War and Peace in Phylogenetics: A Rejoinder on Total Evidence and Consensus

CLAUDINE LEVASSEUR AND FRANÇOIS-JOSEPH LAPOINTE

*Département de sciences biologiques, Université de Montréal, C.P. 6128, Succursale centre-ville, Montréal, Québec, H3C
3J7, Canada; E-mail: claudine.lavasseur@umontreal.ca, Francois-Joseph.Lapointe@umontreal.ca*

Abstract.—For more than 10 years, systematists have been debating the superiority of character or taxonomic congruence in phylogenetic analysis. In this paper, we demonstrate that the competing approaches can converge to the same solution when a consensus method that accounts for branch lengths is selected. Thus, we propose to use both methods in combination, as a way to corroborate the results of combined and separate analyses. This so-called "global congruence" approach is tested with a wide variety of examples sampled from the literature, and the results are compared with those obtained by standard consensus methods. Our analyses show that when the total evidence and consensus trees differ topologically, collapsing weakly supported nodes with low bootstrap support usually improves "global congruence." [Average consensus; combined analysis; global congruence; separate analysis; total evidence.]

For more than a decade, phylogeneticists have been searching for ways to analyze the ever-increasing amount of data (for review, see de Queiroz et al., 1995; Huelsenbeck et al., 1996). The same question has been raised time and time again: Is it better to combine different datasets before phylogenetic reconstruction or not? With the recent advances in and increasing popularity of molecular systematics, this debate opposing total evidence (character congruence) against consensus (taxonomic congruence) approaches has become even more important (e.g., Bull et al., 1993; Chippindale and Wiens, 1994). Supporters of character congruence (sensu Mickevich, 1978) claim that all data should always be combined for phylogenetic analysis (Kluge, 1989, 1998; Barrett et al., 1991; Kluge and Wolf, 1993). On the other hand, proponents of taxonomic congruence (sensu Mickevich, 1978) insist that independent datasets should be analyzed separately and combined by means of consensus techniques a posteriori (Swofford, 1991; Bull et al., 1993; Huelsenbeck et al., 1994; Miyamoto and Fitch, 1995). For others, these competing options are too radical and an intermediate solution has been proposed to decide whether or not to combine data, based on the results of statistical heterogeneity tests (Mickevich and Farris, 1981; Rodrigo et al., 1993; Farris et al., 1995; Huelsenbeck and Bull, 1996).

Numerous studies have declared character congruence to be superior to consensus (see Miyamoto, 1985; Barrett et al., 1991; de Queiroz, 1993), usually providing trees that are more resolved than those obtained by taxonomic congruence. However, consensus methods that consider branch lengths (see Lapointe, 1998a) have been suggested to do better than those based on topological relationships alone, including the strict (Sokal and Rohlf, 1981) and majority rule (Margush and McMorris, 1981) consensus. In particular, the average consensus procedure (Lapointe and Cucumel, 1997) may be more likely than standard consensus methods to produce trees as well resolved as those obtained from total evidence analysis (Lapointe, 1998b; Lapointe et al., 1999).

We could engage in this debate by opting for character congruence, taxonomic congruence, or the conditional combination approach. Rather, we prefer to suggest using combined and separate analyses jointly, as proposed by de Queiroz (1993; see also Larson, 1994). Interestingly, a distance-based procedure relying on the average consensus has been applied successfully by Lapointe et al. (1999) to combine either trees or data matrices in a coherent fashion. This hybrid procedure is defined as a global congruence approach (see Lapointe, 1998b) because it assesses neither the congruence among characters nor that among individual phylogenies; rather, it evaluates the congruence between total evidence *and* consensus trees. This approach can thus be used to cross-corroborate the trees obtained by combined and separate analyses.

In the present paper, we apply the so-called global congruence approach to a wide variety of published datasets sampled from the systematic literature, using a uniform

distance-based procedure (Lapointe et al., 1999). We postulate that (1) total evidence and consensus trees will be congruent when average consensus is used to combine the trees estimated separately from individual datasets; (2) average consensus trees will be more similar to total evidence trees and more resolved than strict and majority rule consensus trees; and (3) when average consensus and total evidence trees differ, the discrepancies will not hold if the clades with low bootstrap support in the total evidence tree are collapsed.

MATERIALS AND METHODS

To test our hypotheses, a diversity of datasets evolving under different models and at different rates of evolution was required. To obtain them, we surveyed all papers published in *Systematic Biology* since Kluge's (1989) seminal paper and selected all those that used multiple character sets. Our initial selection included 26 studies representing a wide variety of taxonomic groups and different types of characters; the number of datasets per study ranged from 2 to 17 and the number of taxa from 9 to 193. From that list, a secondary selection was made according to data availability, which yielded 15 distinct studies with complete character sets. However, because our objectives were quite different from those of the original papers, some datasets were modified before the analyses. In specific cases, the taxa that were not represented in all of the original datasets were deleted, reducing the total number of taxa. In other cases, removing character sets defined for a reduced number of taxa allowed us to proceed with a larger total number of taxa. In each situation, the decision to delete taxa or characters was always made so as to maximize the number of datasets representing the largest possible collection of common taxa. For example, the paper by Mason-Gamer and Kellogg (1996) originally included 41 taxa and four datasets. All comparisons between sets were computed in a pairwise fashion in that publication, and the different combinations did not include all taxa. By considering only taxa for which information was available for all datasets, we reduced the number of common taxa to six. The final list and details about the selected studies are presented in Table 1.

