15

Congruence Among Data Sets: A Bayesian Approach

WARD C. WHEELER

When several independently derived data sets agree, producing the same cladogram, we have high confidence in the result. Such complete agreement, however, is rare and becoming rarer with the multiplication of molecular data sets.

Extreme disagreement has arisen between morphological and molecular studies. Zimmer et al. (1989) have produced results on monocot origins that contradict the morphological results of Donoghue and Doyle (1989). Field et al. (1988) have suggested relationships within and among invertebrate groups that counter the great majority of work on invertebrate systematics. Mammal studies have produced further disagreement (Goodman et al., 1985; Wyss et al., 1987), and studies of bird phylogeny (Cracraft, 1974; Sibley and Ahlquist, 1981) have had their differences. While these conflicts are not limited to comparisons between molecules and morphology, herein lie some of the most prominent cases. I do not mean to overstate the extent of the disagreement or of the acrimony accompanying it. In fact, congruence among data sets may well be more common than incongruence. Nonetheless, in these works, there are important differences between hypotheses of relationships which require explanation.

When molecular data disagree with hypotheses based on morphology, there are two common responses. Either one of the studies is dismissed or some consensus is attempted. When consensus prevails, a consensus "tree" is most often produced. There are, however, shortcomings to this approach.

The most frequently used consensus procedures are Adams (1972) and strict (Rohlf, 1982) consensus. Neither of these methods is entirely satisfying. The Adams procedure may generate groupings of taxa that are found nowhere in the initial cladograms, while the strict method is so ruthlessly conservative that all resolution may be lost by a shift in the position of a single taxon. In the strict consensus, only groups present in all input trees

are included. Lately, a modified procedure has become popular (Margush and McMorris, 1981). The modified method includes groups that appear in some fraction (greater than one half) of the cladograms. In this way, "majority rule" consensus or percentage consensus trees are produced.

There is, however, a problem intrinsic to consensus analysis. These procedures tend to combine the weaknesses of the data sets; disagreements based on few data are accorded equal weight with agreements based on a great deal of information, since only topologies are compared (Miyamoto, 1985). A desirable method would combine the strengths of the data instead, making allowances for the degree to which separate data sets support different nodes. Here, I propose a method that attempts to do this using Bayesian decision theory. The Bayesian approach draws on a philosophy and an analytical procedure that directly apply to the problem of congruence in phylogenetic analysis.

THE MONTY HALL PROBLEM

Before discussing the particulars of Bayesian decision theory, I would like to start with an everyday example of the purpose and power of this type of analysis, known as the "Monty Hall Problem." Imagine yourself on the television game show "Let's Make a Deal." Monty Hall, the host, presents you with three doors. Concealed behind one of these is a fabulous prize, while the others hide objects considerably less attractive. You are invited to choose a door, guessing which one hides the best prize. Suppose you select door number 1. At this point, Monty Hall opens door number 3, revealing an undesirable bauble (by the dictates of the game, Hall never reveals the best prize). You are now faced with a dilemma: do you stay with your original pick or do you go for what's behind door number 2? The proper Bayesian would switch, and double her chances of winning.

This may, at first, seem counterintuitive. After all, the prize is equally likely to be behind any of the doors. How are your odds improved by switching? The trick comes from the use of additional information. Initially, your pick has a one in three chance of being correct. Hence, the probability that the prize is behind one of the other doors is the complement of this—two-thirds. When Monty Hall reveals that there is nothing worthwhile behind one of the two doors you did not pick, the probability of that door being the one with the desirable prize collapses to 0. The total probability, however, that the two unchosen doors contain the best prize is unchanged; and it remains two-thirds. At this point, there is a one-third chance you have picked the correct door and a two-thirds chance that the prize is behind the other door. Of course you should switch.

Put another way, if your first choice, door number 1, is correct and you stand pat, you win. But if the desirable prize lies behind door number 2 or number 3 you must switch to win. Hence, by switching you win two out

of three times. It is true that your chances of finding the prize behind door number 1 do not change when Monty Hall discloses the prize behind door number 3. But your chances of finding the one prize behind door number 2 increase dramatically.

