Syst. Biol. 52(4):554–564, 2003

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150390223730

Stability in Phylogenetic Formulations and Its Relationship to Nodal Support

GONZALO GIRIBET

Museum of Comparative Zoology, Department of Organismic & Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA; E-mail: ggiribet@oeb.harvard.edu

The phylogenetic analysis of nucleic acid sequences, as with other data, is unavoidably based on explicit and implicit assumptions. At the fore are character transformation models—usually transversion–transition ratios–and the relative cost of alignment-derived sequence gaps. These values are the fulcra of sequence analysis. Simple homogeneous weighting does not avoid the issue of arbitrary, yet crucial, assumptions. (Wheeler, 1995:321)

Sensitivity Analysis (SA) is the study of how variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation, and how the given model depends upon the information fed into it. (Saltelli, 2000:3)

Robustness is resistance to variation in some input/assumption. In the context of phylogenetic analysis, the term robust has been traditionally applied to a method that is consistent, even with sizable deviations from a given model (Penny et al., 1992). In a more general sense, robustness could be defined as the stability of a phylogenetic hypothesis to variation in phylogenetic inference procedures, including weighting scheme or model of evolution and the method used to construct the tree. In phylogenetics, robustness is often assessed via measures of nodal support but rarely as the variation resulting from alternative models or tree constructing methodologies. Even in those cases where robustness to model selection is truly evaluated (e.g., Buckley and Cunningham, 2002), still the individual trees are often judged by appealing to nodal support measures. Recently, the notion of (nodal) stability has become important when postulating taxonomic formulations, although this application of the robustness criterion has been mostly restricted to stability to parameter variation through alignment-to-tree construction or via single-step phylogenetic analysis (direct optimization) in the sensitivity analysis framework proposed by Wheeler (1995). In this sense, the stability notion is considered here, as a test of robustness within a method of phylogenetic analysis. My intention is to explicitly discuss the relationship between nodal support (as commonly applied in phylogenetic inference) and nodal stability.

Nodal support, as measured by the methods of bootstrapping (Felsenstein, 1985), Bremer support (Bremer, 1988, 1994), jackknifing (Farris et al., 1996), relative support (Goloboff and Farris, 2001), Bayesian posterior probabilities (Huelsenbeck et al., 2001), or spectral plots (Lento et al., 1995), is generally required for publication. A review of some or all of these methods can be found elsewhere (Swofford et al., 1996; Siddall, 2002), and it is not my intention to discuss in detail these methods of evaluating nodal support. Contrary to the notion of how much support for a given node is derived from the analysis in question, knowing the degree to which those nodes are affected by variation in the analytical conditions under which they are postulated may be interesting; such a measure is here formally defined as nodal stability. The idea of a nodal stability measure is much newer in phylogenetic analysis than that of nodal support, and it is directly related to issues of sensitivity analysis (Wheeler, 1995). Although its application has been mostly in parsimony analyses (especially with the use of direct optimization), its usage could be much broader, and it may become a standard way of assessing phylogenetic relationships. Here, I review the application of nodal stability in phylogenetic studies by using three examples. (Other types of stability, in particular leaf stability, have been used [Thorley and Wilkinson, 1999], although that measure does not directly compare to the one discussed here.) The results presented here support the idea that sensitivity analysis is a useful technique for assessing nodal stability, that the results of a stability analysis are often but not always correlated to values of nodal support, and that the comparison of multiple models for a given method of phylogenetic inference may well constitute the best way to detect nodes supported only under particular analytical conditions.

HISTORICAL PERSPECTIVE

The existence of parameter dependence in multiple sequence alignments was recognized early (Fitch and Smith, 1983; Williams and Fitch, 1989); alternative alignment parameters result in different alignments, and these may result in alternative phylogenetic trees. Dependence in analytical parameters is independent of the tree-construction methodology of choice and thus is a central conundrum of phylogenetic analysis. Dependence theoretically affects pairwise-distance–based alignments (Higgins and Sharp, 1988; Jeanmougin et al., 1998), parsimony-based alignment methods (Wheeler and Gladstein, 1994; Wheeler, 1995), and maximum likelihood–based alignments (Thorne et al., 1991). The fact that alternative alignments may result in alternative trees was elegantly presented by Morrison and Ellis (1997). That study concluded that changes in phylogenetic hypotheses are greater when alternative alignments are compared within methods of tree construction than when the same alignment is analyzed across methods.

Other authors have advocated the use of secondary structure to improve congruence with morphology (Titus and Frost, 1996), to improve accuracy and robustness (Hickson et al., 2000), or simply to improve homology statements across taxa (Giribet, 2001, 2002). I do not discuss sequence alignments further here except to point out that automated multiple alignment approaches require arbitrary decisions in parameter selection (i.e., gap cost and length, transversion/transition cost ratios, etc.), and the sensitivity analysis approach aims to explore these arbitrarily chosen parameter sets (e.g., Wheeler, 1995; Giribet and Wheeler, 1999). However, it is important to note that, some authors (Kjer, 1995; Hickson et al., 2000) have argued that for aligning molecules such as RNA, a single gap cost could not be optimal because the probability of an indel event changes dramatically from one RNA segment to another.

This notion of parameter dependence introduced by Fitch and Smith (1983) leads the investigator toward a fundamental question: which parameter values need to be explored? The necessity of making explicit but somehow arbitrary decisions in parameter specification led Wheeler (1995) to propose a methodology for exploring the data under a range of parameter set values. This parameter exploration was later termed sensitivity analysis because it was equivalent to sensitivity analysis in a statistical sense (Saltelli, 2000) and is the basis of the notion of nodal stability. After Fitch and Smith (1983), logically, parameter exploration was incorporated implicitly into a few phylogenetic analyses, but it was not until 1995 that parameter sensitivity was explicitly incorporated into choosing alignment parameters for the purpose of building phylogenetic trees.

The same article that introduced the parameter sensitivity analysis discussion (Wheeler, 1995) incorporated a congruence assay as the most reasonable metaoptimality criterion for phylogenetic analysis. This congruence assay was aimed at choosing the optimal set of parameters (among the explored ones) for a given data set, i.e., the parameters that maximized precision (the agreement among data). Because the two notions, parameter sensitivity and the congruence assay to choose the best corroborated hypothesis, were introduced simultaneously, many systematists seem to see both exercises as necessarily connected, i.e., one investigator generates *n* phylogenetic hypotheses under *n* combinations of parameters and appeals to congruence (or other criteria) to present the preferred hypothesis. However, these two steps are not necessarily connected and one could explore an infinite number of parameter combinations for parsimony or for likelihood models and display the respective trees without making a decision as to which one yields the preferred hypothesis (Giribet, 2002; Giribet and Wheeler, 1999). Furthermore, a sensitivity analysis is required to make reference to nodal stability, but it is not necessary to choose among the different hypotheses generated under the different parameter/models (e.g., one could display all topologies obtained under different analytical conditions). Sensitivity analysis is therefore independent of hypothesis selection.

