

## Assessing Progress in Systematics with Continuous Jackknife Function Analysis

JEREMY A. MILLER

Department of Systematic Biology—Entomology, National Museum of Natural History, NHB-105, Smithsonian Institution, PO Box 37012, Washington, D.C., 20013-7012, USA, and Department of Biological Sciences, George Washington University, Washington, D.C., 20052, USA;  
E-mail: zmjermey@gwu.edu

**Abstract.**—Systematists expect their hypotheses to be asymptotically precise. As the number of phylogenetically informative characters for a set of taxa increases, the relationships implied should stabilize on some topology. If true, this increasing stability should clearly manifest itself if an index of congruence is plotted against the accumulating number of characters. Continuous jackknife function (CJF) analysis is a new graphical method that portrays the extent to which available data converge on a specified phylogenetic hypothesis, the reference tree. The method removes characters with increasing probability, analyzes the rarefied data matrices phylogenetically, and scores the clades shared between each of the resulting trees and the reference tree. As more characters are removed, the number of shared clades must decrease, but the rate of decrease will depend on how decisively the data support the reference tree. Curves for stable phylogenies are clearly asymptotic with nearly 100% congruence for a substantial part of the curve. Less stable phylogenies lose congruent nodes quickly as characters are excluded, resulting in a more linear or even a sigmoidal relationship. Curves can be interpreted as predictors of whether the addition of new data of the same type is likely to alter the hypothesis under test. Continuous jackknife function analysis makes statistical assumptions about the collection of character data. To the extent that CJF curves are sensitive to violations of unbiased character collection, they will be misleading as predictors. Convergence of data on a reference tree does not guarantee historical accuracy, but it does predict that the accumulation of further data under the sampling model will not lead to rapid changes in the hypothesis. [Accuracy; combined data; comparative biology; congruence; data visualization; jackknife; pseudoreplicate; stability; support; total evidence.]

The importance of evaluating support for phylogenetic hypotheses is evident in the number of indices that have been developed for this purpose, from simple descriptive statistics (e.g., Kluge and Farris, 1969; Archie, 1989b; Farris, 1989; Goloboff, 1991a) to support measures based on the variation of some aspect of the analysis (e.g., Felsenstein, 1985a; Bremer, 1988; Archie, 1989a; Faith, 1991; Faith and Cranston, 1991; Farris et al., 1996). Although support measures used in systematics are diverse in their methodology, all are intended to assess confidence in a clade, tree, or set of trees. Existing support measures are expressed as numerical values that may represent an index of conflict in the data, the frequency with which a clade is recovered from permuted data, the sensitivity of a clade to the relaxation of optimality criteria, or the probability that the data demonstrate more structure than expected under some null distribution. John Tukey, the eminent statistician who pioneered exploratory data analysis, once said that, “the **greatest value of a picture** is when it *forces* us to notice **what we never expected to see**” (1977:vi, emphasis in original). In that vein, a more exploratory approach would be to reveal graphically how the estimate of the tree topology changes with the addition of data. Systematists expect their hypotheses to be asymptotically precise, i.e., at some point addition of more characters should not affect the topology. The process of adding characters usually stops for practical rather than theoretical reasons; morphological and molecular characters both require time, money, and diverse skills to collect. Assessing just where we are on the road to systematic stability is the aim of continuous jackknife function (CJF) analysis.

CJF analysis graphically visualizes progress toward a stable systematic solution. The method tests a specific phylogenetic hypothesis with respect to a cladis-

tic data matrix. The phylogeny under study is usually a most-parsimonious tree or consensus tree based on all available data, referred to herein as the reference tree. CJF analysis simply reveals how the removal of increasing quantities of characters from a data matrix affects the rate of recovery of clades in the reference tree. The graph of nodes recovered against data removed shows the adequacy of data collection to date. As more characters are removed, the number of clades shared between trees from a rarefied matrix and a reference tree should decrease. The shape of the curve indicates the completeness of a phylogenetic investigation toward a stable result. Very stable phylogenies exhibit nested sets of synapomorphies that overwhelm character conflict. Such data can endure the exclusion of substantial portions of characters before phylogenetic signal begins to degrade. Graphing clade congruence as a function of character exclusion generates a curvilinear relationship showing nearly 100% congruence for the asymptotic part of the curve. However, less stable phylogenies will lose congruent nodes quickly as characters are excluded. Graphs of clade congruence from such data will be more nearly linear, or asymptotic with much less than 100% congruence. The shape of the CJF curve predicts the degree to which additional characters drawn from the same statistical universe as characters in the matrix can be expected to change the phylogeny.

CJF analysis is useful when evaluating a phylogenetic hypothesis as a whole. Methods such as bootstrap (Felsenstein, 1985a) and Bremer support (Bremer, 1988, 1994; DeBry, 2001) are commonly used to assess support for specific clades. The ensemble consistency index (CI; Kluge and Farris, 1969) and ensemble retention index (RI; Farris, 1989; Archie, 1989b) are commonly used measures of homoplasy for an entire tree. CJF analysis can

be used likewise to evaluate the support in a data set for a given hypothesis. Comparative biologists can use CJF analysis in cases where an evolutionary hypothesis depends on multiple clades within a tree, such as the repeated evolution of some characteristic. CJF analysis also can be used to evaluate the performance of different data to address the same phylogenetic question, such as with data partitions in a combined data (total evidence) analysis.

Jackknife methods generate pseudoreplicates by excluding a fixed fraction of the data. Either characters or taxa may be culled to make pseudoreplicates. Methods may remove one taxon or character at a time (first order) or more than one (higher order) (Siddall, 1995; see also Miller, 1974). Methods based on the jackknife with applications for systematics have been proposed using a variety of protocols for a variety of objectives (e.g., Mueller and Ayala, 1982; Lanyon, 1985; Davis, 1993; Farris et al., 1996; Farris, 2001). CJF analysis is the only jackknife-based method designed to produce a graph that portrays progress in a phylogenetic analysis. It is more similar in goals and presentation to Hillis's (1996, 1998) simulation studies than to any other jackknife method. Hillis predicted how many characters were necessary to solve a phylogenetic problem based on a model of evolution estimated from a real data set. As with CJF analysis, Hillis looked at how much of the tree was correct rather than how many trees were completely correct as his measure of congruence (contrast with Cummings et al., 1995; Sullivan and Swofford, 2001; Swofford et al., 2001). Unlike Hillis's method, CJF analysis does not predict how many characters will be required to stabilize a phylogenetic problem, but it does show the rate at which the hypothesis is changing as data are added. CJF analysis also differs by using rarefaction of real data rather than a model of character evolution.

Other biological disciplines rely on asymptotic curves to assess progress (Pielou, 1966). Ecologists use species accumulation curves to assess progress in sampling di-

versity. These curves track the total number of species collected over time. An asymptotic function is expected for studies where most species have been encountered (Pielou, 1975; Colwell and Coddington, 1994). CJF analysis analogously assesses progress in the collection of phylogenetically informative characters.

The expectation that accumulated character data should eventually converge on some solution has been explored previously in systematics. Penny and Hendy (1986) successfully tested their expectation that comparisons between nonoverlapping data partitions of increasing size should result in increasingly similar most-parsimonious trees. Other studies have expressed similar relationships graphically, plotting some measure of accuracy against data quantity. All have interpreted asymptotic behavior as indicating good performance (e.g., Hillis, 1996, 1998; Cummings et al., 1995; Sullivan and Swofford, 2001; Swofford et al., 2001).

## METHODS

### Data Sets Analyzed

CJF analysis was performed with a series of simulated data sets, an analysis of vertebrate relationships based on the entire mitochondrial genome (Cummings et al., 1995), an analysis of mammal relationships based on 12S RNA ribosomal (rRNA) (Emerson et al., 1999), selected morphological analyses of spider relationships at various taxonomic levels (Hormiga et al., 1995; Scharff and Coddington, 1997; Griswold et al., 1998), and two studies of ordinal relationships among arachnids and their allies (Shultz, 1990; Wheeler and Hayashi, 1998; Table 1).