The Bayesian decision to switch is mandated by two conditions: (1) the prior information that no door is more likely than any other to contain the desirable prize; and (2) the information gathered by observing the contents of one of the doors not chosen. These are the basic components of a Bayesian analysis, the information prior to the observation and the observations themselves. This type of procedure is the only one to include both kinds of information in the decision process.

The point I will make here is that in the use of morphological and molecular data to determine the credibility of a cladogram, morphology can provide the prior evidence to interpret molecular data.

DEFINITIONS OF PROBABILITY

Since this type of analysis deals with the probabilities of phylogenetic schemes, it is necessary to first define what is meant by the probability of a cladogram. There are three ways of defining probability: logical, empirical, and subjective (Hartigan, 1983). A logical probability is a rational degree of belief in a hypothesis in light of certain evidence, whereas an empirical probability is a statement of fact about the world. The third type of definition, subjective, is similar to the logical in that it represents a degree of belief, but it is an *individual* degree of belief. Two people may assign different probabilities based on the same evidence.

An example of these different concepts of probability can be found in betting ratios. Hartigan (1983) makes the point that a logical bettor (Bayesian) wagers what they ought to, while the empiricist will bet what is profitable in the long run, and the subjectivist will bet what they are willing. In the procedure described here, I will use the logical definition of probability. The probability of a particular cladogram, then, is our degree of belief in that cladogram, in light of evidence from molecular and morphological characters.

BAYESIAN INFERENCE AND DECISION THEORY

Just as there are three types of probability definitions, there are, broadly speaking, three types of inference: "classical," likelihood, and Bayesian. I am using classical inference to refer to most of the standard methods of statistical inference, those involving the calculation of p values and confidence intervals to approximate final probabilities. Classical methods have at least two shortcomings: first, they assume that the uncollected data can affect the distribution of the estimates; and second, they base their conclusions in part on the manner in which the data were collected. Thus, if the collection procedure is not known, or is modified during the gathering of observations, p values have no meaning, and classical methods are invalidated. Likelihood methods, in contrast to classical methods, rely solely on the observed data; their conclusions are not influenced by the way in which the data are gathered. These methods, however, do not make use of all relevant information. Specifically, they ignore information about prior distributions. This omission is quite deliberate. Due to the philosophical difficulties in determining priors—they may be entirely subjective—likelihood approaches pointedly avoid their use, or at least assume that all priors are equal and therefore irrelevant. Bayesian procedures, not bothered by these restrictions, extend likelihood through the inclusion of prior information.

All Bayesian procedures rely on an elementary relation of conditional probability (Bayes, 1763):

$$\mathrm{pr}(\Theta = \theta_i | \mathbf{D}) = \frac{\mathrm{pr}(\mathbf{D}|\theta_i)\mathrm{pr}(\theta_i)}{\Sigma_k[\mathrm{pr}(\mathbf{D}|\theta_k)\mathrm{pr}(\theta_k)]} \quad (1)$$

The "final" or posterior probability that the parameter Θ has the value θ_i given the data $\mathbf{D}$ is constructed by multiplying the probability of the data given θ_i, $\mathrm{pr}(\mathbf{D}|\theta_i)$, with its prior probability, $\mathrm{pr}(\theta_i)$, and normalized by the sum of these terms for all values of θ (θ_k).

If the priors are known or can be reasonably calculated, all statisticians would agree that Bayesian methods should be used to determine the best estimate of Θ. Some priors seem inherently reasonable. An example would be the proportion of defective items on an assembly line. In this case, the performance of the line in the past gives a prior probability of the future production of defective items. Yet, when priors are less clearly linked to observation, there is disagreement about their use. In fact, those who hold to likelihood argue that in all but the most trivial cases, prior probabilities cannot be calculated; thus only likelihoods (or their ratios) can be used.