As mentioned above, alignment parameter sensitivity was made explicit in phylogenetic analysis after publication of Wheeler's (1995) article, although because of computational difficulties in generating multiple sequence alignments for different alignment parameters (Wheeler, 1994; Slowinski, 1998), some authors explored parameter variation solely in the tree-building step (Whiting et al., 1997) and not for constructing alignments. For authors using direct optimization (Wheeler, 1996), sensitivity analysis became standard practice. Under direct optimization, parameters need to be specified only once, and because of computation efficiency the phylogenetic exercise could be repeated multiple times for different parameters in the computer program POY (Wheeler et al., 2002). In contrast, two-step phylogenetic analyses, requiring a multiple sequence alignment step plus a tree search, are more costly because they require two levels of heuristics, one at the alignment level (a sum of two NP-complete problems) and another one for the tree search. The recent incorporation of likelihood functions into the software allows maximum-likelihood analyses using the same criterion for establishing primary homologies, a necessity mostly ignored because of the computational burden (Thorne et al., 1991). Since the publication of the direct optimization method, the use of sensitivity analysis has increased, with publications appearing in all the major systematics journals (Wheeler and Hayashi, 1998; Sorenson et al., 1999; Arnedo et al., 2001; Cognato and Vogler, 2001; Frost et al., 2001a; Giribet et al., 2001; Janies, 2001; Sanchis et al., 2001; Shull et al., 2001; Belshaw and Quicke, 2002; Edgecombe et al., 2002; Hormiga et al., 2003; Whiting et al., 2003). Shull et al. (2001) compared direct optimization to alignment sensitivity analysis followed by phylogenetic analysis.

The appeal of sensitivity analysis is the new notion of stability, originally formulated as robustness in a statistical inference paradigm (Wheeler, 1995:323):

> As with statistical inference, two types of decisions (estimates of parameter values) can be made: best and robust. A best decision is made by choosing the set (or sets) of parameter values at which the optimality criterion is maximized. . . . A robust decision selects a range of parameter values rather than settling on a single set. This range defines a subset of the analysis space in which some statement is supported. For example, an area might be specified in which some group was monophyletic.

But later, when referring to the congruence plots, Wheeler (1995:328) interpreted stability in the more restricted sense here employed: "The size and number of these areas give a measure of the generality and stability of the hypothesis of monophyly."

Stability Analyses

Nodal stability measures are an attempt to show how alignment (or other specifiable) parameters affect phylogenetic hypotheses. In the simple case of a parsimony analysis, stability can measure how alignment parameters (e.g., indel costs) affect the outcome topology and can give an indication of whether a given hypothesis is supported by all-many-few-no alignments. This approach

has been incorporated for several alignments generated under different costs and analyzed via parsimony applying the same step matrices (Wheeler, 1995; Giribet and Carranza, 1999; Giribet and Wheeler, 1999), for a single alignment analyzed under different weighting schemes (Whiting et al., 1997), or for multiple parameters under direct optimization (Wheeler, 1998; Giribet and Wheeler, 1999; Sorenson et al., 1999; Giribet et al., 2000; Arnedo et al., 2001; Cognato and Vogler, 2001; Frost et al., 2001a, 2001b; Janies, 2001; Shull et al., 2001; Edgecombe et al., 2002; Whiting et al., 2003). However, the quest for stability of phylogenetic hypotheses goes beyond the use of parsimony or direct optimization. Authors often present pluralistic analyses with trees generated under different optimality criteria or tree-building methods. Although I have criticized this type of "stability" analysis elsewhere for not achieving its purpose (Giribet et al., 2002a), I still believe that it is done with the intention of testing for stability under different analytical conditions. (The criticism refers to the necessity of exploring the chosen method of analysis thoroughly instead of seeking for congruence among methods under similar conditions.)

The search for stability is also rooted in morphological analysis of phylogenetic data. For example, in a recent taxonomic study of the scorpion superfamily Scorpionoidea, Prendini (2000) evaluated the effect of different concavity values under implied weights and reported topological changes under the different k values. Exploration of a range of successive weighting options has also been presented in an analysis of sphenodontid phylogeny (Wilkinson and Benton, 1996). Clade stability is so important that it drives the current discussion about classification systems.

As mentioned above, sensitivity analyses do not conflict with the notion of choosing the best (most corroborated) phylogenetic hypothesis (i.e., the tree obtained under the most congruent parameter set or under the most likely model). Both types of trees are sometimes presented simultaneously, the preferred tree side by side with the strict component consensus of all the analyzed parameter sets (Wheeler, 1995; Wheeler and Hayashi, 1998; Edgecombe et al., 1999, 2002; Giribet and Boyer, 2002; Giribet and Wheeler, 2002).

Sensitivity analyses are also represented graphically using the "Navajo rugs," a graphic plot where the parameter space is represented as a grid with parameters represented in two (or more) axes. A typical representation shows the indel–change ratio in one axis and transversion–transition cost ratio in the second axis, and color coding is employed for designating monophyly or nonmonophyly of a given group (Wheeler, 1995; Giribet and Wheeler, 1999, 2002; Janies, 2001; Giribet et al., 2002b; Edgecombe and Giribet, 2003).

A Comparison of Nodal Support Measures

One of the most interesting questions concerning most common methods for assessing nodal support in phylogenetic hypotheses is what is their relationship to each other? For example, two of the most common resampling methods for assessing nodal support are bootstrapping and jackknifing, which differ in the way characters are resampled. In bootstrapping (Felsenstein, 1985), characters are resampled with replacement from the original matrix to form a pseudoreplicate matrix with the same number of characters as the original. Jackknifing (Farris et al., 1996) instead employs independent removal, in which each character has the same chance, e^{-1} (about 0.37), of being omitted from a given pseudoreplicate. Jackknifing thus simplifies the relationship between frequency and support; for data with no missing entries, the expected jackknife frequency of a group G set off by k uncontradicted characters is $1 - e^{-k}$ (Farris et al., 1996). Bootstrapping has the same expectation when the total number n of characters is very large (Farris, 1997). The relationship between the two resampling techniques is simple, and the interpretation of the expected results is straightforward: Jackknifing generally shows lower frequencies for a given group. The relationship between character-based measures, such as Bremer support (Bremer, 1988), and the resampling methods is not as well understood.

Bremer support generates absolute values of the degree to which a tree is suboptimal compared with another tree. A defect of that method is that it does not always take into account the relative amounts of evidence contradictory and favorable to the group. For example, according to the Bremer metric, a group supported by 2 uncontradicted characters is less well supported than a group supported by 100 and contradicted by 97. The first group, however, is relatively well supported, whereas in the second case there is about as much evidence in favor of the group as against it. This problem, as pointed out by Goloboff and Farris (2001; see also Lee, 2000), is diminished if the support for the group is calculated as the ratio between the favorable and the contradictory evidence. Potential advantages of the relative supports measures over normal Bremer support are that they vary between 0 and 1 and they provide an approximate measure of the amount of favorable/contradictory evidence (e.g., if the corrected Bremer support [RFD] is 0.25, the amount of contradictory evidence is 75% of the amount of favorable evidence, so it is equivalent to the conflict of four favorable characters versus three contradictory characters) (Goloboff and Farris, 2001). In this sense, the relative support may be directly comparable with the commonly employed resampling techniques. Problems with the interpretation of Bremer support values and their relation to branch lengths have been recently pointed out (DeBry, 2001).

Is There a Connection Between Nodal Support and Node Stability? Three Examples

What, if any, is the relationship between measures of assessing nodal support and the notion of stability discussed here? Since no published article has explicitly discussed this issue, by exploring the relationship between nodal support and nodal stability, I stress the necessity of using sensitivity analysis as a mode to explore nodal

stability in phylogenetics in addition to the common measures of assessing robustness. I illustrate this approach by using some real data sets.