*Simulations.*—The simulated data sets were generated using the Genesis module (Maddison and Maddison, 2001a) in Mesquite (Maddison and Maddison, 2001b). Data sets were composed of 16 ingroup taxa plus 2 outgroup taxa and 180 DNA characters. One series of simulated data sets was based on a pectinate topology, and

TABLE 1. Published data matrices used in this study. Density is the ratio of phylogenetically informative characters to taxa.  $N_{\max}$  refers to the maximum number of resolved nodes in any most-parsimonious tree with a basal trichotomy; this is the denominator used to scale the congruence score. MPTs is the number of most-parsimonious trees for the complete data set.  $S_{50}$  and  $S_{90}$  are the CJF scores set at the 50% and 90% levels, respectively. CJF score is calculated as the minimum amount of data removal at which at the specified congruence level is exceeded. A dash indicates undefined CJF score. CI is the consistency index (calculated using informative characters only), RI is the retention index.

Data set	Authors	No.		Density	$N_{\max}$	No.		$S_{50}$	$S_{90}$	CI	RI
		characters	informative characters			taxa	MPTs				
Vertebrate mitochondria	Cummings et al., 1995	16,075	7,185	10	718.50	7	1	98	88	0.57	0.42
Mammal 12S rRNA <sup>a</sup>	Emerson et al., 1999	1,038	604	118	5.12	115	6	78	0	0.15	0.54
Mammal 12S rRNA <sup>b</sup>	Emerson et al., 1999	1,038	604	118	5.12	115	30	76	—	0.15	0.54
Orbicularian families	Griswold et al., 1998	91	91	31	2.94	28	1	68	10	0.64	0.81
Tetragnathid genera	Hormiga et al., 1995	60	58	22	2.64	19	3	52	4	0.56	0.72
Araneid genera	Scharff and Coddington, 1997	83	81	70	1.16	66	12	18	0 <sup>c</sup>	0.35	0.74
Arachnid orders	Shultz, 1990	64	59	12	4.92	9	1	70	18	0.57	0.64
Chelicerate morphology	Wheeler and Hayashi, 1998	93	79	17	4.65	12	9	52	—	0.52	0.64
Chelicerate molecules	Wheeler and Hayashi, 1998	1,276	378	33	11.45	16	1	0	—	0.39	0.41
Chelicerate combined data	Wheeler and Hayashi, 1998	1,369	470	34	13.82	17	7	74	12	0.41	0.50

<sup>a</sup>Emerson et al. (1999) reported finding 30 MPTs; the 6 trees used here as reference trees are one step shorter. The strict consensus of the six MPTs has the same  $S_{50}$  and  $S_{90}$  scores as the individual MPTs.

<sup>b</sup>Reference tree is the strict consensus of 30 trees reported by Emerson et al. (1999: fig. 1).

<sup>c</sup>The strict consensus of the 16 MPTs of Scharff and Coddington (1997) is undefined for the  $S_{90}$  score.

the other had the ingroup taxa arranged in a completely balanced topology. Simulations were based on an equal frequencies, equal rates model (Jukes and Cantor, 1969), with all branches of equal length. Scaling factors, which modify evolutionary rate, were varied to show how CJF curves should appear under conditions of light, moderate, and extreme homoplasy. Five data sets were generated for each of three scaling factors (0.1, 0.5, and 1.0) and for each topology.

*Empirical data.*—Cummings et al. (1995) investigated the phylogeny of 10 vertebrate taxa based on the entire mitochondrial genome. The aligned matrix had >7,000 phylogenetically informative characters. The modest number of taxa, the large number of characters, and the assessment of stability by Cummings et al. (1995) indicated that this data set should be quite stable to the removal of data. Their single most-parsimonious tree was used as the reference tree (Cummings et al., 1995: fig. 1).

Emerson et al. (1999) investigated 118 complete 12S rRNA sequences representing all 18 eutherian orders and 3 metatherian orders and recovered 30 most-parsimonious trees, all of which indicate controversial results such as polyphyly of both primates and bats. Citing alignment ambiguities and low parsimony jackknife support, Emerson et al. concluded that 12S rRNA sequence data alone lacked the ability to resolve relationships among mammalian orders. My reanalysis of Emerson et al.'s data using the ratchet (Nixon, 1999) as implemented in WinClada (Nixon, 2002) produced 6 most-parsimonious trees one step shorter than those reported by Emerson et al. I performed CJF analyses on Emerson et al.'s data using each of the 6 most-parsimonious trees, the strict consensus of the 6 most-parsimonious trees, and the consensus of 30 trees reported by Emerson et al. (1999: fig. 1) as reference trees.

I investigated progress in the phylogeny of orb-weaving spiders and their relatives by selecting three analyses: relationships among orb-weaving families and their descendents (Griswold et al., 1998), genera of the family Tetragnathidae (Hormiga et al., 1995), and genera of the family Araneidae (Scharff and Coddington, 1997). The orb-weaving families analysis yielded one most-parsimonious tree, which was used as the reference tree (Griswold et al., 1998: fig. 7). The tetragnathid genera analysis yielded three most-parsimonious trees, with conflict at one outgroup node. Hormiga et al. (1995: fig. 30) preferred one of these trees based on its congruence with another analysis, and this preferred tree was used as the reference tree. The araneid genera analysis produced 16 most-parsimonious trees but included problematic taxa that jumped between distant parts of the tree. Scharff and Coddington did select a preferred working hypothesis. This preferred tree (Scharff and Coddington, 1997: fig. 82) and the strict consensus tree (Scharff and Coddington, 1997: fig. 81) were both used as reference trees.

Shultz (1990) investigated arachnid relationships based on morphology. This analysis employed novel character systems and resulted in controversial conclusions, especially concerning the placement of *Scorpiones*

(Selden, 1993; Dunlop and Webster, 1999). The data implied a single most-parsimonious tree, which was used as the reference tree (Shultz, 1991: fig. 4).

Wheeler and Hayashi (1998) considered both molecular (18S rDNA and 28S rDNA) and morphological evidence in their study of chelicerate phylogeny. Their analysis was performed using OY (Gladstein and Wheeler, 1996), which produces phylogenies directly from sequence data (OY is a precursor to POY; Gladstein and Wheeler, 2001). Their preferred phylogeny was generated under a weighting scheme that had insertion-deletion events and morphological characters weighted twice compared with base substitutions and transversions weighted twice compared with transitions. They also generated an implied alignment. I used their implied alignment with all characters and state changes equally weighted and with gaps and polymorphic sites treated as unknown. The equally weighted combined data analysis of Wheeler and Hayashi generated seven most-parsimonious trees, six of which were congruent with their preferred hypothesis of ordinal relationships. The one incongruent tree suggested an alternative placement for palpigrades, which were not represented in the molecular data and may therefore have suffered from missing data problems. I ran three analyses of the Wheeler and Hayashi data: the morphology alone, the molecular data (both genes) alone, and all data in combination. For all three analyses, the reference tree was based on the topology for the combined data analysis (Wheeler and Hayashi, 1998: fig. 10). Because no sequences of Palpigradi were available to Wheeler and Hayashi, this taxon was excluded from the reference tree for the molecular data alone. When molecular data were involved, the reference tree had exemplars from each order collapsed into a polytomy so that only ordinal relationships were reflected; relationships of exemplars within orders were not a focus of Wheeler and Hayashi's study. The reference tree also did not resolve relationships among outgroup taxa.