Character sets were converted to distance matrices for both types of analyses, to be able to combine trees or data in a similar way (Lapointe et al., 1999). To do so, uncorrected (p) distances were computed for sequence data, and mean character differences were calculated for any other types of data (all computations were made with PAUP*; Swofford, 1999). In the case of combined analyses, distances were computed by using all characters at the same time. For separate analyses, distance matrices were computed independently from each individual set. Phylogenetic trees based on combined or separate data were obtained by using an unweighted least-squares method (Cavalli-Sforza and Edwards, 1967) implemented in PAUP* (Swofford, 1999). We then applied a bootstrap procedure to total evidence trees, using the same least-squares method and a fast stepwise addition, for 100 replicates. All weakly supported clades (i.e., those with bootstrap support <50%) were collapsed in the total evidence trees.

To compute average consensus trees (Lapointe and Cucumel, 1997), we recorded the pathlength distance matrices corresponding to the trees derived from the separate datasets. Average pathlength distances were then computed and submitted to a least-squares estimation procedure to construct the consensus solution. The resulting average consensus tree minimizes the sum-of-squared distances to the original phylogenies. To compare the average consensus with other consensus methods that ignore branch lengths, we derived strict (Sokal and Rohlf, 1981) and majority rule (Margush and McMorris, 1981) consensus trees directly from the individual least-squares trees computed in PAUP* (Swofford, 1999).

An important criterion for evaluating trees, particularly consensus trees, is the level of resolution of those trees. A simple way to measure resolution is to count the number of internal branches in a tree (this is the fork index in PAUP*; Swofford, 1999). The relative resolution of a tree is computed as the ratio of the number of internal branches to the maximum possible number of internal branches (i.e., $n - 3$ for unrooted trees); this computation allows one to compare the resolution of trees bearing different numbers (n) of species. Likewise, the relative resolution of bootstrap trees can be expressed as the ratio

TABLE 1. Results and summary of the 15 examples analyzed as part of this study.

References	Data No.[a] taxa	Data No.[a] sets	Data Type[b] of data	Total evidence BS[c] support	Total evidence Original study[d]	Consensus resolution Ave	Consensus resolution Maj	Consensus resolution Strict	Global congruence before bootstrap[e] Ave	Global congruence before bootstrap[e] Maj	Global congruence before bootstrap[e] Strict	Global congruence after bootstrap[f] Ave	Global congruence after bootstrap[f] Maj	Global congruence after bootstrap[f] Strict
Kluge (1989)	10	2	2,6	0.86	y	1.00	0.29	0.29	0.86	0.29	0.29	0.86	0.29	0.29
	(10)	(2)							*0.86*	*1.00*	*1.00*	*1.00*	*0.86*	*0.86*
Olmstead and Sweere (1994)	18	3	1,3	0.53	n	1.00	0.73	0.13	0.73	0.67	0.13	0.53	0.53	0.13
	(18)	(3)							*0.73*	*0.93*	*1.00*	*1.00*	*0.73*	*0.53*
Omland (1994)	9	2	1,2	1.00	y	1.00	0.67	0.67	1.00	0.67	0.67	1.00	0.67	0.67
	(9)	(2)							*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*
Mason-Gamer and Kellogg (1996)	6	4	1	0.67	n	1.00	1.00	0.33	0.67	1.00	0.33	0.67	0.67	0.33
	(41)	(4)							*0.67*	*1.00*	*1.00*	*1.00*	*1.00*	*0.67*
Pennington (1996)	27	2	1,2	0.71	y	1.00	0.13	0.13	0.75	0.13	0.13	0.63	0.13	0.13
	(27)	(2)							*0.75*	*1.00*	*1.00*	*0.92*	*0.71*	*0.71*
Baker and DeSalle (1997)	17	8	1	1.00	y	1.00	0.57	0.00	1.00	0.57	0.00	1.00	0.57	0.00
	(17)	(8)							*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*1.00*
Lutzoni (1997)	30	4	1	0.70	n	1.00	0.22	0.07	0.63	0.22	0.07	0.56	0.22	0.07
	(30)	(4)							*0.63*	*1.00*	*1.00*	*0.81*	*0.70*	*0.70*
Baum et al. (1998)	10	4	1,2,3	0.57	y	1.00	0.14	0.00	0.57	0.14	0.00	0.14	0.14	0.00
	(18)	(4)							*0.57*	*1.00*	*1.00*	*0.57*	*0.57*	*0.57*
Cannatella et al. (1998)	10	5	1,2,4,5	0.86	y	1.00	1.00	0.14	1.00	1.00	0.14	0.86	0.86	0.14
	(10)	(5)							*1.00*	*1.00*	*1.00*	*1.00*	*1.00*	*0.86*
Messenger and McGuire (1998)	26	4	1,2	0.87	y	1.00	0.74	0.30	0.83	0.74	0.30	0.78	0.70	0.30
	(56)	(4)							*0.83*	*1.00*	*1.00*	*0.91*	*0.91*	*0.87*
Flook et al. (1999)	33	3	1	0.70	n	1.00	0.60	0.30	0.63	0.50	0.30	0.53	0.43	0.30
	(35)	(3)							*0.63*	*0.87*	*1.00*	*0.83*	*0.73*	*0.70*
Gatesy et al. (1999a)	12	7	1	0.44	y	1.00	0.67	0.22	1.00	0.67	0.22	0.44	0.33	0.22
	(13)	(17)							*1.00*	*1.00*	*1.00*	*1.00*	*0.78*	*0.56*
Liu and Miyamoto (1999)	26	3	1,2	0.52	y	1.00	0.43	0.22	0.48	0.43	0.22	0.35	0.35	0.22
	(35)	(3)							*0.48*	*1.00*	*1.00*	*0.61*	*0.61*	*0.52*
Quicke and Belshaw (1999)	30	4	1,2,7	0.67	y	1.00	0.44	0.22	0.78	0.44	0.22	0.56	0.44	0.22
	(33)	(4)							*0.78*	*1.00*	*1.00*	*0.85*	*0.67*	*0.67*
Springer et al. (1999)	11	8	1	0.75	y	1.00	0.50	0.00	0.75	0.38	0.00	0.75	0.38	0.00
	(11)	(8)							*0.75*	*0.75*	*1.00*	*1.00*	*0.75*	*0.75*

