J Mol Evol (1983) 19:244–254

Journal of
Molecular Evolution
© Springer-Verlag 1983

An Empirical Evaluation of Qualitative Hennigian Analyses of Protein Electrophoretic Data

John C. Patton[1] and John C. Avise[2]

1 Department of Biology, Baylor University, Waco, Texas 76798, USA
2 Department of Molecular and Population Genetics, University of Georgia, Athens, Georgia 30602, USA

Summary. In an empirical evaluation of a qualitative approach to construction of phylogenetic trees from protein-electrophoretic data, we have employed Hennigian cladistic principles to generate molecular trees for waterfowl, rodents, bats, and other phylads. This procedure of tree construction is described in detail. Branching structures of molecular trees produced by three different algorithms were compared against those of "model" classifications previously proposed by other systematists. In each case, the qualitative cladistic trees provided fits to model phylogenies which were strong and as good or better than those resulting from phenetic-clustering or distance-Wagner trees based on manipulation of quantitative values in matrices of genetic distance.

The qualitative Hennigian approach has several pragmatic (as well as theoretical) advantages for analyzing routine sets of electrophoretic data: (1) the analyses are simple and can be performed by hand; (2) they provide the researcher with a strong "feel" for the data; (3) additional data (from new loci or species) can readily be added to the tree without need to recalculate distance matrices; and (4) the qualitative output of the analyses explicitly defines character states along all branches of the tree, and hence affords a high degree of testability. However, these advantages are counterbalanced by a number of serious disadvantages which will likely limit the general applicability of this qualitative approach. These drawbacks are also discussed in detail.

For a deeper appreciation of electrophoretic-based protein phylogenies, it is suggested that both quantitative phenetic and qualitative cladistic analyses be employed when posssible, and that results of the two approaches be contrasted.

Offprint requests to: J.C. Avise

Key words: Hennigian Cladistics – Distance Wagner procedure – UPGMA clustering – Protein electrophoresis – Phylogeny – Systematics

Introduction

Biological macromolecules contain within their structures information about evolutionary relationships of organisms. Because a variety of molecular techniques are now widely employed to compare proteins and DNA's from different species, there has been considerable interest in developing algorithms for converting these genetic data into estimates of phylogeny. Most often, such algorithms manipulate quantitative values in a summary matrix of genetic similarities or distances between populations or species. For molecular data, the most widely-used methods of phylogenetic tree construction have been those of Fitch-Margoliash (1967; the F-M procedure) and Farris (1972, 1974; the distance-Wagner procedure). Commonly employed phenetic-clustering procedures, and in particular the "unweighted pair-group method analysis" (UPGMA), have been discussed by Sneath and Sokal (1973). Goodnss-of-fit between output distances from the tree, and input distances from the matrix, is a criterion commonly employed to choose the "best" tree (Prager and Wilson 1978).

Some molecular techniques, such as those involving immunological comparisons, provide raw data only in the form of quantitative distance values (i.e., immunological distance units – Gorman et al. 1971; Wilson et al. 1977). However, other molecular techniques, such as conventional protein electrophoresis, provide raw data in the form of distributions of qualitative character states – electromorphs in this case. During the evolutionary process, it is changes in the genetic character

states themselves that underlie any increase or decrease in overall genetic distance. Thus, there is a potential loss of information in first generating a distance matrix before constructing a phylogenetic tree. Furthermore, character state evolution can include convergences, reversals, and parallel retentions of ancestral states. Although these general phenomena are relatively easily calculated where a cladogeny is postulated, the individual characters involved are not easily elucidated solely by reference to a distance matrix. Thus, for many purposes, a more appealing phylogenetic approach would focus upon analyses of the character states themselves.

Hennig (1966) has provided a conceptual framework (now called Hennigian cladistics) for analyzing character state evolution to generate phylogenetic trees. In this approach, ancestral (plesiomorphic) character states are distinguished from derived (apomorphic) states. Since primitive characters may be held in common by distantly related extant species, common ancestries (clades) within the group under consideration are defined solely by possession of shared-derived (synapomorphic) characters. Hennig (1966) suggested several criteria for determining ancestral character states, such as their presence in fossils. Of these criteria, the "sister-group" approach appears most applicable to molecular or other data available only from extant species. By this definiton, a character state is considered plesiomorphic if it is shared with members of a group related to, but phylogenetically distinct from, the assemblage under study.