THE SENSITIVITY OF THE RESULT TO THE PRIOR

One way to measure the effect of prior probability is through the use of simplex space. Take, for example, a parameter Θ, which may have three values, θ_1, θ_2, or θ_3. The sensitivity of the final probability to the prior probabilities can be graphed (Fig. 15-1). The space is divided into three regions. These contain the prior probabilities that determine the choice of θ_1, θ_2, or θ_3 as the estimate of Θ. The values of the priors are plotted ($\mathrm{pr}(\theta_1) + \mathrm{pr}(\theta_2) + \mathrm{pr}(\theta_3) = 1$) and their propinquity to the decision lines noted. If this point is very close to one or several of these lines, small variations in the priors can affect the estimate of Θ, undermining our faith in the result. If, however, the priors map to areas well within a decision space,

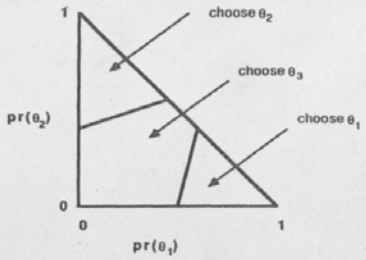

Figure 15-1 Simplex space of a simple inference problem showing the areas in which estimates 1, 2, and 3 are chosen.

we have greater confidence in the conclusions. A more precise measure of the robustness of results is offered by decision theory.

DECISION THEORY

A Bayesian decision is one which minimizes risk. In order to make this type of decision, an appropriate loss function must be specified for all possible values of θ. This loss function, coupled with prior probability and data probability values, allows us to determine the cost or risk of any decision. By definition, a decision is to accept some value θ_i as the estimate of a parameter Θ, or to embark upon a course of action θ_i out of all possible actions.

Unfortunately, the losses must be specified for all possible correct and incorrect decisions. This task may be overwhelming or impossible, given the complexity of the options and their costs. If, however, we are willing to accept that all incorrect hypotheses are equally undesirable and all correct hypotheses are equally desirable in that they are correct, a simple loss function can be used.

In this type of function, the cost of a correct decision is 0 while the cost of an incorrect choice is one (Fig. 15-2). The risk of a decision, then, is the sum of the cost of all the decisions:

$$R_i = \sum_j C_{ij}\mathrm{pr}(\Theta = \theta_i|\mathbf{D}). \qquad (2)$$

The risk of the decision i (R_i) is equal to the sum over all j decisions of the product of the cost of that decision (C_{ij}—if the parameter has the value i) and its final probability. With the simple cost matrix suggested above (Fig. 15-2), the decision risk is merely the complement of its final probability. Since

$$C_{ij} = 1 \text{ if } i \neq j$$

$$C_{ij} = 0 \text{ if } i = j,$$

the risk reduces to:

$$R_i = 1 - \mathrm{pr}(\Theta = \theta_i|\mathbf{D}) \qquad (3a)$$

		Decision						
		1	2	3	-	n-2	n-1	n
	1	0	1	1	-	1	1	1
	2	1	0	1	-	1	1	1
	3	1	1	0	-	1	1	1
True Value	.	.	.	.		.	.	.
	n-2	1	1	1	-	0	1	1
	n-1	1	1	1	-	1	0	1
	n	1	1	1	-	1	1	0

Figure 15-2 A simple cost matrix used in phylogenetic decisions with "n" decisions possible.

for a single cladogram or:

$$R_i = 1 - \sum_k \mathrm{pr}(\Theta = \theta_k|\mathbf{D}) \qquad (3b)$$

for several cladograms $\{k\}$ in i. By these manipulations, the risk of a decision with a final probability of 0.9 would be 0.1. The minimum risk occurs when the probability is maximized.

I propose that this notion of risk be used to evaluate the support for cladograms. Possibly no single cladogram will stand out as having acceptably low risk; in these cases, several cladograms should be considered until their collective risk is sufficiently low. By analogy with p values, we might choose the "risk" of type-I error to be no greater than 5%. Here, we would accept the most probable cladograms until 95% of the risk had been removed.

THE PROCEDURE

Determination of Prior Probabilities $\{\mathrm{pr}(\theta_i)\}$

The first step in this type of analysis is to determine the prior probabilities for all of the possible values of θ. Usually, prior probabilities are calculated using known distributions of continuous or discrete variables. In this case, however, the prior probability of a cladogram in a molecular analysis is its ability to explain morphological data, which is an entirely discrete value.