[a] Number in original study are in parenthesis.
[b] 1: molecular sequences 2: morphology; 3: restriction sites; 4: allozymes; 5: calls; 6: lipid characters; 7: life history.
[c] No of clades with bootstrap value >50%/total no of clades.
[d] Congruence with original study.
[e] Strict (*Cmin*) and semistrict (*Cmax*) congruence. *Cmax* values are in italics.
[f] Strict (*Cmin'*) and semistrict (*Cmax'*) congruence. *Cmax'* values are in italics.

883

of strongly supported clades (i.e., with bootstrap support >50%) to the maximum number of clades in a tree.

When trees need to be compared with one another, one can use consensus trees and indices (Rohlf, 1982; Shao and Sokal, 1986). For example, the global congruence of combined and separate analyses can be visualized with a global consensus tree (sensu Lapointe et al., 1999) bearing the clades corroborated by the different approaches. The relative resolution of that global consensus thus indicates topological agreement among the trees compared (a value of 1 means the trees are identical). Different consensus methods can be applied to assess global congruence. Namely, the strict consensus can be used to measure strict congruence (hereafter referred to as *Cmin*), and the semistrict consensus can be used to indicate semistrict congruence (hereafter referred to as *Cmax*) among trees. Whereas *Cmin* represents an index of topological identity, *Cmax* can be defined in a broader sense as a measure of topological compatibility among partially resolved trees. For that matter, *Cmin* and *Cmax* determine the lower and upper bounds of global congruence. Notice that both indices would give identical results for pairs of fully resolved trees, however.

In our analyses, the resolution of all trees was recorded. Total evidence trees were also compared with the different consensus trees by using strict (*Cmin*) and semistrict (*Cmax*) congruence. To assess the effect of bootstrapping, both indices were measured before (*Cmin* and *Cmax*) and after (*Cmin'* and *Cmax'*) collapsing the weakly supported clades (i.e., with bootstrap support <50%) in the total evidence trees. Our results were also compared with those previously obtained by the authors of the original studies. We wanted to know whether total evidence trees based on distances were topologically different from the previously published trees, when restricted to the same numbers of taxa. We also compared the bootstrap support values obtained in both cases to determine the number of well-supported clades in common. Finally, all individual trees obtained in the separate analyses were compared with the corresponding total evidence tree to detect any differences that could be reflected in the consensus.

RESULTS

All of the total evidence trees were fully resolved, as expected, given that a least-squares procedure was applied to distance matrices computed from the combined datasets. However, comparisons of these trees with those in the original studies revealed similar levels of resolution. In most cases, the least-squares phylogenies were congruent with the previously published trees based on parsimony (Table 1). The proportion of clades with high bootstrap support in our total evidence trees varied from 0.44 (4 of 9) to 1 (14 of 14); similar numbers were also obtained in the published trees.

In all but four cases, the topologies of the total evidence trees based on distances or characters were identical. For Lutzoni's (1997) data, the position of a single taxon was different in our tree, but the bootstrap value for the branch supporting that clade was rather low (61%). In the case of Flook et al.'s (1999) data, the relationship of two small clades differed from our tree. Olmstead and Sweere's (1994) data also differed from ours with respect to the relationships of four taxa within a large clade. Finally, the analysis of Mason-Gamer and Kellogg's (1996) data revealed several discrepancies in our tree relative to those already published; when we compared the relationships among the six taxa for which all sequences were available in the original study, only one clade appeared to be congruent with our total evidence tree.

All of the average consensus trees were fully resolved, just as the total evidence trees were. On the other hand, the resolution of the standard consensus trees was quite variable. The relative number of resolved clades ranged from 0 (0 of 14) to 0.66 (4 of 6) in strict consensus trees and from 0.13 (3 of 24) to 1 (7 of 7) in the majority rule trees. By definition, all clades in the majority rule consensus were obtained in the strict consensus trees; in three cases for which only two separate trees were combined, those consensus tree were identical (Kluge, 1989; Omland, 1994; Pennington, 1996). All of the clades in the average consensus trees were also obtained in the strict consensus trees. In six cases (Mason-Gamer and Kellogg, 1996; Lutzoni, 1997; Flook et al., 1999; Liu and Miyamoto, 1999; Quicke and Belshaw, 1999; Springer et al., 1999), however, unique clades were obtained

in the average consensus trees in comparison with the majority rule trees. In another case (Cannatella et al., 1998), the average and majority rule consensus trees were identical.