#### *Sampling Data*

*NONA phase.*—Rarefied data matrices were created using a batch file executed in NONA (Goloboff, 1993) (see Appendix 1 for settings and search strategy, available online at the *Systematic Biology* website: <http://systbiol.org/>). Characters were randomly removed from the data matrix, with the probability of character removal set between 2% and 98% in increments of 2%. One hundred replicates of each level of character removal were performed. A phylogenetic analysis then was performed on each rarefied matrix. In total, 4,900 rarefied matrices were analyzed for each complete data set. The congruence score of each tree from the rarefied analyses was calculated by comparing it with the reference tree.

*PEST phase.*—Congruence was calculated using PEST (Zujko-Miller and Miller, 2002). Groups were counted as congruent when their composition, but not necessarily their internal arrangement, was identical (Fig. 1). Each

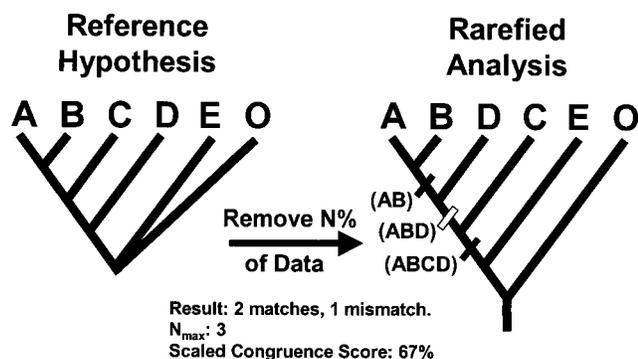


FIGURE 1. Criteria for assessing congruence between a reference tree and a tree from a rarefied analysis. Nodes were counted as congruent or incongruent depending on whether they contained the same set of taxa but without regard for the internal arrangement of those taxa. Thus, node ABCD is considered congruent even though the internal arrangement of C and D are different. This difference is factored into the analysis because there is no node ABC in the rarefied analysis. Reference trees have a basal trichotomy to avoid counting the "free" clade ABCDE. In this example, comparison between the tree from the rarefied analysis and the reference hypothesis identifies two congruent nodes (solid bars). This result is divided by  $N_{\max}$ , the maximum number of resolved nodes in any most-parsimonious tree with a basal trichotomy. Thus, if any most-parsimonious tree is fully resolved,  $N_{\max}$  equals the number of taxa minus 3. The number of congruent nodes divided by  $N_{\max}$  is the scaled congruence score, which is expressed as a percentage.

unique most-parsimonious tree from a rarefied matrix was compared with the reference tree. Thus, each data point in Figures 2–4 represents the average congruence of at least 100 trees (and often more) with the reference tree. The final data point was provided by comparing all most-parsimonious trees from the complete data set with the reference tree.

For CJF, congruence is defined as the number of nodes shared between the rarefied tree and the reference tree (Fig. 1). Congruence is scaled for plotting on a graph by dividing by  $N_{\max}$ , the maximum number of dichotomous nodes found in the most resolved most-parsimonious unrooted tree. This quotient is the scaled congruence score and is expressed as a percentage (Table 1). When the reference tree is fully resolved, the scaled congruence score is equivalent to the average percentage of nodes in the reference tree recovered from the rarefied matrix. This scaling device insures that a poorly resolved reference tree (e.g., a strict consensus tree) will not lead to an inflated curve by excluding conflicted nodes. The scaled congruence score guarantees that a consensus tree used as the reference tree will never perform better on a CJF graph than any of the most-parsimonious trees used to generate it. If congruence were scaled by the maximum possible number of resolved nodes (number of taxa minus 3, including a basal trichotomy), the curve could be inflated by absence of data to discriminate among taxa rather than conflict in the data. For example, the addition of an arbitrary number of identically coded taxa would result in a polytomy (Wilkinson, 1995), driving the curve down. If more characters were added and some of

the taxa continued to be identical for all characters, the tree could become stable in spite of its polytomy, i.e., the tree would not change with the addition of more equivalent data. Only lack of resolution due to data conflict should be able to degrade the CJF curve.

When necessary, reference trees were emended to include a basal trichotomy (Fig. 1), which is equivalent to an unrooted network. NONA always draws the first taxon as sister to the remaining taxa in the analysis, but the convention creates a "free" clade in all trees, regardless of data, and thus inflates the curve.

The results of CJF analysis can also be summarized numerically by reporting the minimum amount of data removal at which the average scaled congruence score reaches a predetermined value,  $P$ . This is the CJF score, and it facilitates the interpretation of CJF curves (Table 1).  $P$  can be set to any appropriate value, such as 50% for average data sets or 90% for more stable data sets. The CJF score is undefined if the analysis never reaches the critical level. The curve itself depicts the robustness of the hypothesis under test in a graphic way that a number cannot match, but the CJF score fulfills the need to compare or describe data sets numerically. The CJF score inevitably discards much of the information and can be misleading if not considered with reference to the graph.

## RESULTS

The results from simulated data sets are plotted in Figure 2, the vertebrate mitochondrial genome (Cummings et al., 1995) and mammal 12S rRNA (Emerson et al., 1999) data sets are plotted in Figure 3, and the spider morphology (Hormiga et al., 1995; Scharff and Coddington, 1997; Griswold et al., 1998) and the arachnid/chelicerate (Shultz, 1990; Wheeler and Hayashi, 1998) data sets are plotted in Figure 4. Using an 800-MHz Pentium III computer, the NONA phase of the analysis usually took between 12 and 48 hr (the mammal 12S rRNA data took about 3 weeks); the PEST phase took between 1 min and 8 hr.

The curves from the simulated data sets illustrate a broad range of the possible results from CJF analysis. All simulations with a scaling factor of 0.1 converged on the reference tree with a substantial portion of the data removed. The  $S_{50}$  (CJF score at the 50% level) was 86–88 for the pectinate topology, and 86–90 for the balanced topology, and the  $S_{90}$  (CJF score at the 90% level) was 56–70 for the pectinate topology and 54–62 for the balanced topology. Topology affected results more under scaling factor 0.5. Pectinate topology CJF curves tended to be more or less straight, showing little sign of concavity. Balanced CJF curves decreased but recovered a smaller proportion of nodes when most of the data set was retained. The  $S_{50}$  ranged from 56 to 60 for the pectinate topology and 70 to 76 for the balanced topology. The  $S_{90}$  ranged from 4 to 24 for the pectinate topology, but the  $S_{90}$  level was achieved in only one of the five replicates for the balanced topology, scoring 10. No simulation at the 1.0 level attained 50% of the nodes in the reference tree. Curves