There are many measures of the relative abilities of cladograms to explain data, such as the length (number of evolutionary steps) of the scheme or its consistency index (Kluge and Farris, 1969). Here, a length difference of one step (or postulated event) will be considered equal to a difference in probability of a factor of e, the base of natural logarithms. (Since these probabilities will be normalized, they need not sum to unity.) This factor of probability reflects the relative degree of belief in the different hypotheses before the molecular observations are made.

This assessment of probability may seem arbitrary at first. However, it can be justified by the link between natural logarithms and phylogenetic character weights (Farris, 1977). Of course, an individual investigator may use different methods to determine prior probabilities based on their con-

fidence in that topology. In an unweighted parsimony analysis, all changes increment the cladogram cost by one step; similarly here, all changes are considered to be equally probable until there is evidence to the contrary.

Thus the prior probability of cladogram i with a length L_i from some previous analysis is:

$$\text{pr}(\theta_i) = e^{(-L_i)}, \tag{4}$$

or to sum to unity:

$$\text{pr}(\theta_i) = e^{(-L_i)} / \sum_k e^{(-L_k)}. \tag{5}$$

Determination of Data Probability $\{\text{pr}(\Theta = \theta_i | D)\}$

The second phase of the procedure involves determining the probability of the data given a certain θ_i, or topology. Any method that yields this value can be used. There are three methods, however, which seem especially appropriate: unweighted or maximum parsimony, weighted parsimony, and maximum likelihood.

The first of these may seem, initially, to be unsuited to statistical interpretation. As with the establishment of priors, a factor of e change in probability can be assigned to each event. In this way, all changes are treated equally, and a quantitative degree of belief is determined for each topology. As with Keynes' (1921) argument about the frequency of colored balls in an urn, in the absence of knowledge to the contrary, we must assume all events are equally likely. Total symmetry is a very simple model of character change, but it offers a starting point for examining the data. One effect of using the same probability factor in determining priors and data probabilities is that the result is the same as if all the data had been collected at one time and pooled. The absolute value of the natural logarithm of the final probabilities will be the lengths of cladograms constructed from these pooled data. The most probable (least risk) cladogram will be the most-parsimonious.

From this entirely symmetrical model of character transformation, the next step is to vary the weights assigned to different types of transformation. Just as these models of character transformation are used in weighted parsimony analyses, they can be used to assign data probabilities. Along these lines, Farris (1977) and DeBry and Slade (1985) have offered a Dollo-type method based on probabilistic thinking. In the analysis of nucleic acid sequences, many have suggested that transitions should be weighted differently from transversions (Brown et al., 1982). Specific cost ratios have been proposed, with obvious probabilistic interpretations. Elsewhere, I have put forward a combinatorial weight scheme for nucleic acids that yields weights for different types of character transformation (Wheeler, 1990). In each of these methods, only the different types of change receive varying probabilities. No assessments of the probability of change itself occurring are attempted (in a character transformation matrix all $a_{ii} = 0$). Instead, these procedures rely on explicit minimization of "cost" values.

$$\begin{array}{c} \\ \text{A} \\ \text{C} \\ \text{G} \\ \text{T} \end{array} \begin{array}{c} \begin{array}{cccc} \text{A} & \text{C} & \text{G} & \text{T} \end{array} \\ \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix} \end{array}$$

Figure 15-3 Markov type character transformation model showing the probability of change from one character state to another.

Likelihood methods differ, in that they seek to incorporate the rate of evolution. Not only the possible avenues of change, but the probability of change is included in the analysis. Smouse and Li (1987) use this approach in their examination of restriction fragment character change (and even discuss prior probabilities from other studies). Markov character transformation models (Fig. 15-3) chart rates of evolution as diagonal values. Since parsimony models lack these values, they offer no information on the probability of change.

With the prior probabilities and the data probabilities in hand, the final probability of each cladogram is calculated by equation (1).

Determination of Utility and Risk

The loss function suggested above and the final probabilities are used to determine the risks of various single and compound decisions [equation (3)]. A single cladogram, or several, may be accepted until risk is sufficiently minimized.