In all comparisons involving total evidence and average consensus trees, strict and semistrict congruence indices were identical because both trees were always fully resolved before bootstrap analysis. Strict congruence ($Cmin$) was better for the average consensus (0.78) than for majority rule (0.52) or strict consensus (0.20) trees, on average. On the other hand, semistrict congruence ($Cmax$) was worse for the average consensus (0.78) than for majority rule (0.97) or strict consensus (1.00) trees, on average. The standard consensus methods produced trees that were perfectly compatible ($Cmax = 1$) with the total evidence trees in 10 and 15 cases, respectively. Given the poor resolution of these consensus trees (see Table 1), such results were not surprising (an unresolved tree is always compatible with any other tree!). The comparisons performed after bootstrap analysis revealed quite different patterns, however. Strict congruence values decreased or remained the same for all consensus methods, following bootstrapping, and $Cmin'$ was again better for average consensus trees (0.64) than for majority rule (0.45) or strict consensus (0.19) trees, on average. Semistrict congruence ($Cmax'$) was also better for the average consensus (0.90) than for majority rule (0.80) or strict consensus (0.73) trees, on average, when the clades with low bootstrap support in total evidence trees were collapsed. Whereas $Cmax'$ always decreased or remained the same for standard consensus methods, it usually increased in the case of average consensus trees.

The results of the global congruence analysis comparing total evidence and average consensus trees can be classified into two categories: the perfectly congruent cases, for which combined and separate analysis provided identical trees ($Cmin = 1$), and the incongruent cases, for which the competing approaches provided topologically different solutions ($Cmin < 1$) before bootstrap analysis. Those incongruent results could be further divided into three subsets: studies that were unaffected by the bootstrap; studies for which bootstrap analysis partially improved the global congruence; and studies for which all topological incompatibilities between total evidence and average consensus trees were caused by weakly supported clades (Table 1).

Identical trees were obtained for combined and separate analyses in four cases, when average consensus trees were used (Omland, 1994; Baker and DeSalle, 1997; Cannatella et al., 1998; Gatesy et al., 1999a). In one of those cases (Cannatella et al., 1998), the majority rule consensus was also identical with the total evidence tree ($Cmin = 1$). As an example, Figure 1 illustrates the total evidence, average consensus, and strict consensus trees obtained with Omland's (1994) data. The total evidence tree is here topologically identical to the average consensus tree, whereas the strict consensus tree differs in terms of resolution. In this specific case, as well as with Baker and DeSalle's (1997) data, all clades were highly supported and global congruence was not affected by the bootstrap analysis. In the two remaining examples (Cannatella et al., 1998; Gatesy et al., 1999a), however, bootstrapping reduced strict congruence ($Cmin'$) between consensus and total evidence trees (see Table 1).

