. 48:559–580.

Received 18 February 1998; accepted 15 November 1998
Associate Editor: R. Olmstead

Syst. Biol. 48(3):596–603, 1999

Sampling Confidence Envelopes of Phylogenetic Trees for Combinability Testing: A Reply to Rodrigo

FRANÇOIS LUTZONI[1,4] AND F. KEITH BARKER[2,3]

[1]*Department of Botany and* [2] *Department of Zoology, Field Museum of Natural History, Roosevelt Road at Lake Shore Drive, Chicago, Illinois, 60605, USA* [3] *Department of Ecology and Evolution, Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois, 60637-1573, USA*

In 1997, Lutzoni pointed out two main caveats of a method (referred to here as RKB3) proposed by Rodrigo et al. (1993) to determine whether trees derived from different data sets are sample estimates of a parametric phylogenetic tree. The first problem was the high and undetermined number of bootstrap replicates necessary to implement correctly the second part of the RKB3 method. The second problem was the difficulty of handling the huge bootstrap tree files for the tree-to-tree comparisons needed. This second problem was most acute when analyzing data sets with differential (low versus high) resolving power. The term "resolving power" refers here to the relative number of equally most-parsimonious trees associated with a given data set.

Rodrigo (1998) acknowledged part of the first problem and proposed, using an adapted mark–capture–recapture approach, to estimate the number of bootstrap replicates necessary to adequately sample 95% of the unique trees in the "confidence envelope" surrounding the optimal tree(s). To avoid the prohibitive amount of time needed to phylogenetically analyze the extremely high number of bootstrapped data sets (e.g., > 1,000,000) that would be needed in many cases, Rodrigo (1998) proposed using distance-based methods (e.g., neighbor-joining) instead of maximum parsimony. In this reply to Rodrigo (1998), we demonstrate that estimates of the appropriate number of bootstrap replicates (*b*),

[4]Address correspondence to this author. E-mail: flutzoni@fmnh.org.

using Rodrigo's (1998) approach, can be extremely high for a given pair of data sets, incurring a significant computational load. Furthermore, these estimates of *b* are sensitive to the model of evolution used and to the resolving power of the data sets compared. We also show that the workaround solution of Rodrigo (1998) using neighbor-joining is unjustified for theoretical and practical reasons, especially when there seem to be more viable alternatives.

The Original RKB3 Method

The RKB3 method proposed by Rodrigo et al. (1993) consisted of a series of three tests, the last two each being contingent on the result of the previous test. The first test of the RKB3 method addresses whether the best trees from two data sets are more similar than pairs of random trees sampled from the universe of all possible bifurcating unrooted trees for a given number of OTUs. Only if the observed symmetric-difference is significantly smaller than expected by chance is it logical to proceed to the second test of the RKB3 method.

The second test of the original RKB3 method was designed to determine if the "confidence envelope" (as used by Rodrigo, 1998) associated with trees derived from the two original data sets were overlapping. Rodrigo et al. (1993) proposed using bootstrap resampling to generate trees for this "confidence envelope." If no trees were common to both bootstrap tree profiles, then the null hypothesis—that trees derived from the original data sets are estimates of the same true tree—is rejected and there is no need to proceed to the third test of the RKB3 method. If at least one topology is shared by the two bootstrap tree files, the null hypothesis is not rejected and the third test can be implemented. Rodrigo (1998) used the term "bootstrap profile" interchangeably with "confidence envelope." We believe that the former term is much more appropriate. Estimating confidence limits on entire topologies is complex, and which criterion should be used to delimit confidence sets is not yet resolved (Sanderson, 1989; Sanderson and Donoghue, 1989; Felsenstein and Kishino, 1993; Hillis and Bull, 1993). For this reason, we do not use "confidence envelope" in this paper when referring to this specific part of the RKB3 method.