An Example

In order to demonstrate this congruence procedure, the case of the monophyly of the Eumetabola will be examined in the light of morphological (Hennig, 1969, 1981; Kristensen, 1975, 1981; Boudreaux, 1979) and molecular (Wheeler, 1989) evidence. There are 15 dichotomous unrooted arrangements of the five insect taxa (Paleoptera, Orthoptera, Hemiptera, Diptera, and Coleoptera, as shown in Fig. 15-4). In the morphological data set, there were five presence/absence characters: wing flexion (neoptery), the jugal bar, larval stemmata, holometaboly, and larval ocelli. The number of evolutionary events required by each of the possible arrangements is shown in Table 15-1. The sequence data of Wheeler (1989) were applied to these same cladograms; their lengths are also shown in Table 15-1, along with the weighted parsimony values. The procedure used to determine the weighted cladogram lengths is described in detail in Wheeler (1990). The likelihoods, or more accurately, their absolute values (Felsenstein, 1978, 1979), are also shown in this table.

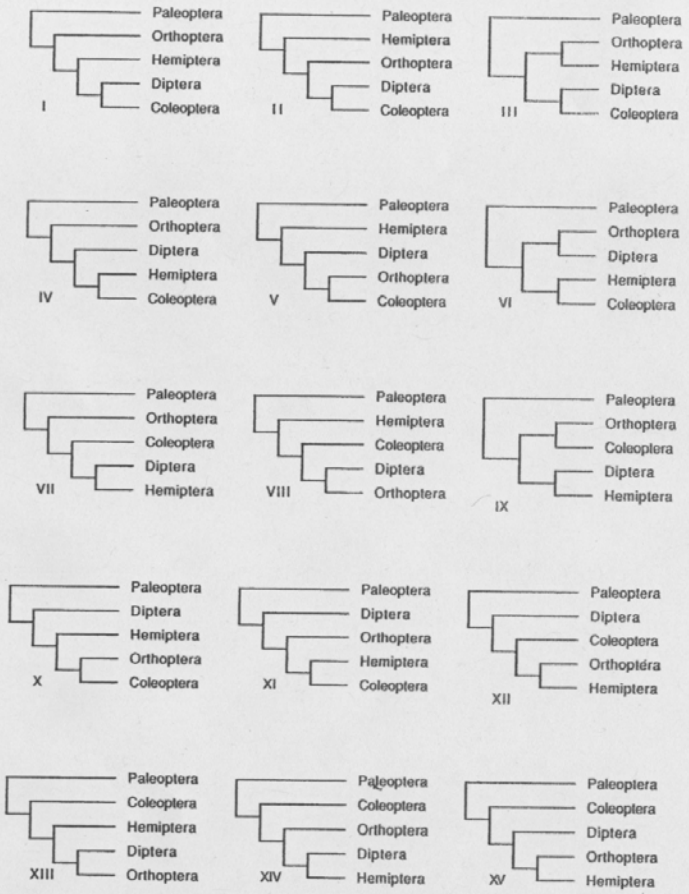

Figure 15-4 The 15 possible topologies for the Coleoptera, Diptera, Hemiptera, Orthoptera, and Paleoptea, with Paleoptera as the outgroup.

The final probabilities determined from the three types of cladogram costs and the morphologically based priors are shown in Table 15-2. Table 15-3 gives the topology decisions and associated risks.

The three types of data probabilities, maximum parsimony, weighted parsimony, and likelihood, agree for the most part. Topology III is the most favored (or at least one of the most favored hypotheses) in each of the analyses. However, the likelihood method includes topology XV and

Table 15-1 Cladogram Length

		Sequence Data		
Cladogram	Morphology	Unweighted Parsimony	Weighted Parsimony	Likelihood
I	5	133	317	494
II	6	137	327	494
III	6	132	314	489
IV	7	135	321	494
V	8	140	332	494
VI	8	136	322	494
VII	7	137	327	494
VIII	8	139	327	494
IX	8	140	334	494
X	8	137	328	493
XI	8	134	321	493
XII	8	132	316	489
XIII	8	136	322	493
XIV	8	138	329	493
XV	8	133	318	490