Ideally, one would expect (or wish) that a well-supported node is not model-dependant. Otherwise, one could ask how good is a well-supported relationship between taxon A and taxon B if it is only obtained under one model out of a whole set of models explored? But what happens if that single model is our preferred one under the chosen (meta-)optimality criterion (e.g., a likelihood ratio test)? Such cases can be found in real data sets, and although they do not necessarily invalidate an entire analysis/hypothesis/tree, they may indicate that something suspicious may be going on and can summarize the conflict between best and robust. Three examples are discussed below:

Example 1: Arthropod Phylogeny

For example, Figure 1 is a tree obtained after analyzing eight molecular loci (18S ribosomal rRNA [rRNA], 28S rRNA, elongation factor 1α, RNA polymerase II, histone H3, U2 snRNA, 16S rRNA, and cytochrome *c* oxidase subunit I) and a large morphological matrix for a set of 48 arthropods plus selected outgroups (Giribet et al., 2001). While the general topology of the tree and the deepest nodes are reasonable from a traditional and modern perspective (i.e., monophyly of Arthropoda, Pycnogonida, Chelicerata, Mandibulata, Myriapoda, and Tetraconata), there is a node that especially attracts the attention of arthropod systematists. This node includes taxa as divergent as a japygid (Hexapoda, Diplura), a fruit fly (Hexapoda, Diptera), and a barnacle (Crustacea, Cirripedia). This group is furthermore supported by a Bremer support (BS) value of 55, the 10th highest value for that tree; the other nine most well-supported groups are Remipedia (BS = 506), Xiphosura (157), Symphyla (126), Archaeognatha (112), Onychophora (112), Cephalocarida + Remipedia (87), Malacostraca (80), Pycnogonida (73), and Opiliones (56) (see Table 1). All nine groups are well supported morphologically, and no sensible phylogenetic hypothesis would question them. These nine clades share with the japygid–fruit fly–barnacle clade a high value of Bremer support, but they have a fundamental difference. All nine clades are quite stable to parameter set variation (indel and change ratios) and are recovered in 76–100% of the 12 analytical conditions examined (a range of gap–change and transversion–transition cost ratios) (see Table 1). However, the spurious clade, dipluran–fruit fly–barnacle, is only obtained in 1 of the 12 parameter sets examined. Given our knowledge of the relationships of diplurans, barnacles, and fruit flies based on a wealth of biological information, the stability criterion would be a better indicator of the lack of reality of the odd japygid–fruit fly–barnacle clade than the Bremer metric.

When the tree from Figure 1 was published, only Bremer support values and stability measures were provided. When adding a standard resampling technique to evaluate nodal support, such is the case of jackknife proportions, we found a somewhat different result. The japygid–fruit fly–barnacle clade received a jackknife value of <50%, whereas all groups receiving high values are well recognized morphologically, such as Onychophora (100), Pycnogonida (100), Symphyla (100), Xiphosura (100), Remipedia (100), Archaeognatha (99), Branchiopoda (94), Chelicerata (92), and Opiliones (92) (see Table 1 for other nodes). Only one group not recognized morphologically received a jackknife proportion value >85%, the clade uniting scorpions, spiders, and vinegaroons, a clade that is also highly stable to parameter variation. However, this clade received a moderate-to-low Bremer support value, in the range of those taxa that received low jackknife support and/or were unstable to parameter variation. In this example, only one clade that was 100% stable to parameter variation received a jackknife value <85%, the clade uniting Remipedia and Cephalocarida (jackknife proportion = 80%).

According to this data set and analysis, there is a closer relationship between the measure of stability here employed and the resampling techniques (jackknife used here) than between either of these measures and Bremer support, which is artificially affected by not taking into account the contradictory evidence (Goloboff and Farris, 2001) or the number of changes accumulated on branches (DeBry, 2001). It is also important to note that the opposite situation may also occur in real data sets; low nodal support values may be related to highly stable clades. For example, the pair Remipedia1 + Remipedia3 received one of the lowest Bremer support values (9) and a jackknife value <50%, although it was consistently obtained under all parameter sets explored, i.e., it is a very stable node. This behavior is logically expected when a clade has just a few unambiguous characters supporting it and no characters that contradict it. Unlike the previous case of spurious

TABLE 1. Values of nodal support and stability analysis measured by Bremer support (BS), parsimony jackknifing (JF), and stability analysis shown as the 50% majority rule consensus of the 12 analytical parameters examined (ST). Nodes and clades are illustrated in Figure 1. Empty cells indicate JF or ST values <50%.

		Index		
Node	Clade	BS	JF	ST
1	Remipedia	506	100	100
2	Xiphosura	157	100	100
3	Symphyla	126	100	100
4	Onychophora	112	100	100
5	Archaeognatha	112	99	100
6	Cephalocarida + Remipedia	87	80	100
7	Malacostraca	80	87	75
8	Pycnogonida	73	100	100
9	Opiliones	56	92	100
10	*Drosophila* + Japygidae + Balanidae	55		
11	Branchiopoda	46	94	83
12	Juliformia	42	79	100
13	Chelicerata	38	92	100
14	Tetraconata	37		66
15	Scorpiones + Araneae + Uropygi	35	86	100
16	Arthropoda	34		
17	Chilopoda	30	80	