Apart from these four examples, the other 11 studies provided topologically different trees ($Cmin < 1$) when using combined or separate analyses before bootstrap analysis. In one situation (Baum et al., 1998), collapsing weakly supported clades in the total evidence tree did not affect the compatibility ($Cmax'$) of the average consensus but did reduce the topological compatibility of standard consensus trees (Table 1). In six other cases (Lutzoni, 1997; Messenger and McGuire, 1998; Flook et al., 1999; Liu and Miyamoto, 1999; Quicke and Belshaw, 1999; Springer et al., 1999), the $Cmax'$ of average consensus trees was improved after bootstrapping (Table 1). For example, Figure 2 presents the total evidence and consensus trees obtained with Messenger and McGuire's (1998) data. Before the bootstrapping, four clades in the total evidence tree were not compatible with the average consensus tree ($Cmax = 0.83$). Collapsing two clades with low bootstrap support (27% and 46%) reduced conflicts among the trees while increasing topological compatibility ($Cmax' = 0.91$), however. In this specific example, the majority rule consensus was as compatible ($Cmax' = 0.91$) as the average consensus tree, but more compatible than the strict consensus solution ($Cmax' = 0.87$). In three other cases (Pennington, 1996;
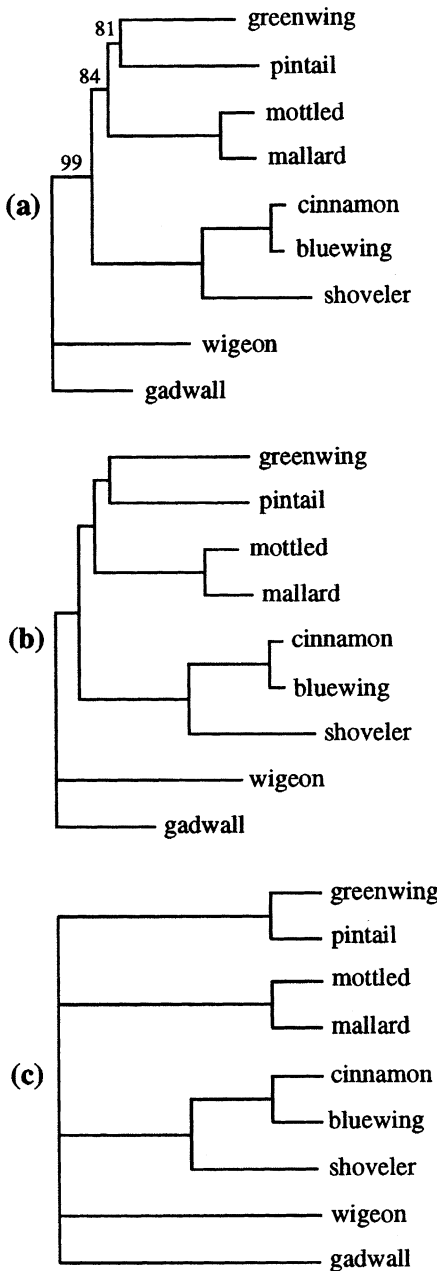
FIGURE 1. Comparison of (a) the total evidence tree, (b) the average consensus tree, and (c) the strict consensus tree obtained from Omland's (1994) data. Numbers above branches in the total evidence tree are bootstrap support values, when different from 100. The majority rule consensus tree is identical to the strict consensus tree. See text for more details.
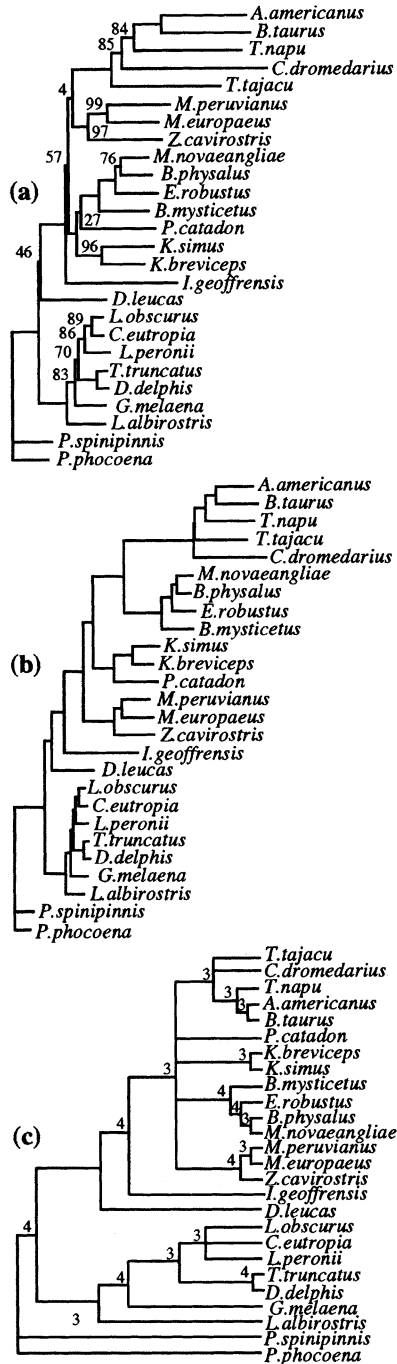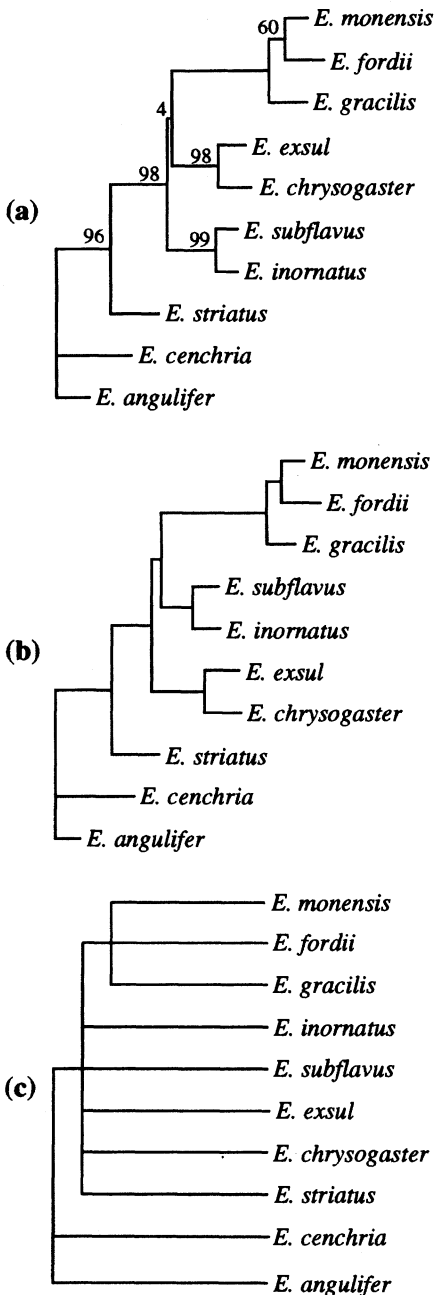

FIGURE 2. Comparison of (a) the total evidence tree, (b) the average consensus tree, and (c) the majority rule consensus tree obtained from Messenger and McGuire's (1997) data. Numbers above branches in the total evidence tree are bootstrap support values, when different from 100. Numbers above branches in the majority rule consensus tree are the numbers of individual trees containing that clade. The strict consensus tree can be obtained by collapsing all branches with numbers <4. See text for more details.