Table 15-2 Final Probabilities

Cladogram	Unweighted Parsimony	Weighted Parsimony	Likelihood
I	4.5×10^{-1}	1.2×10^{-1}	1.5×10^{-2}
II	3.0×10^{-3}	2.0×10^{-6}	5.5×10^{-3}
III	4.5×10^{-1}	8.7×10^{-1}	8.1×10^{-1}
IV	8.2×10^{-3}	2.9×10^{-4}	2.0×10^{-3}
V	2.0×10^{-5}	1.8×10^{-9}	7.4×10^{-4}
VI	1.1×10^{-3}	3.9×10^{-5}	7.4×10^{-4}
VII	1.1×10^{-3}	7.2×10^{-7}	2.0×10^{-3}
VIII	5.5×10^{-5}	2.6×10^{-7}	7.4×10^{-4}
IX	2.0×10^{-5}	2.4×10^{-10}	7.4×10^{-4}
X	4.1×10^{-4}	9.7×10^{-8}	2.0×10^{-3}
XI	8.2×10^{-3}	1.1×10^{-4}	2.0×10^{-3}
XII	6.1×10^{-2}	1.6×10^{-2}	1.1×10^{-1}
XIII	1.1×10^{-3}	3.9×10^{-5}	2.0×10^{-3}
XIV	1.5×10^{-4}	3.6×10^{-8}	2.0×10^{-3}
XV	2.2×10^{-2}	2.1×10^{-3}	4.1×10^{-2}

Table 15-3 Topology Decisions

	Cladograms		
Risk Level	Unweighted Parsimony	Weighted Parsimony	Likelihood
20%	I, III	III	III
10%	I, III	I, III	III, XII
5%	I, III, XII	I, III	III, XII, XV

excludes topology I in the 5% risk decision, while the two parsimony-based methods include I at the expense of XV.

CONCLUSIONS

At present there are two commonly used congruence procedures: global parsimony and consensus (Miyamoto, 1985; Kluge, 1989). In a global parsimony analysis, all of the characters are lumped together, and the most-parsimonious solution for the combined data is chosen as the most-favored hypothesis. Consensus procedures, on the other hand, treat the data sets distinctly; individual characters and their combinations do not play a direct role in the topology decision made from comparing the data sets. In this respect, consensus procedures treat characters as if they were members of noncomparable categories; hence, only topologies are compared. The method presented here attempts to extract the best qualities of both approaches through Bayesian analysis.

If the simple *e* factor transformation is used for both the prior distribution and the determination of the data probability, the result will be the most-parsimonious solution for the combined data. Since all characters are weighted equally, it does not matter from which analysis (morphological or molecular) they were derived. If, however, the investigator believes that character classes exist, or feels that some distinction can be made between the types of data used, this method allows him to separately determine the probabilities that influence the final decision.

The procedure draws its strength from this flexibility. These probabilities offer the investigator a quantitative assessment of the degree of belief. If several analyses favor a certain hypothesis, but each only marginally, the final probability of that hypothesis should reflect the increased confidence which comes from this corroboration.

The risk inherent in the decision to accept or reject topologies gives a measure of the level of agreement between data sets. If there is little discrimination, or the various observations differ greatly, no hypothesis will have sufficiently low risk to be accepted by itself. In these cases, several or many topologies will have to be considered to reduce the risk to a satisfactory level.

Congruence is the final test of any hypothesis. The type of procedure presented here gives a logical means of assessing the agreement and disagreement between different sets of data that bear on the same phylogenetic question.

ACKNOWLEDGMENTS

I would like to acknowledge Kirk Fitzhugh, Paul Vrana, and especially Elise Broach for reading drafts of this manuscript. I would also like to thank two anonymous reviewers. I thank Christopher Urheim for making clear to me the subtleties of the "Monty Hall Problem." This research was supported by the Alfred P. Sloan Foundation.

REFERENCES

Adams, E. N. III. (1972) Consensus techniques and the comparison of phylogenetic trees. *Syst. Zool.* **21**: 390–397.

Bayes, T. (1763) An essay towards solving a problem in the doctrine of chances. *Philos. Trans. R. Soc.* **53**: 370–418.