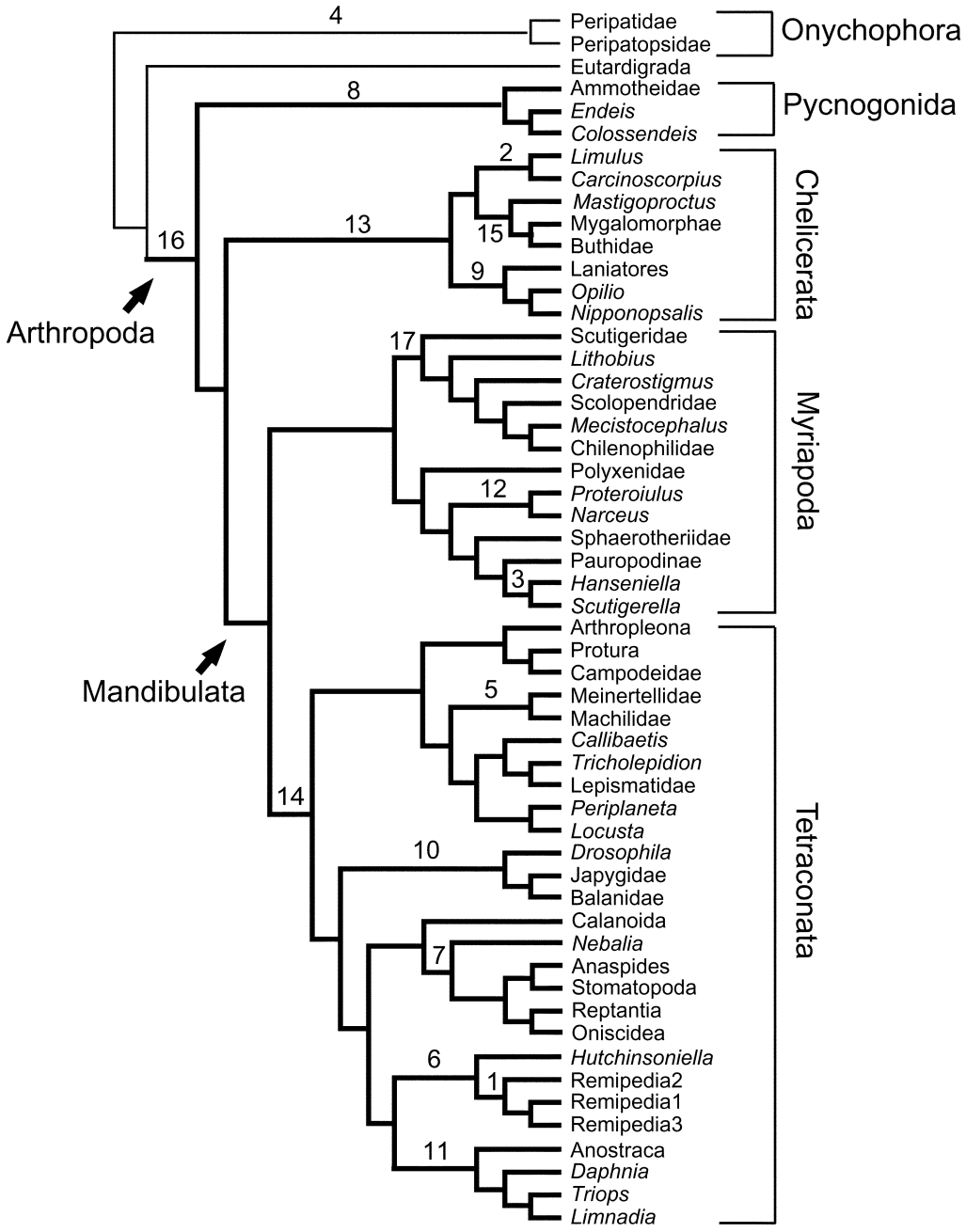

FIGURE 1. Phylogenetic tree of selected arthropods based on the combined analysis of a morphological data set and eight molecular loci (analysis based on direct optimization when all transformations, including gaps, are equally weighted; the parameter set that minimized incongruence among partitions). Numbers correspond to nodes listed in Table 1. Terminal taxa are velvet worms (Peripatidae and Peripatopsidae), tardigrades (Eutardigrada), sea spiders (Ammotheidae, *Endeis* and *Colossendeis*), horseshoe crabs (*Limulus* and *Carcinoscorpius*), vinegaroon (*Mastigoproctus*), spider (Mygalomorphae), scorpion (Buthidae), daddy-long-legs (Laniatores, *Opilio*, *Nipponopsalis*), centipedes (Scutigeridae, *Lithobius*, *Craterostigmus*, Scolopendridae, *Mecistocephalus*, Chilenophilidae), millipedes (Polyxenidae, *Proteroiulus*, *Narceus*, Sphaerotheriidae), pauropod (Pauropodinae), symphylans (*Hanseniella*, *Scutigerella*), springtail (Arthropleona), proturan (Protura), diplurans (Campodeidae, Japygidae), bristletails (Meinertellidae, Machilidae), silverfish (Lepismatidae, *Tricholepidion*), mayfly (*Callibaetis*), cockroach (*Periplaneta*), grasshopper (*Locusta*), fruit fly (*Drosophila*), crustaceans (Remipedia, *Hutchinsoniella*, Calanoida, Anostraca, *Triops*, *Limnadia*, *Daphnia*, *Nebalia*, Balanidae, Stomatopoda, *Anaspides*, Oniscidea, Reptantia). For more details, see Giribet et al. (2001).

nodes (i.e., long branch attraction), this case does not need to be an artifact but rather reflects the reality of many data sets; having few characters supporting a node may not necessarily indicate that the node is not well corroborated, especially in the case of lack of contradictory characters (see also Bandelt et al., 1995; Knox and Palmer, 1995). This has also been noticed when comparing quartet puzzling (QP) support values with those of bootstrapping for recently evolved taxa, where unlike bootstrap values the QP values are usually high when sequence

data show few informative sites but relatively few incompatibilities (Lockhart et al., 2001). A similar conclusion is found when comparing bootstrap support and Bayesian posterior probability values (Jordan et al., 2003).

While some systematists may pursue the goal of support, others may prefer to stick to the notion of stability, and hence could, for example, name a node that may seem poorly supported because there are no analytical conditions that may overturn it. (But it is not the point of this article to make a choice between stability or support.)

Bremer support has been criticized as a measure for assessing nodal support for taking into account only evidence in favor and not evidence against a given node (Lee, 2000; Goloboff and Farris, 2001). While this is certainly so, a direct relationship between the corrected Bremer support of Goloboff and Farris (RFD) and the notion of stability cannot be explored for the present data set because of the lack of implementation. I believe the RFD to be more closely related to stability measures than Bremer support, but this is so far speculation. However, the same disconnection between stability and support may occur for other measures of nodal support based on resampling techniques such as bootstrapping or parsimony jackknifing (for a clear comparison of these two methods, see Farris, 1997).

Example 2: Relationships Among Henicopid Centipedes

A second example of instability is found in an analysis of five molecular markers (18S rRNA, 28S rRNA, cytochrome *c* oxidase subunit I, 16S rRNA, and 12S rRNA) and 58 morphological characters for the relationships of the centipede family Henicopidae (Edgecombe and Giribet, 2003). The analysis contains 47 stone centipedes, and one of the studied genera, the Gondwanan *Paralamyctes*, is represented by 22 terminals (16 morphologically recognized species). *Paralamyctes* includes the previously recognized genus of small henicopids *Haasiella*, which clearly shows overlap with *Paralamyctes* in geographical distribution (Edgecombe et al., 2002). But the morphological delimitation of *Paralamyctes* Pocock, 1901, sensu Edgecombe (2001), is ambiguous for morphological data alone, and when the molecular data are analyzed under the parameter set that minimized overall incongruence (in this case, when all transformations [indels, transversions, and transitions] are weighted equally), the monophyly of *Paralamyctes* is violated by *Cermatobius* (= *Esastigmatobius*) falling inside *Paralamyctes* (Fig. 2). Current taxonomy places the Japanese genus *Cermatobius* with the American genus *Zygethobius* in the Laurasian tribe Zygethobiini; molecular analysis under equal weights places *Cermatobius* within the Gondwanan clade *Paralamyctes*, with a jackknife support frequency of 81% (Fig. 2). While this analysis may suggest that support for the inclusion of *Cermatobius* within *Paralamyctes* is not overwhelming, it is still a well-corroborated hypothesis because it is the least contradicted one under the analytical conditions examined.

The stability analysis of the henicopid molecular data set tells a completely different story. Figure 3 illustrates some graphic plots of sensitivity analyses ("Navajo rugs") for indel/change ratios of 1, 2, and 4, and transversion–transition ratios of 1, 2, 4, and infinity (transversion parsimony), where a solid square indicates monophyly, an open square indicates nonmonophyly, and a shaded square indicates monophyly for some of the shortest trees but not all. The first plot shows the genus *Paralamyctes* as monophyletic under all analytical conditions with the exception of equal weights. The second plot identifies monophyly of *Cermatobius* and the Henicopinae for 10 sets of parameters. As in the previous example, hypothesis choice is difficult in this case because *Paralamyctes* is not monophyletic under our favorite analytical conditions (the parameter set that maximizes overall congruence among partitions), but all the other models show its monophyly, agreeing with current taxonomy as well as with biogeographical hypotheses for henicopid phylogeny. The choice is tough, but showing the conflict seems extremely important to taxonomists. Given the choice, I would not include the genus *Cermatobius* within *Paralamyctes* in a revised taxonomy of the group unless more evidence for doing so is gathered. Stability under different parameters/models may well become the preferred criterion for taxonomic revision.