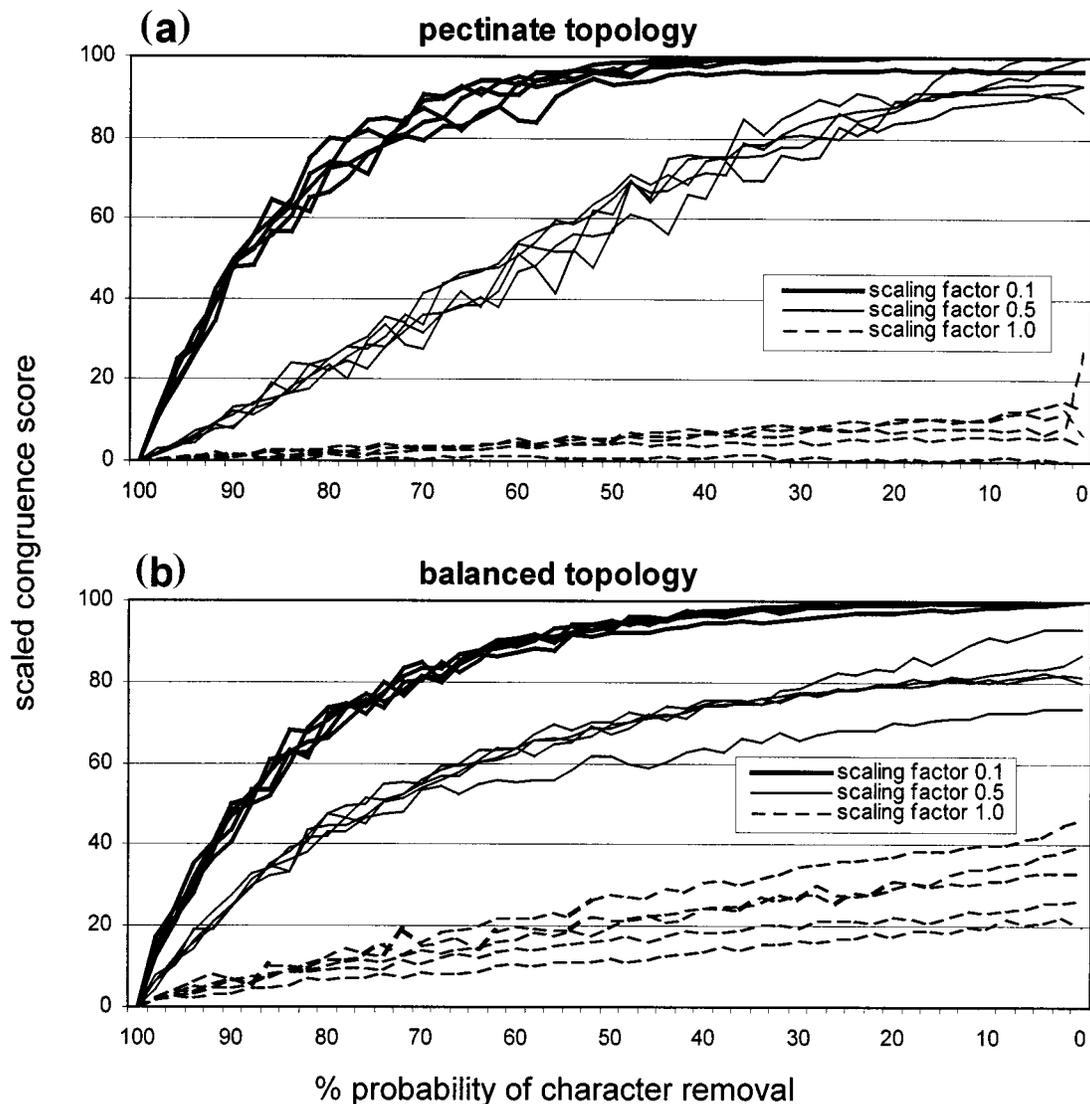


FIGURE 2. Continuous jackknife function graphs of simulated data arranged to give a pectinate topology (a) and arranged to give a balanced topology (b). Curves show the average scaled congruence score with respect to a reference tree as the percent probability of removing each character from the data matrix increases. Simulated data sets were composed of 16 ingroup taxa plus 2 outgroup taxa and 180 DNA characters. Simulated data were generated using an equal rates, equal base frequencies, equal branch lengths model with the scaling factor set to 0.1, 0.5, or 1.0. Five replicate data sets using each of the three scaling factors were generated for both topologies.

in the pectinate topology increased at a markedly slower rate compared with curves in the balanced topology.

The phylogenetic analysis of 10 vertebrates using the complete mitochondrial genome (Cummings et al., 1995) clearly converged on the reference tree (Fig. 3a). The  $S_{50}$  was calculated as 98, but the graph indicates that the actual value is higher. Less than 2% of the mitochondrial genome is necessary to recover more than 50% of the nodes supported by the entire genome for these 10 taxa. A more precise  $S_{50}$  value requires finer deletion steps than 2%. The  $S_{90}$  was 88. The CJF scores for the Cummings et al. data exceed those for all other real data sets examined for this study. This stability is due in part to the data matrix dimensions, with >700 times as many informative characters as taxa.

Despite poor parsimony jackknife support and some peculiar relationships, CJF analysis indicates a fair amount of structure in Emerson et al.'s (1999) data (Fig. 3b). Although these data do not quickly converge on the reference trees, they might well make a useful contribution to a larger analysis incorporating other genes and/or morphology. Problems with alignment ambiguity reported by Emerson et al. might be alleviated through reference to secondary structure (e.g., Kjer, 1995; Titus and Frost, 1996; Hickson et al., 1996, 2000; Morrison and Ellis, 1997; Buckley et al., 2000; Page, 2000). A total of eight reference trees were tested: 6 most-parsimonious trees one step shorter than those reported by Emerson et al., the strict consensus of these 6 trees, and the strict consensus of 30 trees reported by Emerson et al.