The null hypothesis of the third test of the RKB3 method is that the observed symmetric-difference between trees derived from the two different data sets is not different from symmetric-differences expected when comparing trees obtained from two data sets known to sample the same phylogenetic history. The null distribution of this third test, which consists of symmetric-differences between trees derived from two bootstrap data sets resampled from the same original data set, produces two null distributions, one for each of the two data sets. If the observed symmetric-difference is > 95% of the symmetric-differences among trees within the two bootstrap profiles, the null hypothesis is rejected. In such a case, Rodrigo et al. (1993) suggested going back to the original data set to identify the OTU(s) causing the conflict, and to remove them one by one from the two data sets, reimplementing the three tests after each exclusion. This procedure is repeated until the null hypothesis is not rejected. Once this was achieved, Rodrigo et al. (1993) recommended that trees derived from the two series of bootstrapped data sets be summarized by using least-squares consensus rather than combining the data sets and performing a new search.

First Modification of the RKB3 Method

In 1995, Lutzoni and Vilgalys performed an empirical comparison of three methods that had been proposed by Bull et al. (1993) as potentially useful in assessing whether different data sets are samples of the same phylogenetic history or not (i.e., homogeneous vs. heterogeneous data sets). The three tests compared were the RKB3 method by Rodrigo et al. (1993), an adapted version of Faith's (1991) T-PTP test, and Kishino and Hasegawa's (1989) likelihood test. At the time when Lutzoni and Vilgalys (1995) compared these methods, they had never been specifically used as tests of homogeneity within the paradigm of conditional data combination (Bull et al., 1993), and some changes in procedure were required. For the RKB3 procedure, Lutzoni and Vilgalys (1995) suggested implementing a new search on the combined data sets when the hypothesis of a shared parametric tree could not be rejected, rather than summarizing trees from individual data sets via a consensus method (as suggested in Rodrigo et al., 1993). This decision was based on results suggesting that, when presented with individual data sets that do not have significantly different phylogenetic signals, reimplementing a phylogenetic search on these data sets pooled together is more likely to converge on the correct tree than is analyzing data sets separately and combining the resulting trees by using a consensus method (Miyamoto, 1985; Kluge, 1989; Barrett et al., 1991; Bull et al., 1993).

To some extent, we agree with Rodrigo (1998) that "pooling different data types can easily present problems." This will always be true for inherently incombinable data types, such as DNA hybridization distances and discrete character data. However, when data sets are combinable in a single analysis, we recommend that they be analyzed together, provided the data sets are found to be homogeneous. Within a parsimony framework, programs such as PAUP* 4.0 (Swofford, 1998) offer significant flexibility to the user, accommodating heterogeneity in the properties of different data sets through character and character-state weighting schemes for as many different partitions of the combined data set as the user wants to recognize. Likewise, simultaneous maximum likelihood analysis of nucleotide sequence data with mixed models of sequence evolution (Yang, 1996) has been implemented, although the available search algorithms are inefficient (Yang, 1998). This situation is likely to improve as more researchers seek to increase the sophistication of their phylogenetic analyses.

Number of Bootstrap Replicates and Neighbor-Joining

Number of Bootstrap Replicates Needed for the Second Part of the RKB3 Method

In addition to proposing a modification to the RKB3 protocol to combine data sets when they are found to be homogeneous, rather than combining the derived phylogenetic trees, Lutzoni and Vilgalys (1995) pointed out a problem associated with the second test of RKB3. They found that, when dealing with small numbers of replicates, the number of shared trees depends on the number of bootstrap replicates. Therefore, they suggested that simulation studies were needed to explore the "bias" of Rodrigo's method associated with the number of bootstrap replicates. We agree with Rodrigo (1998) that use of the term "bias" by Lutzoni (1997; Lutzoni and Vilgalys, 1995) to qualify the second part of the RKB3 method was inappropriate, because the number of resampling replicates was the issue. The reason Lutzoni (1997) used this qualifier was that when the RKB3 method was first introduced, Rodrigo et al. (1993) rejected the null hypothesis that bootstrap profiles derived from their two exemplar data sets were overlapping. Their decision was based on only 100 bootstrap replicates and was followed by removing taxa from their data sets without ever addressing the possibility that the rejection of the null hypothesis could be attributable to the low number of bootstrap replicates used.