Example 3: Major Bivalve Clades

A third example of the difference between nodal stability and nodal support is based on an analysis of three loci (18S rRNA, 28S rRNA, and cytochrome *c* oxidase subunit I) and 183 morphological characters for a data set including 65 bivalves and 12 outgroup taxa attempting to address the relationships among the major bivalve clades (Giribet and Wheeler, 2002). There are several contentious issues in bivalve systematics, but two strike for their positive resolution using these new sources of data. The first issue refers to the Palaeoheterodonta, which includes the well-known freshwater pearl mussels (Unionoida) and the Trigonioida, an exclusive marine group consisting of the sole living circumaustralian genus *Neotrigonia* but having a rich fossil record including up to eight families (Beesley et al., 1998). The two groups have long been considered monophyletic by paleontologists based on shell ultrastructure and hinge morphology, but these characteristics are not shown in living species of unionoids, and thus most neontologists do not recognize the taxon Palaeoheterodonta (e.g., Purchon, 1987; Salvini Plawen and Steiner, 1996). Monophyly of Palaeoheterodonta has been recently examined based on sperm morphology (Healy, 1989) and molecular (Hoeh et al., 1998; Giribet and Wheeler, 2002) and morphological (Salvini Plawen and Steiner, 1996; Giribet and Wheeler, 2002) data analyses. While morphological analyses of living taxa fail to obtain monophyly of the Palaeoheterodonta, even when including the sperm characters that support such monophyly (Healy, 1989), molecular data clearly recognize the group. But more importantly, Palaeoheterodonta is the most stable group for every partition and combination of partitions for

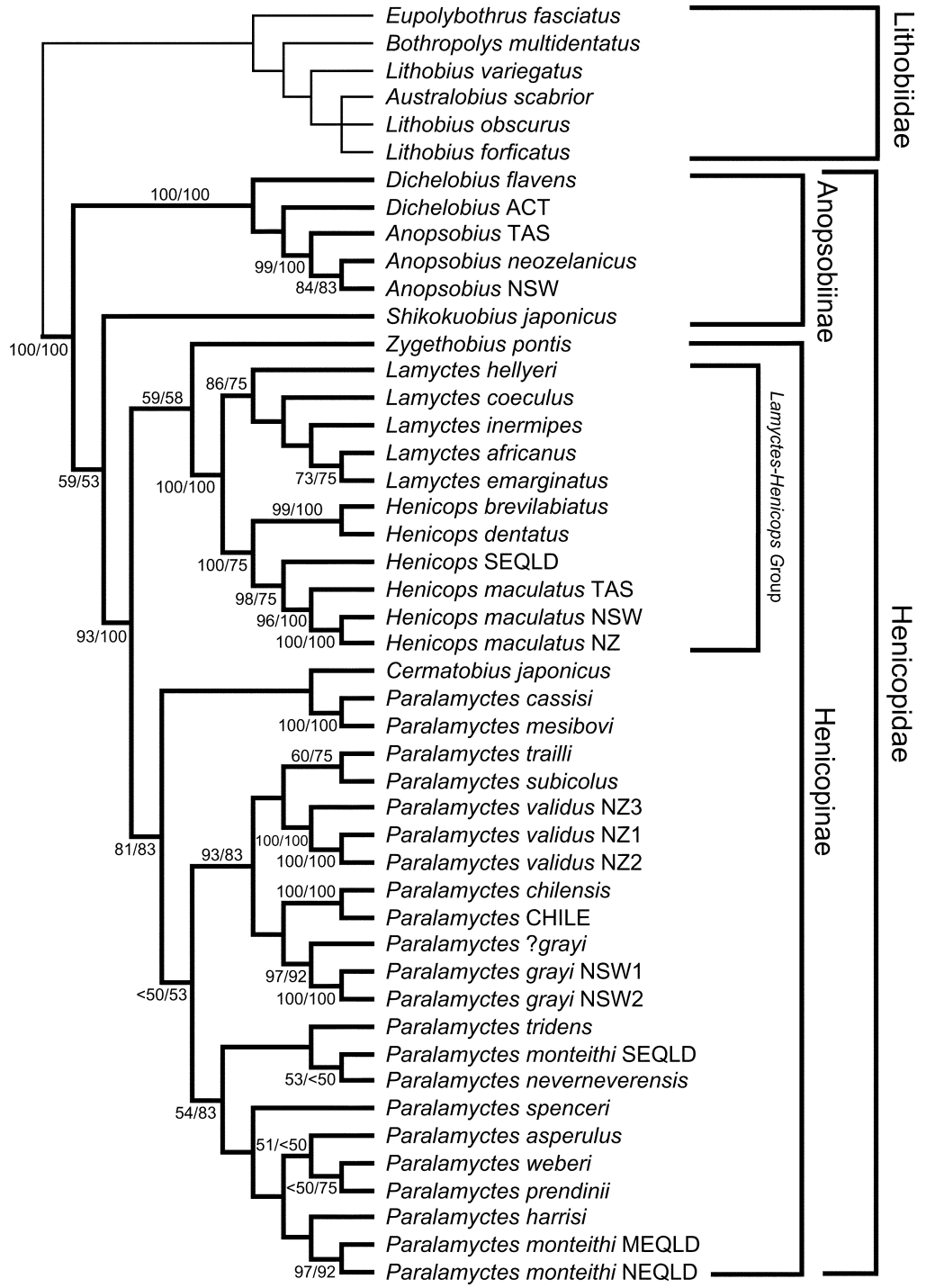

FIGURE 2. Phylogenetic tree of stone centipedes belonging to the Lithobiomorpha based on the combined analysis of five molecular loci (analysis based on direct optimization when all transformations, including gaps, are equally weighted; the parameter set that minimized incongruence among partitions). Thicker branches represent the ingroup taxa, the members of the family Henicopidae. Numbers on nodes represent jackknife percentages/stability values (for 12 parameter sets evaluated). For more details on the analyses and taxa, see Edgecombe and Giribet (2003).

each explored parameter set in the phylogenetic study of bivalve relationships from Giribet and Wheeler (2002), even when morphological analysis of 183 traits is unable to recognize this group. Palaeoheterodonta is thus a higher taxon (currently listed as a subclass) that benefits from a stability analysis. Monophyly of the group is however only moderately supported by jackknifing in the optimal tree for all data analyzed in combination; Palaeoheterodonta is found in 77% of subreplicates, with 25 other clades receiving higher support.

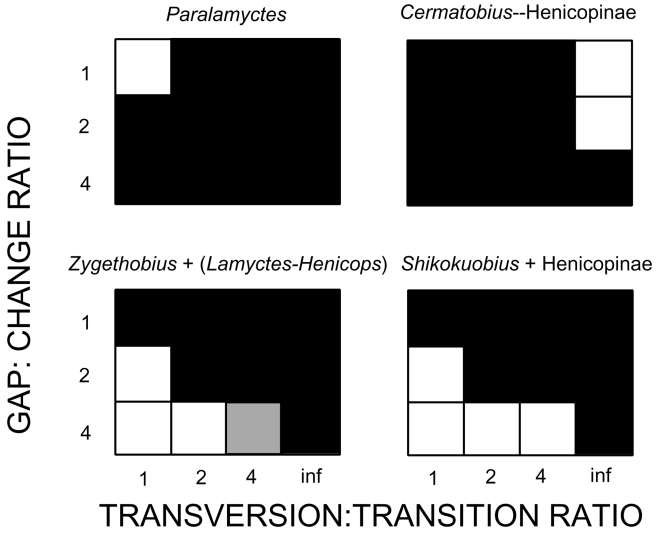

FIGURE 3. Four selected stability plots ("Navajo rugs") from the analysis of the centipede family Henicopidae illustrated in Figure 2. The two axes show the gap–change cost ratio ranging from 1 to 4 and the transversion–transtition cost ratio ranging from 1 to 4 and up to infinity (transversion parsimony). Solid squares represent monophyly for the taxon listed above each box; open squares indicate nonmonophyly; shaded squares represent monophyly in some but not all of the shortest trees: i.e., the group formed by *Zygethobius* + (*Lamyctes–Henicops*) is monophyletic (solid) under eight parameter sets and nonmonophyletic (open) under three parameter sets; one parameter set yields two alternative hypothesis for the position of *Zygethobius*.