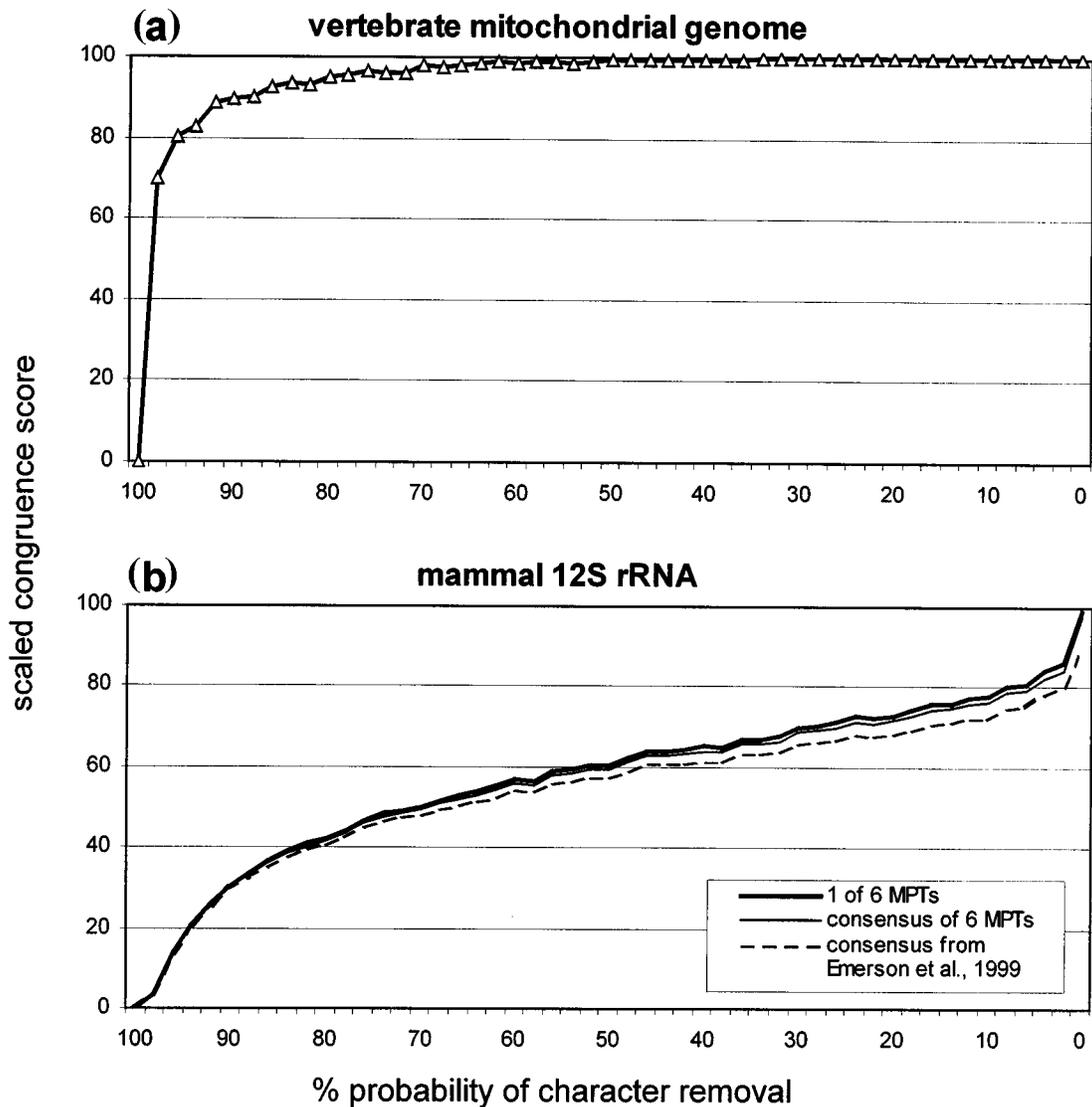


FIGURE 3. Continuous jackknife function graphs for real data. (a) Analysis of vertebrate relationships based on the entire mitochondrial genome (Cummings et al., 1995). (b) Analysis of mammal relationships based on 12S rRNA (Emerson et al., 1999). Eight different topologies were tested using the Emerson et al. data: 6 most-parsimonious trees (MPTs; one shown), the strict consensus of 6 most-parsimonious trees, and the strict consensus of 30 trees (one step longer than the most-parsimonious trees) reported by Emerson et al. (1999). Curves for all six most-parsimonious trees were nearly identical. Curves drawn as in Figure 2.

Curves for all six most-parsimonious trees are nearly identical (curve for one of six trees shown in Fig. 3b). The curve based on the strict consensus of six most-parsimonious trees is slightly lower than that for any single most-parsimonious tree but is indistinguishable using  $S_{50}$  or  $S_{90}$  values. The curve based on the strict consensus of 30 most-parsimonious trees reported by Emerson et al. performs slightly worse than the most-parsimonious trees or their consensus. All six most-parsimonious trees and their consensus have an  $S_{50}$  score of 78 and an  $S_{90}$  score of 0 (i.e., congruence exceeds 90% only when no data are excluded). The strict consensus of 30 longer trees has an  $S_{50}$  score of 76; the  $S_{90}$  score is undefined.

The spider data sets show that some aspects of spider phylogeny are understood better than others (Fig. 4a). The orbicularian families (Griswold et al., 1998) CJF curve is slightly concave and never becomes flat ( $S_{50} = 68$ ,  $S_{90} = 10$ ). Although relationships among the orbicularian taxa represented by Griswold et al. (1998) are reasonably stable, they may be expected to change in minor ways as new data are contributed to the problem. The tetragnathid genera (Hormiga et al., 1995) CJF curve is a more or less straight line ( $S_{50} = 52$ ,  $S_{90} = 4$ ). Again, additional data will probably change the topology, probably more drastically than in the former example. The araneid data set (Scharff and Coddington, 1997) is clearly the most incomplete analysis and substantially more data

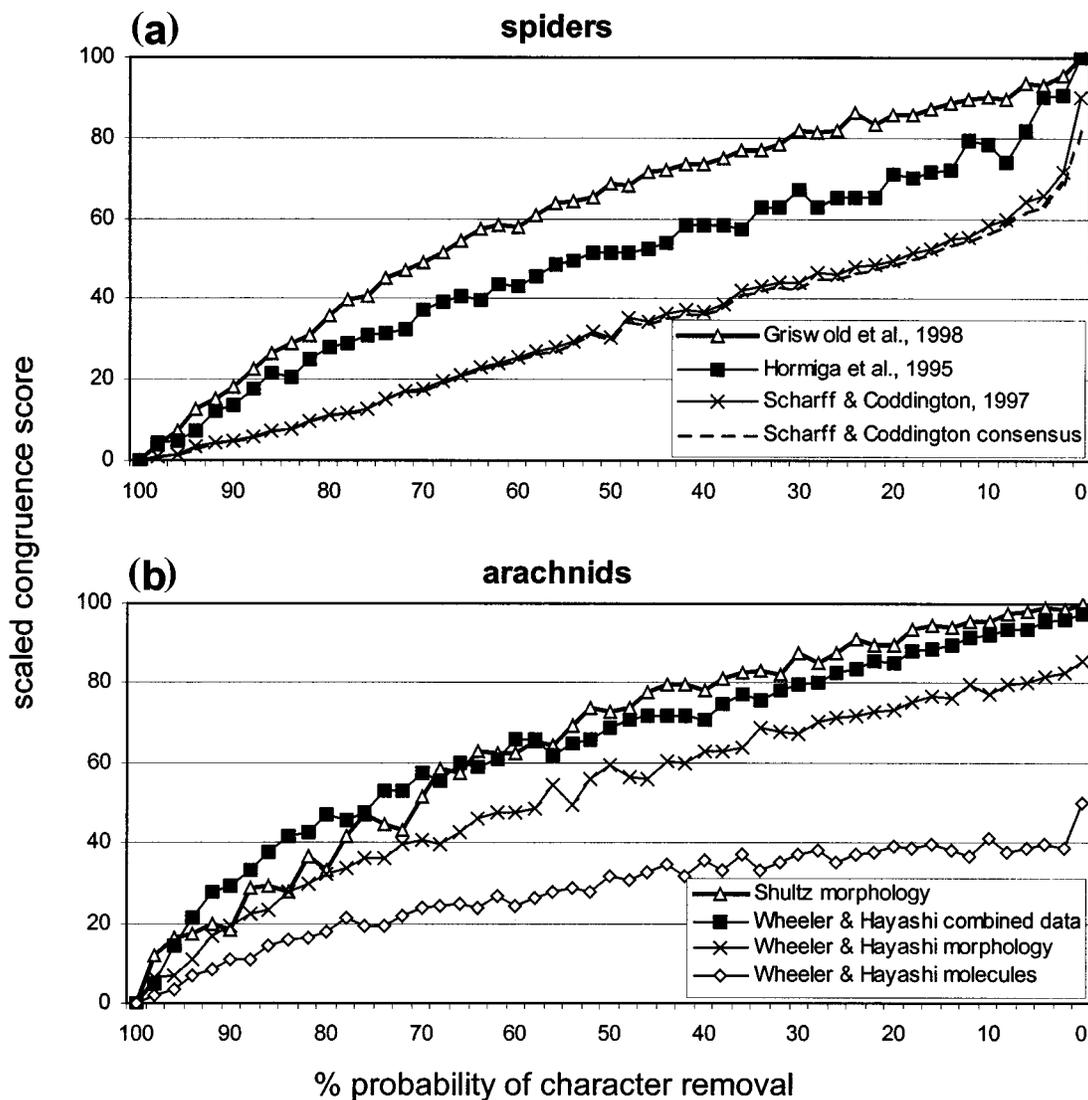


FIGURE 4. Continuous jackknife function graphs for real data. (a) Three analyses of spider relationships: orbicularian families (Griswold et al., 1998), tetragnathid genera (Hormiga et al., 1995), and araneid genera (Scharff and Coddington, 1997). Two trees were tested using the Scharff and Coddington (1997) data, the preferred of 16 most-parsimonious trees and the strict consensus tree. (b) Two analyses of arachnid/chelicerate ordinal relationships: Shultz's (1990) analysis based on morphological characters and Wheeler and Hayashi's (1998) analysis based on both morphology and nuclear ribosomal genes (18S and 28S rDNA). The Wheeler and Hayashi data were partitioned into morphology alone, molecules alone (both genes), and all data combined. Curves drawn as in Figure 2.

are required for topological stability. Scharff and Coddington were well aware that their analysis was unlikely to be the last word on araneid phylogeny. They acknowledged that "support for many nodes is weak" (1997:419) and that "the phylogeny presented here may soon require revision" (1997:404). The CJF curve for the strict consensus tree is only slightly worse than the preferred hypothesis. Both curves have an  $S_{50}$  score of 18. The curve based on the preferred tree crosses the  $S_{90}$  threshold only when no data are excluded from the matrix;  $S_{90}$  is undefined for the curve based on the strict consensus tree.