Patton et al. (1981) used principles of Hennigian cladistics to analyze qualitative, uncoded electrophoretic data from New World rodents, Cricetidae. The cladistic tree provided a slightly better fit to a proposed "model" phylogeny for these taxa than did standard phenetic analyses of the same data. We have since applied Hennigian priniciples to several other electrophoretic data sets, with varying degrees of success. The purposes of this paper are to (1) review this method of Hennigian tree generation, by reference to a particular set of electrophoretic data on waterfowl (Anatidae); (2) compare results of Hennigian cladistic analysis of several data sets with two of the most commonly employed quantitative analyses (UPGMA (Sneath and Sokal 1973) and distance-Wagner (Farris 1972)) of these same data; (3) provide a candid appraisal of the strengths and weaknesses of Hennigian cladistics to analyze qualitative electrophoretic data.

Methods

To exemplify our method of application of Hennigian principles, we will review in detail the analysis of data composed of observed electromorph frequencies at 17–19 genetic loci in 26 species of North American waterfowl, Anatidae (Patton and Avise, submitted). These species are listed and numbered in Table 1, and the electromorphs observed at polymorphic loci are presented in Table 2.

The well-delimited, presumably monophyletic family Anatidae is divided by almost all authorities into two major subfamilies, the Anserinae (geese and swans) and the Anatinae (most ducks). The subfamilies probably represent related but independent evolutionary lineages, dating from the early Cenozoic when fossil remains for both groups first appear (Howard 1964; Romer 1966). If the two subfamilies do indeed represent sister-groups, then by the above definition, an electromorph shared by any member of Anatinae with any member of Anserinae is plesiomorphic to the original Anatinae-Anserinae split. This is the only assumption about phylogenetic relationships of waterfowl underlying the following cladistic analysis.

We have found it convenient to group loci into the following categories for purposes of tree generation. Analysis of these groupings provides a general protocol for manipulating qualitative data bases.

Group I, Single Electromorph Present. Six loci (*CK-3, GOT-2, IDH-2, IPO-1, LDH-2,* and *PEPT-1*) appeared monomorphic for the same allelic product throughout Anatidae. Hence they provide no useful information about cladistic relationships within the family. Electromorphs at these loci were apparently present in basal anatid stock (Fig. 1, I).

Group II, Plesiomorph Determined, Other Electromorphs Define Autapomorphic Character States. At each of four loci (A-D, Table 2), the plesiomorph could be determined, and all other electromorphs yield character states unique (autapomorphic) to a single species. These loci are also uninformative about cladistic affinities of species within Anatidae (Fig. 1, II).

Group III, Plesiomorph Determined, Distinct Clades Defined By Synapomorphs. Several loci contribute to clade definition by the Hennigian criterion that they exhibit derived electromorphs shared by various taxa. In many cases, electromorphs define clades whose boundaries are not strictly inconsistent with information from any other asssayed locus. Only these electromorphs are included within Group III. For example, because Pgm^{50} is shared by some members of Anatinae and Anserinae, it is a plesiomorph for Anatidae; the derived electromorph Pgm^{100} was observed in all species numbered 6–15 (genus *Anas*, Table 1), and no other locus exhibits synapomorphs linking any *Anas* species with any non-*Anas* waterfowl. In addition, three other autapomorphic character states were generated by alleles observed at PGM. The addition of loci E-G defines the growing Hennigian tree pictured in Fig. 1, III.

Similarly, Pgd^{100} defines the plesiomorphic state at PGD for Anatidae (Table 2), so Pgd^{200} becomes a derived electromorph which defines a clade composed of all five assayed members of Anserinae. Gpd^{75} constitutes another anatid plesiomorph (Table 2), and the derived electromorph Gpd^{100} defines a clade composed of species numbered 6–23. Finally, $Idh\text{-}1^{90}$ appears to be ancestral to Anatidae, and the derived allele $Idh\text{-}1^{120}$ delineates a clade composed of species 2–5. The addition of these synapomorphic characters, plus the plesiomorphic and autapomorphic characters at loci H-L, results in the tree pictured in Fig. 1, IV.

To this point in the data analysis, tree generation has been objective, governed solely by strict application of Hennigian thought. However, some ambiguity accompanies the interpretation of the next cohort of electromorphs.