A second interesting clade, and a complete different story, is the Anomalodesmata, a group of bivalves previously ranked as a subclass that includes mostly a set of deep-sea species with incredible adaptations such as the loss of ctenidia or predatory behavior utilizing the siphons as a suction pump. These and other drastic morphological changes are associated with accelerated molecular change (Giribet and Wheeler, 2002). However, all molecular data gathered so far point toward a phylogenetic position of the Anomalodesmata within the subclass Heterodonta (the largest group of bivalves, including many edible species such as cockles and littlenecks). Again, this stable result allowed the elimination of the subclass Anomalodesmata, which is now classified as a clade within the subclass Heterodonta. Inclusion of Anomalodesmata within a subclade of heterodonts is furthermore supported by a jackknife value of 83.

Finally, this tree has one clade that receives a jackknife value of <50% while is 100% stable to parameter variation: the Pteriomorphia, a clade containing oysters and mussels, among other "primitive" bivalves. Several other relationships are also represented in 83% of the parameter sets analyzed (10 of the 12 parameter sets) but receive jackknife support values below 50% (Fig. 4).

FINAL REMARKS

A linear relationship between nodal stability and certain measures of nodal support is certainly the case in many phylogenetic analyses. However, many examples exist in the literature showing cases like the ones presented above, decoupling support from stability in both directions: high stability with low support and low stability with high support (Table 1; Fig. 4). While the first example can be explained under the circumstance of few but uncontradicted evidence, the second may be a way to detect the pernicious long-branch-attraction (Felsenstein, 1978) phenomenon, certainly immune to the method of choice under more realistic situations than a four-taxon case (Pol and Siddall, 2001). The use of multiple models for a given method of tree analysis may well constitute the best and most secure way to detect artificially well-supported nodes, as shown in some of these examples illustrated above. It is also clear that in some instances support measures based on resampling techniques do better than Bremer support in detecting these situations, and in many respects they perform in a similar way to stability analyses. However, there are other cases where decoupling between the two measures is clear. It should also be noted that unlike support measures, stability could decrease as more and more parameters/models are explored, especially when departing from the areas of maximum congruence/agreement among data sets.

Some investigators may feel that resampling techniques (bootstrapping and jackknifing) fit the definition of sensitivity analysis because these affect the output of a model by providing sources of variation affecting the information fed into it. However, these techniques are not aimed at proposing alternative relationships but rather are designed to show the support for given nodes when the data (not all the analytical conditions) are altered. While any analytical variation (e.g., weights of characters) may fit the definition of sensitivity analysis, here we refer to a special case where alternative analytical parameters or models are explored simultaneously, the outcome of which is the stability analysis defined above.

Stability measures, sensitivity analysis, and model exploration have its detractors based on epistemology (Kluge, 1997a, 1997b; Frost et al., 2001b) and have even been described as unscientific (Grant, 2003). I see sensitivity analysis as the strictest possible test for a given phylogenetic hypothesis in the sense that it incorporates the greatest amount of possible refutation. Certain resampling techniques seem to be a fairly good approximation of this test, although not under all experimental situations. I hope to convince the reader that regardless of the "final" phylogenetic hypothesis that one aims to present, whether it is based on an arbitrary choice or on a more sophisticated or explicit way to choose models (e.g., congruence or a likelihood ratio test), it is a necessity to explore how those relationships may be affected by model/parameter choice. Using such measures in combination with the more traditional measures for nodal support clearly provides a better picture of the relationships presented.

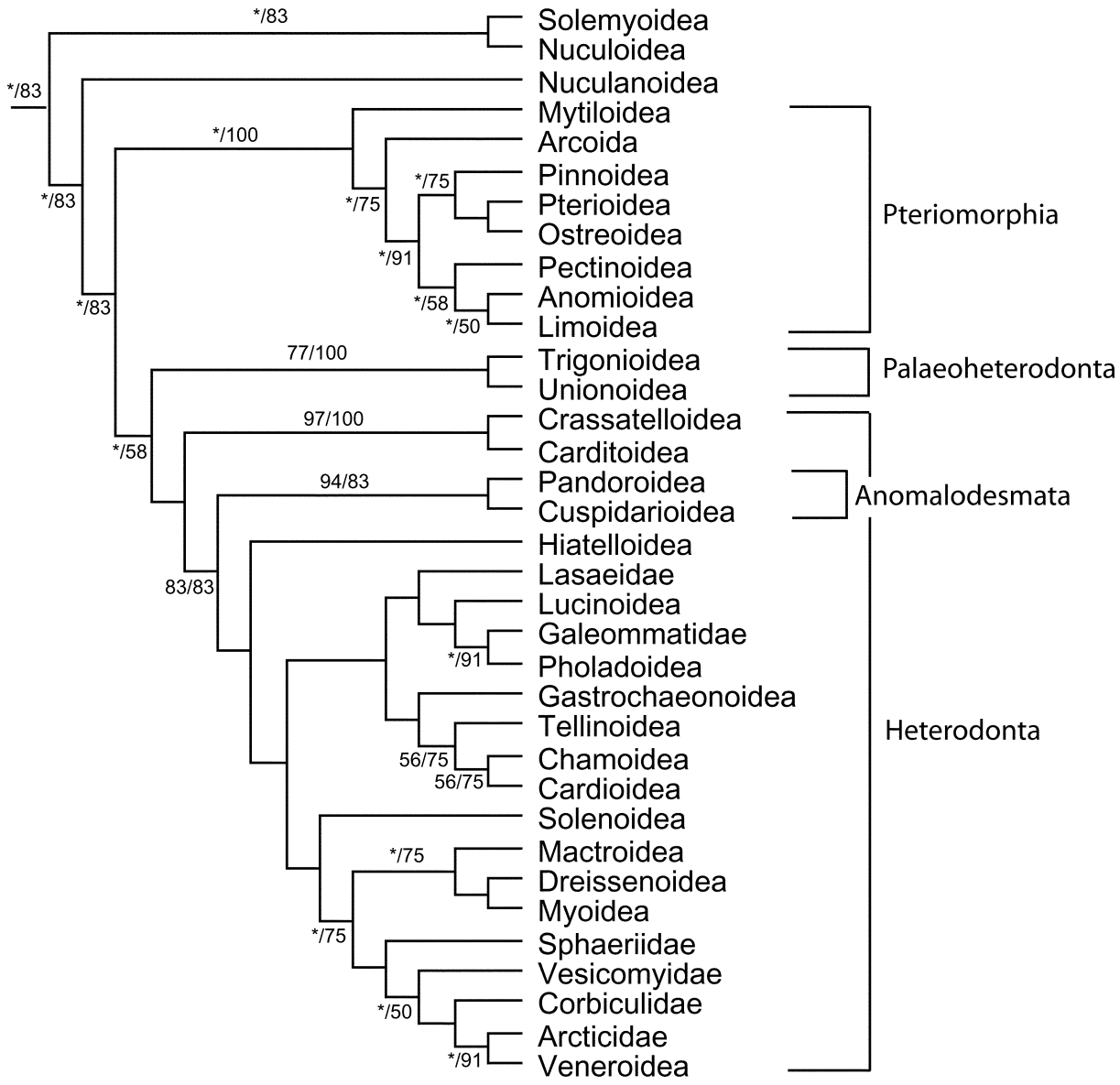

FIGURE 4. Tree summarizing the relationships among the main bivalve superfamilies (or other categories, when necessary) as published in the combined analysis of morphology and two nuclear ribosomal genes by Giribet and Wheeler (2002). Numbers on nodes represent jackknife percentages/stability values (for 12 parameter sets evaluated). An asterisk indicates values <50%.