The chelicerate CJF curves show that Wheeler and Hayashi's (1998) incorporation of ribosomal gene data

did not substantially improve on Shultz's (1990) analysis using only morphology (Fig. 4b). The morphological component of Wheeler and Hayashi's analysis was drawn largely from Shultz's analysis but also from other literature sources. Wheeler and Hayashi also expanded the representation of outgroup taxa. Shultz rooted his analysis using a hypothetical ancestor based on several living and fossil taxa. The use of a hypothetical ancestor may have led to an underestimate of the amount of homoplasy that is relevant to the question of early chelicerate evolution. Wheeler and Hayashi's morphological matrix included extant outgroup taxa and contained more character conflict than Shultz's. The additional character conflict rendered several deep nodes in the arachnid tree

ambiguous. Wheeler and Hayashi's morphology data alone ( $S_{50} = 52$ ,  $S_{90}$  undefined) perform distinctly worse than Shultz's morphology data ( $S_{50} = 70$ ,  $S_{90} = 18$ ), although both curves are slightly concave. The molecular data alone cross the  $S_{50}$  threshold only when no data are excluded from the matrix. These data consistently recover just <40% of the nodes when up to a third of the data are removed. This type of leveling off may indicate that the molecular data are consistently recovering a fraction of the nodes in the chelicerae tree. However, this analysis indicates that the genes selected are quite inefficient for recovering the entire reference tree on their own. Despite the relatively poor performance of Wheeler and Hayashi's data partitions independently, the combined data analysis indicates a degree of stability that is roughly on a par with Shultz's analysis ( $S_{50} = 74$ ,  $S_{90} = 12$ ). The superior performance of the combined data over both the morphological and molecular data as indicated by the curves generated may be a visualization of secondary signal, an idea advocated as an advantage of simultaneous data analysis (Kluge and Wolf, 1993; Nixon and Carpenter, 1996). The only topological difference between the conclusions of Shultz and those of Wheeler and Hayashi is in the relationships between the Araneae and Amblypygi. In Shultz's analysis, Araneae is sister to a clade containing Amblypygi and two other orders; in the analysis of Wheeler and Hayashi, Araneae and Amblypygi are sister taxa.

My reanalysis of Wheeler and Hayashi's (1998) data did not employ their favored weighting scheme. Up-weighting morphology and tranversions relative to transitions probably would have improved the combined data and molecular curves, but this approach was not possible to duplicate using NONA. Instead, the CJF curves assess agreement between the reference tree and the accumulated morphological and aligned ribosomal gene characters, counting each line of evidence equally (Kluge, 1997a).

#### DISCUSSION

The examples above show how CJF can be applied in real cases. The most general application is as a method to evaluate support (Fig. 3a). Unlike the bootstrap (Felsenstein, 1985a) or Bremer support (Bremer, 1988, 1994), which assess support for specific clades, CJF analysis graphically dissects the depth of support for the overall phylogeny. CI and RI statistics, in contrast, capture the final support for an overall phylogeny and express it numerically. Interpretation of the CI and RI can be difficult, especially when one is interested in comparisons between data sets (Sanderson and Donoghue, 1989; Goloboff, 1991b), but CJF facilitates such comparisons (Figs. 4a, 4b). CJF is intended to supplement numerical support indices and to enrich the ways in which systematists can understand the relationship between data and results.

Comparative biologists can use CJF analysis in cases where an evolutionary hypothesis depends on multiple clades within a tree, such as the repeated evolution of

some characteristic or the evolution of a continuous variable. Hormiga et al. (2000; see also Coddington et al., 1997) used the topologies of the three data sets depicted in Figure 4a to discuss the evolution of sexual size dimorphism in orbicularian spiders. Bond and Opell (1998) relied in part on Griswold et al.'s (1998) topology for their study of silk and web architecture traits and clade imbalance. Phylogenetic hypotheses are fundamental to such work, and a measurement of the strength of those foundations, numerically or graphically, is important.

CJF analysis also can be used to evaluate how different data or studies contribute to the same phylogenetic question (Fig. 4b). The Wheeler and Hayashi (1998) combined data analysis built upon and modified the results of Shultz (1990). CJF analysis can also parse the contribution of components in a combined data analysis. There is considerable disagreement over when, if ever, data should be analyzed separately versus simultaneously (e.g., Kluge, 1989, 1997b; Barrett et al., 1991; Bull et al., 1993; Eernisse and Kluge, 1993; Kluge and Wolf, 1993; Chippindale and Wiens, 1994; Huelsenbeck et al., 1994; de Queiroz et al., 1995; Miyamoto and Fitch, 1995). Data visualization using CJF analysis could be a useful tool for evaluating this question.

#### Interpretation

Does CJF analysis predict that as more data are added curves will be increasingly asymptotic? That presumes a sampling model in which new characters are drawn randomly from the same universe of characters as the original data, certainly an erroneous supposition in many contexts. Systematists do not gather characters at random, and this fact has long been recognized as a fundamental challenge to the application of statistics in systematics (Felsenstein, 1985a; Sanderson, 1989; Carpenter, 1992, 1996; Kluge and Wolf, 1993; Cummings et al., 1995). However, the nonrandom character sampling critique applies to many statistical support measures, including the bootstrap. To the extent that new work samples from drastically different new universes, or from old universes differently, CJF curves will fail to predict the future. Averaged over many replicates, the CJF curve represents some mean data collecting effort under the assumption of random character sampling. Addition of characters in small increments more closely approximates the process of accumulating morphological data than accumulating sequence data, which involves adding blocks of contiguous regions from the genome. Nevertheless, CJF analysis should be equally applicable to both types of data.

Two attributes of CJF curves are useful in their interpretation: the degree to which the curve is decreasing and the height of the curve above the abscissa. Decreasing functions that flatten out with nearly 100% of nodes recovered should be interpreted as superadequate efforts to address a phylogenetic problem. Curves that flatten out with far less than 100% of nodes recovered may be reliably recovering a set of specific nodes while other parts of the tree are ambiguous. Curves that are essentially straight or even increasing demonstrate

instability and suggest that more data should be added. The final data point, with no data excluded, deviated strongly in a positive direction from the overall trend in several data sets probably because the final data point is not subjected to permutation and represents the very data set that generates the reference tree. This artifact can be neglected when evaluating the overall trend of the CJF curve.

Experience indicates that equally parsimonious reference trees tend to result in similar curves. By design, strict consensus trees never have a higher scaled congruence score than that of any most-parsimonious tree, but their score does not seem to be much lower either. This is probably because nodes lost in the strict consensus tree based on all the data are also usually the first to go in rarefied matrices.

#### *Stability or Accuracy?*

CJF analysis does not assess historical accuracy; it simply indicates whether the data tend to converge on a particular reference tree. Although the reference tree should be the best available estimate of the phylogeny, there is no guarantee that it is the historically accurate phylogeny, regardless of the shape of the curve. The well-known long branch attraction scenario is a theoretical case where parsimony becomes inconsistent and therefore stability and accuracy diverge (Felsenstein, 1987). Under this scenario, as more data are accumulated, they favor the historically inaccurate tree with increasing decisiveness. Under these conditions, a CJF curve will indicate data convergence on historically inaccurate reference tree. Maximum likelihood has been advocated as a solution to the long branch attraction problem under some circumstances (e.g., Felsenstein, 1987; Hendy and Penny, 1989; Huelsenbeck, 1995; Swofford et al., 2001; but see Siddall, 1998; Farris, 1999; Siddall and Whiting, 1999). It may be possible to modify CJF analysis for maximum likelihood, but at a computational cost. CJF analysis for maximum likelihood would have to address both convergence on a topology and convergence on a choice of model because choice of model and parameter values is integral to maximum likelihood analysis and the rate of change in these variables is expected to decrease as more data from a given source are accumulated. As more and more data were excluded, it would become harder to reject simpler models using a likelihood ratio test. It may be reasonable to retain the model structure used for the complete data set for each rarefied data set and only recalculate the parameter values. This approach would probably give approximately the same results for most real data while saving computation time.