Lutzoni, 1997; Quicke and Belshaw, 1999), majority rule and strict consensus trees provided identical $Cmax'$ values after bootstrapping (Table 1).

(a)

(b)

(c)

FIGURE 3. Comparison of (a) the total evidence tree, (b) the average consensus tree, and (c) the strict consensus tree obtained from Kluge's (1989) data. Numbers above branches in the total evidence tree are bootstrap support values, when different from 100. The majority rule consensus tree is identical to the strict consensus tree. See text for more details.

The last four studies (Kluge, 1989; Olmstead and Sweere, 1994; Mason-Gamer and Kellogg, 1996; Pennington, 1996) certainly represent the most interesting examples (Table 1). In all of these cases, the observed discrepancies between total evidence and average consensus trees were not well supported, and collapsing those branches resulted in perfect topological compatibility. As an example, the results obtained from Kluge's (1989) data are presented in Figure 3. Before the bootstrap, one clade in the total evidence tree was not compatible with the average consensus tree ($Cmax = 0.86$), but collapsing that weakly supported clade (4%) resulted in perfectly compatible trees ($Cmax' = 1$). Conversely, standard consensus trees were compatible ($Cmax = 1$) with the total evidence tree before bootstrap analysis, but compatibility decreased after bootstrapping ($Cmax' = 0.86$). The same results were obtained for the other examples, except for Mason-Gamer and Kellogg's (1996) data. In this single case, strict congruence ($Cmin$) was better for the majority rule tree than for the average consensus tree before the bootstrap, but perfect compatibility ($Cmax' = 1$) was obtained with both consensus methods when the one clade with low bootstrap support was collapsed in the total evidence tree.

DISCUSSION

The main objective of this paper was to evaluate the generality and applicability of the global congruence approach (Lapointe et al., 1999), using a wide variety of datasets gathered from the literature. We postulated that a coherent distance-based approach would lead to congruent solutions, regardless of whether data or trees were combined. Our results supported that claim and showed that total evidence and consensus can provide very similar trees, if branch lengths are taken into account when combining the individual phylogenies. The average consensus procedure (Lapointe and Cucumel, 1997) and other consensus techniques for weighted trees (see Lapointe, 1998a) thus represent interesting alternatives to standard consensus methods, if the user is interested in capturing not only topological agreement among the trees combined but also pathlength differences.

We also predicted that average consensus trees would be more similar to total evidence

trees and more resolved than strict and majority rule consensus trees, as shown by Lapointe et al. (1999). Our analyses verified that prediction and demonstrated that topological methods are indeed more likely to produce less resolved trees. This is an important observation because the poor resolution of consensus trees has been repeatedly used to illustrate the superiority of a total evidence approach. In the light of the results obtained with average consensus, combined analysis does not appear to be superior to separate analyses. The trees produced by both approaches can be equally resolved, and in a large number of cases topologically identical, when branch lengths are taken into account.

On a related matter, we claimed that apparent conflicts between trees might be reduced if the weakly supported clades with low bootstrap support are collapsed in total evidence phylogenies. Our results confirm that assertion. As a matter of fact, the comparison of total evidence and consensus trees was greatly affected by bootstrap analysis. Strict ($Cmin'$) and semistrict congruence ($Cmax'$) decreased when the clades that were weakly supported in total evidence trees were also obtained in consensus trees. On the other hand, $Cmax'$ increased when the clades with low bootstrap support were incongruent with the consensus solutions (i.e., topological compatibility was improved when collapsing these clades). Consequently, the global congruence of total evidence and average consensus was always superior to that of standard consensus methods, after bootstrapping.

The different consensus methods used in this paper have distinct properties and can produce different solutions. On the one hand, topological techniques were initially developed to indicate corroboration by independent datasets (Nelson, 1979)—and this notion continues to be an important argument for using consensus (e.g., Miyamoto and Fitch, 1995). On the other hand, consensus with branch lengths (sensu Lapointe, 1998a) focuses on corroboration by ·distances—an approach implying that average consensus trees may greatly differ from strict (or majority rule) consensus trees. For instance, average consensus trees could contain clades that were not present in any of the individual trees (e.g., as was the case with Kluge's data; see Figure 3). The effect of long terminal branches may also affect consen-

sus solution when branch lengths are taken into account. However, several arguments favor consensus methods that incorporate such branch lengths. First, compared with standard consensus methods, average consensus uses more information and is usually more resolved. Second, the individual trees combined can be weighted according to the number of characters in the separate datasets. Third, various least-squares algorithms can be used to construct average consensus trees that satisfy different evolutionary models (e.g., by enforcing a molecular clock; see Kirsch et al., 1997). Finally, the robustness of average consensus trees could be assessed with other resampling procedures or with specific randomization techniques. For instance, Lapointe et al. (1999) used a taxonomic-jackknife procedure (Lapointe et al., 1994) to evaluate the effect of taxon sampling on the stability of consensus trees. Likewise, Cucumel and Lapointe (2000) developed a statistical test to determine the probability that a consensus is representative of a set of individual trees. Such methods could be used to assess the robustness of total evidence and consensus trees similarly (see Lapointe and Kirsch, in press).