ACKNOWLEDGMENTS

Roderic Page, Paul Lewis, and Mark Wilkinson reviewed the submitted version of the manuscript and provided criticisms that helped to improve or make clearer this point of view. Greg Edgecombe, Steve Farris, and Ward Wheeler provided comments and discussion on different stages of the preparation of the article. Chris Simon patiently dealt with the different revisions and provided insightful comments and discussion on many different points. The examples presented here are the fruit of my long-term collaborations with Greg Edgecombe and Ward Wheeler.

REFERENCES

ARNEDO, M. A., P. OROMÍ, AND C. RIBERA. 2001. Radiation of the spider genus *Dysdera* (Araneae, Dysderidae) in the Canary Islands: Cladistic assessment based on multiple data sets. Cladistics 17:313–353.

BANDELT, H. J., P. FORSTER, B. C. SYKES, AND M. B. RICHARDS. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141:743–753.

BEESLEY, P. L., G. J. B. ROSS, AND A. WELLS. 1998. Mollusca: The southern synthesis. CSIRO, Melbourne.

BELSHAW, R., AND D. L. J. QUICKE. 2002. Robustness of ancestral state estimates: Evolution of life history strategy in inchneumonid parasitoids. Syst. Biol. 51:450–477.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803.

BREMER, K. 1994. Branch support and tree stability. Cladistics 10:295–304.

BUCKLEY, T. R., AND C. W. CUNNINGHAM. 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. Mol. Biol. Evol. 19:394–405.

COGNATO, A. I., AND A. P. VOGLER. 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of *Ips* (Coleoptera: Scolytinae). Syst. Biol. 50:758–780.

DEBRY, R. W. 2001. Improving interpretation of the decay index for DNA sequence data. Syst. Biol. 50:742–752.
EDGECOMBE, G. D. 2001. Revision of *Paralamyctes* (Chilopoda: Lithobiomorpha: Henicopidae), with six new species from eastern Australia. Rec. Austr. Mus. 53:201–241.
EDGECOMBE, G. D., AND G. GIRIBET. 2003. Relationships of Henicopidae (Chilopoda: Lithobiomorpha): New molecular data, classification and biogeography. Afr. Invertebr. (in press).
EDGECOMBE, G. D., G. GIRIBET, AND W. C. WHEELER. 1999. Filogenia de Chilopoda: Combinando secuencias de los genes ribosómicos 18S y 28S y morfología [Phylogeny of Chilopoda: Combining 18S and 28S rRNA sequences and morphology]. Pages 293–331 *in* Filogenia y evolución de Arthropoda (A. Melic, J. J. de Haro, M. Méndez, and I. Ribera, eds.). Sociedad Entomológica Aragonesa, Zaragoza, Spain.
EDGECOMBE, G. D., G. GIRIBET, AND W. C. WHEELER. 2002. Phylogeny of Henicopidae (Chilopoda: Lithobiomorpha): A combined analysis of morphology and five molecular loci. Syst. Entomol. 27:31–64.
FARRIS, J. S. 1997. The future of phylogeny reconstruction. Zool. Scr. 26:303–311.
FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124.
FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.
FITCH, W. M., AND T. F. SMITH. 1983. Optimal sequence alignments. Proc. Natl. Acad. Sci. USA 80:1382–1386.
FROST, D. R., R. ETHERIDGE, D. JANIES, AND T. A. TITUS. 2001a. Total evidence, sequence alignment, evolution of polychrotid lizards, and a reclassification of the Iguania (Squamata: Iguania). Am. Mus. Novit. 3343:1–38.
FROST, D. R., M. T. RODRIGUES, T. GRANT, AND T. A. TITUS. 2001b. Phylogenetics of the lizard genus *Tropidurus* (Squamata: Tropiduridae: Tropidurinae): Direct optimization, descriptive efficiency, and sensitivity analysis of congruence between molecular data and morphology. Mol. Phylogenet. Evol. 21:352–371.
GIRIBET, G. 2001. Exploring the behavior of POY, a program for direct optimization of molecular data. Cladistics 17:S60–S70.
GIRIBET, G. 2002. Relationships among metazoan phyla as inferred from 18S rRNA sequence data: A methodological approach. Pages 85–101 *in* Molecular systematics and evolution: Theory and practice (R. DeSalle, G. Giribet, and W. Wheeler, eds.). Birkhäuser Verlag, Basel.
GIRIBET, G., AND S. BOYER. 2002. A cladistic analysis of the cyphophthalmid genera (Opiliones, Cyphophthalmi). J. Arachnol. 30:110–128.
GIRIBET, G., AND S. CARRANZA. 1999. What can 18S rDNA do for bivalve phylogeny? J. Mol. Evol. 48:256–261.
GIRIBET, G., R. DESALLE, AND W. C. WHEELER. 2002a. 'Pluralism' and the aims of phylogenetic research. Pages 141–146 *in* Molecular systematics and evolution: Theory and Practice (R. DeSalle, G. Giribet, and W. Wheeler, eds.). Birkhäuser Verlag, Basel.
GIRIBET, G., D. L. DISTEL, M. POLZ, W. STERRER, AND W. C. WHEELER. 2000. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: A combined approach of 18S rDNA sequences and morphology. Syst. Biol. 49:539–562.
GIRIBET, G., G. D. EDGECOMBE, AND W. C. WHEELER. 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413:157–161.
GIRIBET, G., G. D. EDGECOMBE, W. C. WHEELER, AND C. BABBITT. 2002b. Phylogeny and systematic position of Opiliones: A combined analysis of chelicerate relationships using morphological and molecular data. Cladistics 18:5–70.
GIRIBET, G., AND W. C. WHEELER. 1999. On gaps. Mol. Phylogenet. Evol. 13:132–143.
GIRIBET, G., AND W. C. WHEELER. 2002. On bivalve phylogeny: A high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. Invertebr. Biol. 121:271–324.
GOLOBOFF, P. A., AND J. S. FARRIS. 2001. Methods for quick consensus estimation. Cladistics 17:S26–S34.
GRANT, T. 2003. Against sensitivity analysis in phylogenetic systematics. Abstracts of the 21st annual meeting of the Willi Hennig Society. Cladistics 19:153.
HEALY, J. M. 1989. Spermiogenesis and spermatozoa in the relict bivalve genus *Neotrigonia*: Relevance to trigonioid relationships, particularly with Unionoidea. Mar. Biol. 103:75–85.
HICKSON, R. E., C. SIMON, AND S. W. PERREY. 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol. Biol. Evol. 17:530–539.
HIGGINS, D. G., AND P. M. SHARP. 1988. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244.
HOEH, W. R., M. B. BLACK, R. G. GUSTAFSON, A. E. BOGAN, R. A. LUTZ, AND R. C. VRIJENHOEK. 1998. Testing alternative hypotheses of *Neotrigonia* (Bivalvia: Trigonioida) phylogenetic relationships using cytochrome *c* oxidase subunit I DNA sequences. Malacologia 40:267–278.
HORMIGA, G., M. ARNEDO, AND R. G. GILLESPIE. 2003. Speciation on a conveyor belt: Sequential colonization of the Hawaiian Islands by *Orsonwelles* spiders (Araneae, Linyphiidae). Syst. Biol. 52:70–88.
HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, AND J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.
JANIES, D. 2001. Phylogenetic relationships of extant echinoderm classes. Can. J. Zool. 79:1232–1250.
JEANMOUGIN, F., J. D. THOMPSON, M. GOUY, D. G. HIGGINS, AND T. J. GIBSON. 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. 23:403–405.
JORDAN, S., C. SIMON, AND D. POLHEMUS. 2003. Molecular systematics and adaptive radiation of Hawaii's endemic damselfly genus *Megalagrion* (Odonata: Coenagrionidae). Syst. Biol. 52:89–109.
KJER, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: An example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4:314–330.
KLUGE, A. G. 1997a. Sophisticated falsification and research cycles: Consequences for differential weighting in phylogenetic systematics. Zool. Scr. 26:349–360.
KLUGE, A. G. 1997b. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13:81–96.
KNOX, E. B., AND J. D. PALMER. 1995. Chloroplast DNA variation and the recent radiation of the giant senecios (Asteraceae) on the tall mountains of eastern Africa. Proc. Natl. Acad. Sci. USA 92:10349–10353.
LEE, M. S. Y. 2000. Tree robustness and clade significance. Syst. Biol. 49:829–836.
LENTO, G. M., R. E. HICKSON, G. K. CHAMBERS, AND D. PENNY. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol. Biol. Evol. 12:28–52.
LOCKHART, P. J., P. A. MCLENACHAN, D. HAVELL, D. GLENNY, D. HUSON, AND U. JENSEN. 2001. Phylogeny, radiation, and transoceanic dispersal of New Zealand alpine buttercups: Molecular evidence under split decomposition. Ann. MO. Bot. Gard. 88:458–477.
MORRISON, D. A., AND J. T. ELLIS. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14:428–441.
PENNY, D., M. D. HENDY, AND M. A. STEEL. 1992. Progress with methods for constructing evolutionary trees. Trends Ecol. Evol. 7:73–79.
POL, D., AND M. E. SIDDALL. 2001. Biases in maximum likelihood and parsimony: A simulation approach to a 10-taxon case. Cladistics 17:266–281.
PRENDINI, L. 2000. Phylogeny and classification of the superfamily Scorpionoidea Latreille 1802 (Chelicerata, Scorpiones): An exemplar approach. Cladistics 16:1–78.
PURCHON, R. D. 1987. Classification and evolution of the Bivalvia: An analytical study. Philos. Trans. R. Soc. Lond. B 316:277–302.
SALTELLI, A. 2000. What is sensitivity analysis? Pages 3–13 *in* Sensitivity analysis (A. Saltelli, K. Chan, and E. M. Scott, eds.). John Wiley & Sons, Chichester, U.K.
SALVINI PLAWEN, L. V., AND G. STEINER. 1996. Synapomorphies and plesiomorphies in higher classification of Mollusca. Pages 29–51 *in*