Similarly, parsimony analysis based on any form of *a posteriori* weighting should recalculate weights for each rarefied data. Because weighting schemes, too, should change at a decreasing rate with the accumulation of characters, assigning weights to rarefied data sets based on a large number of characters should inflate the curve. Such a curve would be a biased estimate of the effect of

adding new data and would not take into account rate of change for the weighting scheme.

The continuous jackknife method may also give misleading results when characters are nonindependent (i.e., linked by structural or functional constraints; see Pagel, 1994; Wollenberg and Atchley, 2000; O'Keefe and Wagner, 2001) or when gene trees and species trees diverge via paralogy, hybridization, or lateral transfer (e.g., Moritz and Hillis, 1996). Under these conditions, CJF curves may overestimate stability or indicate convergence on a historically inaccurate reference tree. Data sets containing conflicting phylogenetic signal may never converge on any single tree. Again, such problems afflict most, if not all, comparable measures of support. Even when CJF results are misleading in terms of historical accuracy, they should still be able to predict the rate at which the accumulation of new data will change the phylogenetic results.

Some systematists have argued that all relevant and available phylogenetic observations should be considered simultaneously in phylogenetic analysis and that the process of adding data is continuous (Farris, 1983; Eernisse and Kluge, 1993; Kluge and Wolf, 1993; Kluge, 1997a, 1997b; Wenzel, 1997). However, simply maximizing data does not allow scientists to evaluate the adequacy of their sampling efforts. Clearly some phylogenetic questions are more conclusively answered than others. Although some have argued against the relevance of support measures and the idea that stability is a goal of cladistics (Sober, 1988; Kluge, 1989, 1997b), it seems reasonable for systematists to seek some indication of the stability of their phylogenies (e.g., Mueller and Ayala, 1982; Sanderson, 1989, 1995; Siddall, 1995). After all, phylogenetic analyses are fundamental to biology in general through classification systems (Hennig, 1965), comparative biology (Felsenstein, 1985b), conservation (Vane-Wright et al., 1991), and other avenues. Support is an important systematic concept, yet the best way to quantify it remains controversial. Some statistical methods of assessing support are uninterpretable, nonsensical, or misleading, at least under some circumstances (e.g., Goloboff, 1991b; Carpenter, 1992; Källersjö et al., 1992; Farris et al., 1994; Swofford et al., 1996; Carpenter et al., 1998; de Laet and Smets, 1999; Wenzel and Siddall, 1999; Peres-Neto and Marques, 2000; DeBry, 2001). Perhaps support measures can only approximately measure stability. In spite of potential problems, systematists must make informed judgments about their progress toward the resolution of phylogenetic problems. Graphical procedures such as CJF analysis are an underutilized class of methods for making that assessment.

#### ACKNOWLEDGMENTS

Jonathan Coddington indulged me with long conversations in which he freely shared his ideas about the issues discussed here. I thank Gustavo Hormiga, Robert Robbins, John Wenzel, Mike Sharkey, Martín Ramírez, Ingi Agnarsson, Matjaž Kuntner, Fernando Alvarez-Padilla, Chris Simon, Jeffrey Thorne, Hirohisa Kishino, and an anonymous reviewer for their useful comments on earlier drafts of this

manuscript. This article also greatly benefited from discussions and correspondence with Steve Farris, Pablo Goloboff, David Swofford, Diana Lipscomb, Marc Allard, Marymegan Daly, Vickie Funk, Jeff Shultz, Wayne Maddison, Martín Ramírez, Cliff Cunningham, and the Systematics Discussion Group at George Washington University. Acknowledgment does not necessarily imply agreement with the ideas expressed here. Ward Wheeler, Mike Cummings, and Ginny Emerson kindly provided electronic copies of their published alignments. Institutional support was provided by the Department of Systematic Biology—Entomology at the National Museum of Natural History and the Department of Biological Sciences at George Washington University. Financial support was provided by an NSF-PEET grant to Gustavo Hormiga and Jonathan Coddington (DEB-9712353). I also thank Cynthia Zujko-Miller for her help in creating the computer program that made this investigation possible.

#### REFERENCES

- ARCHIE, J. W. 1989a. A randomization test for phylogenetic information in systematic data. *Syst. Zool.* 38:239–252.
- ARCHIE, J. W. 1989b. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Syst. Zool.* 38:253–269.
- BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. *Syst. Zool.* 40:486–493.
- BOND, J., AND B. OPELL. 1998. Testing adaptive radiation and key innovation hypotheses in spiders. *Evolution* 52:403–414.
- 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., C. SIMON, P. K. FLOOK, AND B. MISOF. 2000. Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol. Biol.* 9:565–580.
- 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.
- CARPENTER, J. M. 1992. Random cladistics. *Cladistics* 8:147–153.
- CARPENTER, J. M. 1996. Uninformative bootstrapping. *Cladistics* 12:177–181.
- CARPENTER, J. M., P. A. GOLOBOFF, AND J. S. FARRIS. 1998. PTP is meaningless, T-PTP is contradictory: A reply to Trueman. *Cladistics* 14:105–116.
- CHIPPINDALE, P. T., AND J. J. WIENS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* 43:278–287.
- CODDINGTON, J. A., G. HORMIGA, AND N. SCHARFF. 1997. Giant female or dwarf male spiders? *Nature* 285:687–688.
- COLWELL, R. K., AND J. A. CODDINGTON. 1994. Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Ser. B* 345:101–118.
- CUMMINGS, M. P., S. P. OTTO, AND J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* 12:814–822.
- DAVIS, J. I. 1993. Character removal as a means of assessing stability of clades. *Cladistics* 9:201–210.
- DEBRY, R. W. 2001. Improving interpretation of the decay index for DNA sequence data. *Syst. Biol.* 50:742–752.
- DE LAET, J., AND E. SMETS. 1999. Data decisiveness, missing entries and the DD index. *Cladistics* 15:25–37.
- DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annu. Rev. Ecol. Syst.* 26:657–681.
- DUNLOP, J. A., AND M. WEBSTER. 1999. Fossil evidence, terrestrialization and arachnid phylogeny. *J. Arachnol.* 27:86–93.
- EERNISSE, D. J., AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Mol. Biol. Evol.* 10:1170–1195.
- EMERSON, G. L., C. W. KILPATRICK, B. E. MCNIFF, J. ORRENWALDER, AND M. W. ALLARD. 1999. Phylogenetic relationships of the order Insectivora based on complete 12S rRNA sequences from mitochondria. *Cladistics* 15:221–230.
- FAITH, D. P. 1991. Cladistic permutation tests for monophyly and non-monophyly. *Syst. Zool.* 40:336–375.
- FAITH, D. P., AND P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. *Cladistics* 7:1–28.
- FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *In* *Advances in cladistics*, Volume 2 (N. I. Platnick and V. A. Funk, eds.). Columbia Univ. Press, New York.
- FARRIS, J. S. 1989. The retention index and the rescaled consistency index. *Cladistics* 5:417–419.
- FARRIS, J. S. 1999. Likelihood and inconsistency. *Cladistics* 15:199–204.
- FARRIS, J. S. 2001. Support weighting. *Cladistics* 17:389–394.
- FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12:99–124.
- FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Permutations. *Cladistics* 10:65–76.
- FELSENSTEIN, J. 1985a. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- FELSENSTEIN, J. 1985b. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- FELSENSTEIN, J. 1987. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- GLADSTEIN, D., AND W. C. WHEELER. 1996. OY. Program and documentation. American Museum of Natural History, New York.
- GLADSTEIN, D., AND W. C. WHEELER. 2001. POY. Program and documentation. Available online at <ftp://ftp.amnh.org/pub/molecular/poy/>. American Museum of Natural History, New York.
- GOLOBOFF, P. 1991a. Homoplasy and the choice among cladograms. *Cladistics* 7:215–232.
- GOLOBOFF, P. 1991b. Random data, homoplasy and information. *Cladistics* 7:395–406.
- GOLOBOFF, P. 1993. NONA (a bastard son of Pee-Wee), version 2.0. Program and documentation. Available online at <http://www.cladistics.com>. Published by the author, Tucumán, Argentina.
- GRISWOLD, C. E., J. A. CODDINGTON, G. HORMIGA, AND N. SCHARFF. 1998. Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea). *Zool. J. Linn. Soc.* 123:1–99.
- HENDY, M. D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38:297–309.
- HENNIG, W. 1965. Phylogenetic systematics. *Annu. Rev. Entomol.* 10:97–116.
- HICKSON, R. E., C. SIMON, A. COOPER, G. S. SPICER, J. SULLIVAN, AND C. PENNY. 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* 13:150–169.
- 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.
- HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* 383:130–131.
- HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* 47:3–8.
- HORMIGA, G., W. G. EBERHARD, AND J. A. CODDINGTON. 1995. Web-construction behavior in Australian *Phonognatha* and the phylogeny of nephiline and tetragnathid spiders (Araneae: Tetragnathidae). *Aust. J. Zool.* 43:313–364.
- HORMIGA, G., N. SCHARFF, AND J. A. CODDINGTON. 2000. The phylogenetic basis of sexual size dimorphism in orb-weaving spiders (Araneae, Orbiculariae). *Syst. Biol.* 49:435–462.
- HUELSENBECK, J. P. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* 44:17–48.
- 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.
- JUKES, T. H., AND C. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 *in* *Mammalian protein metabolism* (M. N. Munro, ed.). Academic Press, New York.
- KÄLLERSJÖ, M., J. S. FARRIS, A. G. KLUGE, AND C. BULT. 1992. Skewness and permutation. *Cladistics* 8:275–287.