In 1997, Lutzoni was confronted with a special case associated with the number of bootstrap replicates used in the second test of the RKB3 protocol. This involved the comparison of one data set with high resolving power and one with low resolving power. As demonstrated in Table 4 of Lutzoni (1997), this kind of case required by far the highest number of bootstrap replicates to determine whether two bootstrap profiles overlap or not. This is because the bootstrap profile for the tree(s) derived from the data set with high resolving power will be very small compared with the bootstrap profile associated with the data set with low resolving power. The degree of overlap between the two bootstrap profiles will be determined mostly by the smaller of the two profiles, the size of which is expected to be only a minute fraction of the larger bootstrap profile. The probability that the trees from the small bootstrap profile will be recovered by any single bootstrap replicate from the large bootstrap profile is minute and decreases as the number of OTUs or the disparity in resolving power between the two data sets increases. The essential problem is that the probability of a type I error occurring in part two of the RKB3 method is dependent on the number of bootstrap replicates used. The number of replicates (or time needed to do them) required to reduce this error to an acceptable level can be prohibitively high (or long). Moreover, the tree files containing bootstrap profiles for data sets with low resolving power are too large and difficult to handle on most computers.

Estimating Profile Size: A Potential Solution to the Prohibitively High Number of Bootstrap Replicates Needed for the Second Part of the RKB3 Method?

Rodrigo (1998) suggested a potential solution for the difficulty encountered by Lutzoni (1997). Specifically, he suggested estimation of the size of bootstrap profiles via a resampling technique. This procedure would place an upper limit on the number of bootstrap replicates required to achieve a given degree of certainty. As an example, he estimated the minimum number of bootstrap replicates necessary to sample 95% of the unique trees in the bootstrap profile (a measure of the type I error of the test) of one of the data sets published by Lutzoni (1997). Unfortunately, he used the full "pruned" 25S nrDNA data set of Lutzoni (1997), which had the highest resolving power and therefore the smallest bootstrap profile (thus requiring a smaller number of replicates to achieve an acceptable type I error rate). Even when Rodrigo (1998) used the "ideal" case, involving a data set with high resolving power and a relatively small bootstrap profile, he concluded that >1,350,000 bootstrap replicates (*b*) would be needed to guarantee sampling 95% of the unique trees in the profile for this data set of 30 species. As acknowledged by Rodrigo (1998), generation of this many bootstrap replicates would be prohibitively costly in terms of the computer time required, especially with processor-intensive techniques such as maximum likelihood. As a potential workaround to this problem, Rodrigo (1998) suggested the use of distance-based tree reconstruction methods, such as neighbor-joining.

Using neighbor-joining with the F81 model (Felsenstein, 1981), as Rodrigo (1998) did, we reestimated the number of bootstrap replicates necessary for this specific case (see the H subsection of the F81 section of Table 1). Our best estimate of *b* was 7.3684×10^5, a value much smaller than the 1.35×10^6 of Rodrigo (1998). We do not know the reason for this large difference between Rodrigo's (1998) and our estimates of *b* for the H case. Nevertheless, what we want to demonstrate here is that *b* will change greatly depending on which evolutionary model is chosen and the resolving power of the data sets compared. For example, using the more general HKY85 model (Hasegawa et al., 1985) with heterogeneous rates across sites following a gamma distribution (see the H subsection of the HKY85 + Γ section of Table 1) requires a much larger number of bootstrap replicates ($b = 2.4954 \times 10^7$ replicates, compared with 7.3684×10^5 by the F81 model for the same H comparison). This effect is less pronounced for the low-resolution data (Table 1). From the empirical evidence (results not shown), we expect that *b* is also sensitive to the optimization criterion and branch-swapping algorithm used. Therefore, it is important to be aware that estimates of the number of bootstrap replicates required to sample 95% of all unique trees need to be evaluated for different types of optimization criteria, branch-swapping algorithms, and evolutionary models. These three factors are part of the complexity associated with establishing a valid criterion to delimit confidence sets around optimal topologies.