Boudreaux, H. B. (1979) *Arthropod Phylogeny with Special Reference to Insects.* John Wiley & Sons, New York.

Brown, W. M., E. M. Praeger, A. Wang, and A. C. Wilson . (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Mol. Evol.* **18**: 225–233.

Cracraft, J. (1974) Phylogeny and evolution of the ratite birds. *Ibis* **116**: 494–521.

DeBry, R. W., and N. A. Slade. (1985) Cladistic analysis of restriction endonuclease cleavage maps with a maximum likelihood framework. *Syst. Zool.* **34**: 21–34.

Donoghue, M. J., and J. A. Doyle. (1989) Phylogenetic studies of seed plants and angiosperms based on morphological characters. Pp. 181–194 in *The Hierarchy of Life. Molecules and Morphology in Phylogenetic Analysis* (B. Fernholm, K. Bremer, and H. Jörnvall, eds.). Elsevier Press Science Publishers B.V., Amsterdam.

Farris, J. S. (1977) Phylogenetic analysis under Dollo's law. *Syst. Zool.* **26**: 77–88.

Felsenstein, J. (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401–410.

Felsenstein, J. (1979) Alternate methods of phylogenetic inference and their interrelationship. *Syst. Zool.* **28**: 49–62.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace, and R. A. Raff. (1988) Molecular phylogeny of the animal kingdom. *Science* **239**: 748–753.

Goodman, M., J. Czelusniak, and J. E. Beeber. (1985) Phylogeny of primates and other eutherian orders: a cladistic analysis using amino acid and nucleotide sequence data. *Cladistics* **1**: 171–185.

Hartigan, J. A. (1983) *Bayes Theory*. Springer-Verlag, New York.

Hennig, W. (1969) Die stammesgeschichte der Insekten. *Seckenberg-Büchern*. **49**: 1–436.

Hennig, W. (1981) *Insect Phylogeny*. John Wiley & Sons, New York.

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., and J. S. Farris. (1969) Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**: 1–32.

Kristensen, N. P. (1975) The phylogeny of hexapod "orders": a critical review of recent accounts. *Z. Zool. Syst. Evol. Forsch.* **13**: 1–44.

Kristensen, N. P. (1981) Phylogeny of insect orders. *Ann. Rev. Entomol.* **26**: 135–157.

Margush, T. and F. R. McMorris. (1981) Consensus *n*-trees. *Bull. Math. Biol.* **43**: 239–244.

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

Rohlf, F. J. (1982) Consensus indices for comparing classifications. *Math. Biosci.* **59**: 131–144.

Sibley, C., and J. E. Ahlquist. (1981) The phylogeny and relationships of the ratite birds as indicated by DNA-DNA hybridization. Pp. 301–335 in *Evolution Today* (G. G. E. Scudder and J. L. Reveal, eds.) Carnegie-Mellon University, Pittsburgh.

Smouse, P. E., and W.-H. Li. (1987) Likelihood analysis of mitochondrial restriction-cleavage patterns for the human-chimpanzee-gorilla trichotomy. *Evolution* **41**: 1162–1176.

Wheeler, W. C. (1989) Evolution and systematics of insect rDNA. Pp. 307–321 in *The Hierarchy of Life. Molecules and Morphology in Phylogenetic Analysis* (B. Fernholm, K. Bremer, and H. Jörnvall, eds.). Elsevier Science Publishers B. V., Amsterdam.

Wheeler, W. C. (1990) Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. *Cladistics* **6**: 269–275.

Wyss, A. R., M. J. Novacek, and M. C. McKenna. (1987) Amino acid sequence versus morphological data and the interordinal relationships of mammals. *Mol. Biol. Evol.* **4**: 99–116.

Zimmer, E. A., R. K. Hamby, M. L. Arnold, D. A. Leblanc, and E. L. Theriot. (1989) Ribosomal phylogeny and flowering plant evolution. Pp. 205–214 in *The Hierarchy of Life. Molecules and Morphology in Phylogenetic Analysis* (B. Fernholm, K. Bremer, and H. Jörnvall, eds.) Elsevier Science Publishers B. V., Amsterdam.

Index