The differences among consensus methods that ignore or incorporate branch lengths are also apparent in our results. Whereas topological identity ($Cmin = 1$) was obtained only for average consensus trees (except in one case), topological compatibility ($Cmax = 1$) was common for standard consensus techniques, before bootstrap analysis. Note that these congruence indices measure different things, however. $Cmin$ measures topological identity as a strict consensus would represent it; its maximum value is obtained for fully resolved trees containing exactly the same clades. $Cmax$, on the other hand, is related to what a semistrict consensus (Bremer, 1990) aims at representing; its maximum value is obtained when the consensus is fully resolved and no clades are incompatible in the trees compared. Both indices would provide identical results in the case of fully resolved trees, but $Cmax$ is more liberal when polytomies are found in any of the input trees, as it is the case when branches are collapsed after bootstrapping, or when a consensus is not well resolved (e.g., strict consensus trees).

Despite the wide variety of studies considered in the present work, we observed no relationships between global congruence

indices (*Cmin* and *Cmax*) and the characteristics of the datasets. For instance, the perfectly congruent cases (Omland, 1994; Baker and DeSalle, 1997; Cannatella et al., 1998; Gatesy et al., 1999a) were based on different taxa (i.e., birds, insects, amphibians, and mammals, respectively), with the number of species ranging from 9 to 17 and the number of datasets ranging from 2 to 8, including sequence, morphological, and allozyme data, among others. There was also no relationship between bootstrap support and congruence, neither in those four cases nor in other examples. However, five of the studies for which bootstrapping did not result in perfect compatibility (Lutzoni, 1997; Messenger and McGuire, 1998; Flook et al., 1999; Liu and Miyamoto, 1999; Quicke and Belshaw, 1999) were among the largest in terms of number of taxa. Global congruence may well be decreased with increasing numbers of taxa, but further analyses with a greater number of studies would be required to support or reject that claim.

All of our combined analyses were based on distance matrices, computed from character data or molecular sequences. Similarly, the consensus analyses were based on the combination of pathlength distance matrices, corresponding to the different trees obtained from individual datasets. For that matter, our approach could be blind to clade support possibly hidden in the analyses of separate datasets, whereas a character congruence approach would be able to identify such "hidden support" (see Gatesy et al., 1998b). Interestingly, the total evidence trees computed with a least-squares algorithm were in most cases congruent with the most-parsimonious total evidence trees published in the original studies. For the most part, those trees were in turn congruent with average consensus trees. Consequently, our results do not seem to have been affected by the type of phylogenetic estimation method, nor by the use of distances, and our conclusion still holds when compared with the original phylogenies.

Future work could generalize the global congruence approach to supertrees defined on overlapping sets of taxa (Lapointe and Cucumel, 1997), instead of reducing the datasets to include only common taxa. In such applications (see Kirsch et al., 1997; Lapointe and Kirsch, 2001), the average consensus procedure represents an alternative to the more commonly used supertree methods

based on topological relationships (for a review, see Sanderson et al., 1998), while considering branch lengths. As stated above, the effect of various weighting schemes could also be evaluated to determine optimal ways to combine phylogenies derived from datasets with different numbers of characters. Taxonomic congruence automatically attributes equal weights to each dataset regardless of the numbers of characters each contains, a procedure considered by some as arbitrary (Hillis, 1987; Donoghue and Sanderson, 1992). However, assigning unit weights to every character—molecular as well as morphological—in a total evidence framework also causes problems (see Doyle, 1992). This dual effect can be addressed with a weighted-average consensus method, accounting for the number of characters in the individual datasets when computing the consensus tree (Lapointe et al., 1999). Finally, we would like to assess whether a global congruence approach increases phylogenetic accuracy over that obtained when using either one of the competing methods independently. The agreement between trees obtained with combined and separate analysis could then be visualized in a global consensus tree (e.g., Lapointe et al., 1999; Lapointe and Kirsch, 2001) bearing the clades corroborated by the different approaches; one could then argue that those clades are more likely to be real (see Kim, 1993).

REFERENCES

BAKER, R. H., AND R. DESALLE. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Syst. Biol. 46:654–673.

BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. Syst. Zool. 40:486–493.

BAUM, D. A., R. L. SMALL, AND J. F. WENDEL. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. Syst. Biol. 47:181–207.

BREMER, K. 1990. Combinable component consensus. Cladistics 6:369–372.

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.

CANNATELLA, D. C., D. M. HILLIS, P. T. CHIPPINDALE, L. WEIGT, A. S. RAND, AND M. J. RYAN. 1998. Phylogeny of frogs of the *Physalaemus pustulosus* species group, with an examination of data incongruence. Syst. Biol. 47:311–335.

CAVALLI-SFORZA, L. L., AND A. W. F. EDWARDS. 1967. Phylogenetic analysis: Models and estimation procedures. Evolution 32:550–570.

CHIPPINDALE, P. T., AND J. J. WIENS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43:278–287.

CUCUMEL, G., AND F.-J. LAPOINTE. 2000. A general approach to test the pertinence of a consensus classification. Pages 125–130 in Data analysis, classification, and related methods (H. A. L. Kiers, J.-P. Rasson, P. J. F. Groenen, and M. Schader, eds.). Springer-Verlag, Berlin.

DE QUEIROZ, A. 1993. For consensus (sometimes). Syst. Biol. 42:368–372.

DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.

DONOGHUE, M. J., AND M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pages 340–368 in Molecular systematics in plants (P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds.). Chapman & Hall, New York.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Syst. Bot. 17:144–163.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Testing significance of incongruence. Cladistics 10:315–319.