As mentioned earlier, the example used by Rodrigo (1998) was based on a case where both data sets had a relatively high resolving power (H, Table 1). The comments by Lutzoni on the RKB3 method in his 1997 paper addressed a specific case in which one of the two data sets had a relatively low resolving power compared with the other (Table 2). Tables 1 and 2 demonstrate, for the same "pruned" 25S nrDNA data set, and using neighbor-joining with the F81 model, how the resolving power of data sets influences

TABLE 1. Estimated number of unique trees in bootstrap profiles (N) and estimated number of bootstrap replicates needed to sample 95% of the trees that are part of the bootstrap profile (b) associated with analyses under different evolutionary models, and data sets with different resolving power, using the adapted mark–capture–recapture approach presented by Rodrigo (1998). All estimates were obtained by using Lutzoni's (1997) pruned 25S nrDNA data set.

Comparison	No. of bootstrap replicates (n)	No. of condensed trees in the first bootstrap set (m_1)	No. of condensed trees in the second bootstrap set (m_2)	No. of unique overlapping trees (r)	Estimated no. of unique trees in bootstrap profile (N)	Estimated no. of bootstrap replicates needed (b)
Neighbor-joining with F81 Model						
H[a]	100	100	100	0	NA	NA
	500	498	498	3	8.2668×10^4	2.4864×10^5
	1,000	994	992	9	1.0956×10^5	3.3053×10^5
	2,000	1,979	1,986	25	1.5721×10^5	4.7512×10^5
	3,000	2,969	2,970	57	1.5470×10^5	4.6820×10^5
	4,000	3,937	3,933	90	1.7205×10^5	5.2392×10^5
	10,000	9,628	9,705	393	2.3776×10^5	7.3684×10^5
L[b]	100	100	100	0	NA	NA
	500	500	500	0	NA	NA
	1,000	1,000	1,000	1	1.0000×10^6	2.9957×10^6
	2,000	2,000	2,000	1	4.0000×10^6	1.1983×10^7
	3,000	3,000	2,999	4	2.2493×10^6	6.7393×10^6
	4,000	3,997	3,996	4	3.9930×10^6	1.1972×10^7
	10,000	9,987	9,993	24	4.1583×10^6	1.2470×10^7
	100,000	98,801	98,932	1,829	5.3442×10^6	1.6193×10^7
Neighbor-joining with HKY85 + Γ						
H[a]	100	100	100	0	NA	NA
	500	500	500	0	NA	NA
	1,000	1,000	1,000	0	NA	NA
	2,000	1,999	2,000	0	NA	NA
	3,000	2,999	3,000	1	8.9970×10^6	2.6957×10^7
	4,000	3,999	3,999	1	1.5992×10^7	4.7920×10^7
	10,000	9,999	9,993	12	8.3267×10^6	2.4954×10^7
L[b]	100	100	100	0	NA	NA
	500	500	500	0	NA	NA
	1,000	1,000	1,000	0	NA	NA
	2,000	1,999	1,999	1	3.9960×10^6	1.1977×10^7
	3,000	3,000	3,000	0	NA	NA
	4,000	3,999	3,998	3	5.3293×10^6	1.5971×10^7
	10,000	10,000	10,000	0	NA	NA
	100,000	99,503	99,444	857	1.1546×10^7	3.4772×10^7

[a] H = complete data set (1,264 sites), with high resolving power.

[b] L = half of the data set, with low resolving power.

the number of bootstrap replicates needed to detect overlapping trees. When comparing data sets with high resolving power (in this case, replicates of the same data set H in Table 1), overlapping trees were found with only 500 bootstrap replicates, and the estimated number of replicates required for 95% sampling of the bootstrap profile of the data set was 7.3684×10^5. When comparing data sets with low resolving power (in this case, replicates of the same data set L in Table 1), 1,000 bootstrap replicates were needed to find overlapping trees, and the estimated number of replicates required for

TABLE 2. Actual number of trees required to detect overlapping profiles in comparisons of data sets with high and low resolving power (where the data set with low resolving power is a jackknifed subset of that with high resolving power). The estimated number of replicates required for 95% sampling of both profiles was 1.6930×10^7 for neighbor-joining with the F81 model, and 5.9726×10^7 with the HKY85 model with Γ.