Origin and evolutionary radiation of the Mollusca (J. D. Taylor, ed.). Oxford Univ. Press, Oxford, U.K.

SANCHIS, A., J. M. MICHELENA, A. LATORRE, D. L. QUICKE, U. GARDENFORS, AND R. BELSHAW. 2001. The phylogenetic analysis of variable-length sequence data: Elongation factor-1α introns in European populations of the parasitoid wasp genus *Pauesia* (Hymenoptera: Braconidae: Aphidiinae). Mol. Biol. Evol. 18:1117–1131.

SHULL, V. L., A. P. VOGLER, M. D. BAKER, D. R. MADDISON, AND P. M. HAMMOND. 2001. Sequence alignment of 18S ribosomal RNA and the basal relationships of adephagan beetles: Evidence for monophyly of aquatic families and the placement of Trachypachidae. Syst. Biol. 50:945–969.

SIDDALL, M. E. 2002. Measures of support. Pages 80–101 *in* Techniques in molecular systematics and evolution (R. DeSalle, G. Giribet, and W. Wheeler, eds.). Birkhäuser Verlag, Basel.

SLOWINSKI, J. B. 1998. The number of multiple alignments. Mol. Phylogenet. Evol. 10:264–266.

SORENSON, M. D., A. COOPER, E. E. PAXINOS, T. W. QUINN, H. F. JAMES, S. L. OLSON, AND R. C. FLEISCHER. 1999. Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. Proc. R. Soc. Lond. B 266:2187–2193.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.

THORLEY, J. L., AND M. WILKINSON. 1999. Testing the phylogenetic stability of early tetrapods. J. Theor. Biol. 200:343–344.

THORNE, J. L., H. KISHINO, AND J. FELSENSTEIN. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. J. Mol. Evol. 33:114–124.

TITUS, T. A., AND D. R. FROST. 1996. Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: Iguania). Mol. Phylogenet. Evol. 6:49–62.

WHEELER, W. C. 1994. Sources of ambiguity in nucleic acid sequence alignment. Pages 323–352 *in* Molecular ecology and evolution: Approaches and applications (B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, eds.). Birkhäuser Verlag, Basel.

WHEELER, W. C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44:321–331.

WHEELER, W. C. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12:1–9.

WHEELER, W. C. 1998. Sampling, groundplans, total evidence and the systematics of arthropods. Pages 87–96 *in* Arthropod relationships (R. A. Fortey and R. H. Thomas, eds.). Chapman & Hall, London.

WHEELER, W. C., AND D. S. GLADSTEIN. 1994. MALIGN: A multiple sequence alignment program. J. Hered. 85:417–418.

WHEELER, W. C., D. GLADSTEIN, AND J. DELAET. 2002. POY, version 3.0. Program and documentation available at ftp.amnh.org/pub/molecular. American Museum of Natural History, New York.

WHEELER, W. C., AND C. Y. HAYASHI. 1998. The phylogeny of extant chelicerate orders. Cladistics 14:173–192.

WHITING, M. F., S. BRADLER, AND T. MAXWELL. 2003. Loss and recovery of wings in stick insects. Nature 421:264–267.

WHITING, M. F., J. M. CARPENTER, Q. D. WHEELER, AND W. C. WHEELER. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46:1–68.

WILKINSON, M., AND M. J. BENTON. 1996. Sphenodontid phylogeny and the problems of multiple trees. Philos. Trans. R. Soc. Lond. B 351:1–16.

WILLIAMS, P. L., AND W. M. FITCH. 1989. Finding the minimal change in a given tree. Pages 453–470 *in* The hierarchy of life. Molecules and morphology in phylogenetic analysis (B. Fernhölm, K. Bremer, and H. Jörnvall, eds.). Elsevier Science, Amsterdam.

First submitted 20 September 2002; reviews returned 15 December 2002; final acceptance 25 March 2003

Associate Editor: Rod Page