FLOOK, P. K., S. KLEE, AND C. H. F. ROWELL. 1999. Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. Syst. Biol. 48:233–253.

GATESY, J., M. MILINKOVITCH, V. WADDELL, AND M. STANHOPE. 1999a. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. Syst. Biol. 48:6–20.

GATESY, J., P. O'GRADY, AND R. H. BAKER, 1999b. Corroboration among data sets in simulataneous analysis: Hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15:271–314.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18:23–42.

HUELSENBECK, J. P., AND J. J. BULL. 1996. A likelihood ratio test for detection of conflicting phylogenetic signal. Syst. Biol. 45:92–98.

HUELSENBECK, J. P., J. J. BULL, AND C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:152–158.

HUELSENBECK, J. P., D. L. SWOFFORD, C. W. CUNNINGHAM, J. J. BULL, AND P. J. WADDELL. 1994. Is character weighting a panacea for the problem of data heterogeneity in phylogenetic analysis? Syst. Biol. 43:288–291.

KIM, J. 1993. Improving the accuracy of phylogenetic estimation by combining different methods. Syst. Biol. 42:331–340.

KIRSCH, J. A. W., F.-J. LAPOINTE, AND M. S. SPRINGER. 1997. DNA-hybridisation studies of marsupials and their implications for metatherian classification. Aust. J. Zool. 45:211–280.

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

KLUGE, A. G. 1998. Total evidence or taxonomic congruence: Cladistics or consensus classification. Cladistics 14:151–158.

KLUGE, A. G., AND A. J. WOLF. 1993. Cladistics: What's in a word? Cladistics 9:183–199.

LAPOINTE, F.-J. 1998a. 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. 1998b. How to validate phylogenetic trees? A stepwise procedure. Pages 71–88 in Data science, classification, and related methods: Studies in classification, data analysis, and knowledge optimization (C. Hayashi, H. H. Bock, K. Yajima, Y. Tanaka, N. Oshumi, and Y. Baba, eds.). Springer-Verlag, Tokyo.

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., AND J. A. W. KIRSCH. Construction and verification of a large phylogeny of marsupials. Aust. Mammal. (in press).

LAPOINTE, F.-J., J. A. W. KIRSCH, AND R. BLEIWEISS. 1994. Jackknifing of weighted trees: Validation of phylogenies reconstructed from distance matrices. Mol. Phylogenet. Evol. 3:256–267.

LAPOINTE, F.-J., J. A. W. KIRSCH, AND J. M. HUTCHEON. 1999. Total evidence, consensus, and bat phylogeny: A distance-based approach. Mol. Phylogenet. Evol. 11:55–66.

LARSON, A. 1994. The comparison of morphological and molecular data in phylogenetic systematics. Pages 371–390 in Molecular ecology and evolution: Approaches and applications (B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, eds.). Birkhauser Verlag, Basel.

LIU, F. R., AND M. M. MIYAMOTO. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. Syst. Biol. 48:54–64.

LUTZONI, F. M. 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.

MARGUSH, T., AND F. R. MCMORRIS. 1981. Consensus *n*-trees. Bull. Math. Biol. 43:239–244.

MASON-GAMER, R. J., AND E. A. KELLOGG. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. 45:524–545.

MESSENGER, S. L., AND J. A. MCGUIRE. 1998. Morphology, molecules and the phylogenetics of Cetaceans. Syst. Biol. 47:90–124.

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. 27:143–158.

MICKEVICH, M. F., AND J. S. FARRIS. 1981. The implications of congruence in *Menidia*. Syst. Zool. 30:351–370.

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

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

NELSON, G. 1979. Cladistics analysis and synthesis: Principles and definitions, with a historical note on Adanson's Familles des Plantes (1763–1764). Syst. Zool. 28:1–21.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43:467–481.

OMLAND, K. E. 1994. Character congruence between a molecular and a morphological phylogeny for dabbling ducks (Anas). Syst. Biol. 43:369–386.

PAGE, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12:357–358.

PENNINGTON, R. T. 1996. Molecular and morphological data provide phylogenetic resolution at different hierarchical levels in Andira. Syst. Biol. 45:496–515.

ROBINSON, D. F., AND L. R. FOULDS. 1981. Comparison of phylogenetic trees. Math. Biosci. 53:131–147.

RODRIGO, A. G., M. KELLY-BORGES, P. R. BERGQUIST, AND P. L. BERGQUIST. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N. Z. J. Bot. 31:257–268.

ROHLF, F. J. 1982. Consensus indices for comparing classifications. Math. Biosci. 59:131–144.

SANDERSON, M. J., A. PURVIS, AND C. HENZE. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol. Evol. 13:105–109.

SHAO, K., AND R. R. SOKAL. 1986. Significance tests of consensus indices. Syst. Zool. 35:582–590.

SOKAL, R. R., AND F. J. ROHLF. 1981. Taxonomic congruence in the Leptopodomorpha re-examined. Syst. Zool. 30:309–325.

STEEL, M. A., AND D. PENNY. 1993. Distributions of tree comparison metrics—Some new results. Syst. Biol. 42:126–141.

SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 in Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. J. Cracraft, eds.). Oxford Univ. Press, Oxford.

SWOFFORD, D. L. 1999. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4, Sinauer Associates, Sunderland, Massachusetts.