No. of bootstrap replicates (n)	No. of condensed trees in the first bootstrap set (m_1)	No. of condensed trees in the second bootstrap set (m_2)	No. of unique overlapping trees (r)	% of required sampling achieved
Neighbor-joining with F81 model				
100	100	100	0	1.18×10^{-3}
500	500	500	0	1.00×10^{-2}
1,000	994	1,000	0	1.00×10^{-2}
2,000	1,988	2,000	0	2.00×10^{-2}
3,000	2,962	2,999	0	4.00×10^{-2}
4,000	3,934	3,996	0	5.00×10^{-2}
10,000	9,677	9,983	3	1.20×10^{-1}
100,000	98,801	87,043	232	1.18×10^{0}
Neighbor-joining with HKY85 + Γ				
100	100	100	0	3.35×10^{-4}
500	499	500	0	1.67×10^{-3}
1,000	999	1,000	0	3.35×10^{-3}
2,000	1,997	2,000	0	1.00×10^{-2}
3,000	2,995	2,999	0	1.00×10^{-2}
4,000	3,985	3,999	0	1.00×10^{-2}
10,000	9,932	9,999	0	3.00×10^{-2}
100,000	94,930	99,476	65	3.30×10^{-1}

95% sampling of the bootstrap profile of the data set was 1.6193×10^7. However, when comparing the data set with low resolving power and that with high resolving power, 10,000 bootstrap replicates were needed to detect overlapping trees (Table 2). The difference between the estimates of b for H (2.4954×10^7) and L (3.4772×10^7) was less pronounced when using the HKY85 model with heterogeneous rates across sites following a gamma distribution (Table 1), although 100,000 replicates were required to find overlap between profiles from data sets with high and low resolving power (Table 2).

If we accept that the number of bootstrap replicates required for sampling 95% of the bootstrap profile b is sensitive to the phylogenetic optimization criterion, the branch-swapping algorithm, and the assumptions of the evolutionary model used, the question then becomes: Which optimization criterion and evolutionary model should be used to evaluate whether two data sets are sampling the same parametric tree? We think it most logical that the same optimization criterion, branch-swapping algorithm, and evolutionary model judged most appropriate for the phylogenetic analysis of the original data should also be used to explore the bootstrap profile associated with the best trees. In this context, the suggestion by Rodrigo (1998) to use neighbor-joining to accelerate the implementation of the second part of the RKB3 method is very disconcerting, because this would also dictate which optimization criterion and evolutionary model systematists should use for the phylogenetic analysis of their original data.

Progress in DNA sequencing technology is resulting in rapid increases in the size of molecular data matrices. As shown here, the number of bootstrap replicates needed to compare two data sets—one with low and the other with high resolving power—is already impracticable for only 30 taxa (Table 1). How many bootstrap replicates will be required to properly implement the second part of the RKB3 method in the same circumstance but for >60 taxa? In defense of the high number of bootstrap replicates that are likely to be needed to sample 95% of all unique trees, Rodrigo (1998) proposed an incremental procedure in which (1) both bootstrap profiles are generated with initially small numbers of bootstrap replicates, (2) the tree files are compared, and (3) if there is no overlap, they are incremented with more replicates until either an overlap is detected or the predetermined number of replicates is reached. This approach is the most efficient way to implement the second part of the RKB3 method. In Table 2—a "best case" scenario, in which the data set with low resolving power is a formal subset of that with high resolving power—only 3.30×10^{-1}% of the total number of bootstrap replicates required for sampling 95% of the trees in the two profiles (calculated from Table 1) was needed to find an overlap. This result agrees with the intuition expressed by Rodrigo (1998), regarding the probability that a much lower number of replicates than that required for 95% confidence may suffice to discover overlap between profiles. However, this is likely to be true only when two data sets share an underlying parametric tree. When data sets are samples of different phylogenetic histories, it will be necessary to generate and compare all of the replicates required to sample 95% of the bootstrap profiles in order to reject the null hypothesis of homogeneity with a reasonable error rate.