- 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. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25.
- KLUGE, A. G. 1997a. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. *Zool. Scripta* 26:349–360.
- KLUGE, A. G. 1997b. Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* 13:81–96.
- KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1–32.
- KLUGE, A. G., AND A. J. WOLF. 1993. Cladistics: What's in a word? *Cladistics* 9:183–199.
- LANYON, S. M. 1985. Detecting internal inconsistencies in distance data. *Syst. Zool.* 34:397–403.
- MADDISON, D. R., AND W. P. MADDISON. 2001a. Genesis: Models of character evolution, version 0.96. Available online at <http://mesquiteproject.org>.
- MADDISON, W. P., AND D. R. MADDISON. 2001b. Mesquite: A modular system for evolutionary analysis, version 0.98. Available online at <http://mesquiteproject.org>.
- MILLER, R. G. 1974. The jackknife—A review. *Biometrika* 61:1–15.
- MİYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44:64–76.
- MORITZ, C., AND D. M. HILLIS. 1996. Molecular systematics: Context and controversies. Pages 1–13 in *Molecular systematics*, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.
- 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.
- MUELLER, L. D., AND F. J. AYALA. 1982. Estimation and interpretation of genetic distance in empirical studies. *Genet. Res. Camb.* 40:127–137.
- NIXON, K. C. 1999. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15:407–414.
- NIXON, K. C. 2002. WinClada, version 1.00.04. Program and documentation. Available online at <http://www.cladistics.com>. Published by the author, Ithaca, New York.
- NIXON, K. C., AND J. M. CARPENTER. 1996. On simultaneous analysis. *Cladistics* 12:221–241.
- O'KEEFE, R. F., AND P. J. WAGNER. 2001. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Syst. Biol.* 50:657–675.
- PAGE, R. D. M. 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. *Nucleic Acids Res.* 13:131–144.
- PAGEL, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B* 13:131–144.
- PENNY, D., AND M. HENDY. 1986. Estimating the reliability of evolutionary trees. *Mol. Biol. Evol.* 3:403–417.
- PERES-NETO, P. R., AND F. MARQUES. 2000. When are random data not random, or is the PTP test useful? *Cladistics* 16:420–424.
- PIELOU, C. E. 1966. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* 13:131–144.
- PIELOU, C. E. 1975. *Ecological diversity*. Wiley & Sons, New York.
- SANDERSON, M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* 5:113–129.
- SANDERSON, M. J. 1995. Objections to bootstrapping phylogenies: A critique. *Syst. Biol.* 44:299–320.
- SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43:1781–1795.
- SCHARFF, N., AND J. A. CODDINGTON. 1997. A phylogenetic analysis of the orb-weaving spider family Araneidae (Arachnida, Araneae). *Zool. J. Linn. Soc.* 120:355–434.
- SELLEN, P. A. 1993. Fossil arachnids—Recent advances and future prospects. *Mem. Queensl. Mus.* 33:389–400.
- SHULTZ, J. W. 1990. Evolutionary morphology and phylogeny of Arachnida. *Cladistics* 6:1–38.
- SIDDALL, M. E. 1995. Another monophyly index: Revisiting the jackknife. *Cladistics* 11:33–56.
- SIDDALL, M. E. 1998. Success of parsimony in the four-taxon case: Long branch repulsion by likelihood in the Farris zone. *Cladistics* 14:209–220.
- SIDDALL, M. E., AND M. F. WHITING. 1999. Long-branch abstractions. *Cladistics* 15:9–24.
- SOBER, E. 1988. *Reconstructing the past: Parsimony, evolution and inference*. Massachusetts Institute of Technology Press, Cambridge.
- SULLIVAN, J., AND D. L. SWOFFORD. 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Syst. Biol.* 50:723–729.
- SWOFFORD, D. L., J. L. THORNE, J. FELSENSTEIN, AND B. M. WIEGMANN. 1996. The topology-dependent permutation test for monophyly does not test for monophyly. *Syst. Biol.* 45:575–579.
- SWOFFORD, D. L., P. J. WADDELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, AND J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50:525–539.
- 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.
- TUKEY, J. W. 1977. *Exploratory data analysis*. Addison-Wesley, Reading, Massachusetts.
- VANE-WRIGHT, R. I., C. J. HUMPHRIES, AND P. H. WILLIAMS. 1991. What to protect?—Systematics and the agony of choice. *Biol. Conserv.* 55:235–254.
- WENZEL, J. W. 1997. When is a phylogenetic test good enough? *Mém. Mus. Natl. Hist. Nat. Paris* 173:31–45.
- WENZEL, J. W., AND M. E. SIDDALL. 1999. Noise. *Cladistics* 15:51–64.
- WHEELER, W. C., AND C. Y. HAYASHI. 1998. The phylogeny of the extant chelicerate orders. *Cladistics* 14:173–192.
- WILKINSON, M. 1995. Coping with abundant missing entries in phylogenetic inference using parsimony. *Syst. Biol.* 44:501–514.
- WOLLENBERG, K. R. AND W. R. ATCHLEY. 2000. Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. *Proc. Natl. Acad. Sci. USA* 97:3288–3291.
- ZUJIKO-MILLER, C., AND J. A. MILLER. 2002. PEST: Precision estimated by sampling traits, version 1.0. Program and documentation. Available online at <http://www.gwu.edu/~spiders/pest.htm>. Published by the authors, Washington, D.C.

First submitted 21 April 2002; reviews returned 3 July 2002;

final acceptance 17 October 2002

Associate Editor: Jeffrey Thorne