Even if using a distance tree-generation technique were an acceptable alternative method for producing bootstrap profile trees, we found that the time required for the calculation of symmetric-differences among trees within a profile (for the purpose of finding the unique trees), and between profiles (for the purpose of finding overlap), was fairly significant for the 100,000 bootstrap replicates case (~1.5 h on a Macintosh G3). Because this calculation burden will increase as $0.5(n^2 - n)$, where n = the number of trees compared, the time required for tree comparisons should quickly overtake the time required for the bootstrap tree searches as the most computationally intensive portion of the test.

We do agree with Rodrigo (1998; Rodrigo et al., 1993) that doing something like part 2 of the RKB3 method is important. It was designed to prevent accepting the null hypothesis—that the observed symmetric-difference is not significantly different from what is expected when two data sets are known to be samples of the same phylogenetic history—when two bootstrap profiles do not overlap (see Rodrigo, 1998: Figs. 1 and 2). It seems to us there is a more sensible way to determine whether a given set of trees is part of the "confidence envelope" associated with the optimal tree(s) derived from another data set. This can perhaps be tested using the Kishino and Hasegawa (1989) maximum likelihood test, the Templeton (1983) test using maximum parsimony, or an improved version of these tests (N. Goldman, pers. comm.). If at least one of the trees that is not significantly different from the optimal tree(s) from the data set with the highest resolving power also is found to be not significantly worse than one of the optimal tree(s) from a different data set, this means that the variance associated with each data set is overlapping. Both tests can be quickly implemented in PAUP* 4.0 (Swofford, 1998) and do not require producing the huge tree files that will often be required by the second test of the RKB3 protocol.

We hope that in this paper we have demonstrated that Lutzoni had already, as early as 1995, suggested that inadequate bootstrap sampling was a problem with the second part of the RKB3 method, and that the proposed modification by Rodrigo (1998) is still not a valid solution to this problem. Too much variation is associated with the estimate of b. Even with neighbor-joining, it is unrealistic to implement part 2 of the RKB3 method for data sets with a high number of OTUs, that have different levels of resolution, or that are heterogeneous. Although we agree that skipping the second test is not an ideal solution, when one of the two data sets compared has low resolving power and generates low bootstrap values, we believe that the computational costs associated with executing the test properly far outweigh the benefits, especially when potentially better alternatives, such as those suggested here, exist. Our work in progress aims to propose a more general combinability testing approach that will not be so likely to fail when dealing with different degrees of resolution or homoplasy intrinsic to different data sets.

Acknowledgments

We thank Kathleen Pryer, Allen Rodrigo, and one anonymous reviewer for their critical comments. This work was supported in part by NSF grant DEB-9615542 to FL.

References

Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. Syst. Zool. 40:486–493.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Faith, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. 40:366–375.

Felsenstein, J. 1981. Evolutionary trees from DNA sequence: A maximum likelihood approach. J. Mol. Evol. 17:368–376.

Felsenstein, J., and H. Kishino. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. Syst. Biol. 42:193–200.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

Lutzoni, F. 1997. Phylogeny of lichen- and non-lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. Syst. Biol. 46:373–406.

Lutzoni, F., and R. Vilgalys. 1995. Integration of morphological and molecular data sets in estimating fungal phylogenies. Can. J. Bot. 73(suppl. 1):S649–S659.

Miyamoto, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1:186–189.

Rodrigo, A. G. 1998. Combinability of phylogenies and bootstrap confidence envelopes. Syst. Biol. 47:727–733.

Rodrigo, A. G., M. Kelly-Borges, P. R. Bergquist, and P. L. Bergquist. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N. Z. J. Bot. 31:257–268.

Sanderson, M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. Cladistics 5:113–129.

Sanderson, M. J., and M. J. Donoghue. 1989. Patterns of variation in levels of homoplasy. Evolution 43:1781–1795.

Swofford, D. L. 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and apes. Evolution 37:221–244.

Yang, Z. 1996. Maximum-likelihood models for combined analysis of multiple sequence data. J. Mol. Evol. 42:587–596.

Yang, Z. 1998. Phylogenetic analysis by maximum likelihood (PAML), Version 1.4. University College, London.

Received 22 December 1998; accepted 5 March 1999
Associate Editor: R. Olmstead