Group IV, Plesiomorph Determined, "Ambiguous" Clades Defined by Synapomorphs. At loci H-L, a few electromorphs appear to define contradictory clades. There are four such instances in the waterfowl data, and differing interpretations of how to resolve these character conflicts will inevitably result in the formulation of different Hennigian trees. Thus it could be that one or more electromorphs were incorrectly hypothesized as defining a derived character state, or it could be that an electromorph has

arisen more than once, independently by convergence. As Wiley (1981) states: "conflicting characters must be explained as plesiomorphies or nonhomologies".

For example, rare but hypothesized synapomorphs were shared by *Bucephala albeola* – *Aythya americana* (Pgd^{120}), *B. albeola* – *A. affinis* (Gpd^{70}), and *B. albeola* – *B. clangula* ($Idh\text{-}1^{180}$). If we follow taxonomic guidelines and link to two *Bucephala* species by the hypothesized synapomorphic character state $Idh\text{-}1^{180}$, then Pgd^{120} and Gpd^{70} may have been present in ancestral *Aythya* – *Bucephala* stock. This is the interpretation presented in Fig. 2. Other interpretations would suggest a polyphyletic origin for *Aythya* with respect to *Bucephala*, or else the convergent evolution of Pgd^{120} and Gpd^{70} by *Bucephala* and *Aythya*.

A second ambiguity concerns *Anas cyanoptera* (species #15), which is linked with *A. discors* and *A. clypeata* by the proposed synapomorphic character state $Ldh\text{-}1^{95}$ and with *A. acuta* by the synapomorphic character state Pgd^{110}. Since $Ldh\text{-}1^{95}$ was monomorphic in three *Anas* species and Pgd^{110} appeared only as a rare variant, we have chosen to picture the former clade in Fig. 2. A third and minor area of ambiguity concerns the placement of *Anas acuta*, which is linked to *A. cyanoptera* by Pgd^{110} and to *A. platyrhynchos* by Pgi^{200}. Arbitrarily, we have chosen the latter interpretation in Fig. 2, which thus places Pgd^{110} as ancestral to most *Anas* species. Of course it is also quite possible that Pgd^{100} (and/or Pgi^{200}) arose independently in the species in which it was observed.

Finally, in 16 of the 21 species of Anatinae assayed, both $Idh\text{-}1^{100}$ and $Idh\text{-}1^{200}$ were present. In the remaining five species, either $Idh\text{-}1^{100}$ or $Idh\text{-}1^{200}$, but not both, were observed. Since these latter species were represented by very small sample sizes in the study (Patton and Avise, 1983), it is likely that larger samples would have revealed the presence of both *Idh-1* alleles. Alternatively, $Idh\text{-}1^{100}$ or $Idh\text{-}1^{200}$ could have been lost in some Anatinae species. In either event, these two electromorphs appear to define the Anatinae clade as shown in Fig. 2.

Group V, Plesiomorph Undetermined. For a single locus (ALB), the anatid plesiomorph could not be determined since no electromorphs were shared by Anatinae and Anserinae. Hence electromorphs at this locus cannot strictly define clades, although they can in some cases provide supporting evidence for previously characterized clades. For example, Alb^{105} was shared by all five assayed species of Anserinae, a clade earlier defined by Pgd^{200}. Many other ALB electromorphs defined autapomorphic character states. A final Hennigian tree, based upon observed distributions of electromorphs at all 17–19 assayed loci, is presented in Fig. 2.

Group V characters will normally include electromorphs of the more rapidly evolving loci. Since plesiomorphic character states for loci in this category cannot be defined by out-group criteria, these character states cannot actively contribute to phylogenetic tree generation, nor can a theory of evolution for each group V locus be explicitly defined. At these loci, character states can only be superimposed on the previously generated phylogenetic tree, and by this superimposition, some elements of the pattern of character-state transitions may be hypothesized. An example can be found in Fig. 2. Based on the previously stated assumptions for group IV loci, we have depicted *A. acuta* as sharing a node with *A. platyrhynchos*. Whereas *A. acuta* has the autapomorphic character state Alb^{98}, *A. platyrhynchos* shares the character state Alb^{100} with three other species of *Anas*. In this phylogeny Alb^{98} is hypothesized to be derived from Alb^{100}. Therefore Alb^{100} is denoted by a circle in Fig. 2, and Alb^{98} would be considered a derivation (or transition) of Alb^{100}. In an alternative scenario presented in Patton and Avise (submitted), *A. acuta* is depicted outside of the four species assemblage which shares Alb^{100}. In that representation both Alb^{100} and Alb^{98} are considered to be derived characters and no direction to the transition can be postulated.

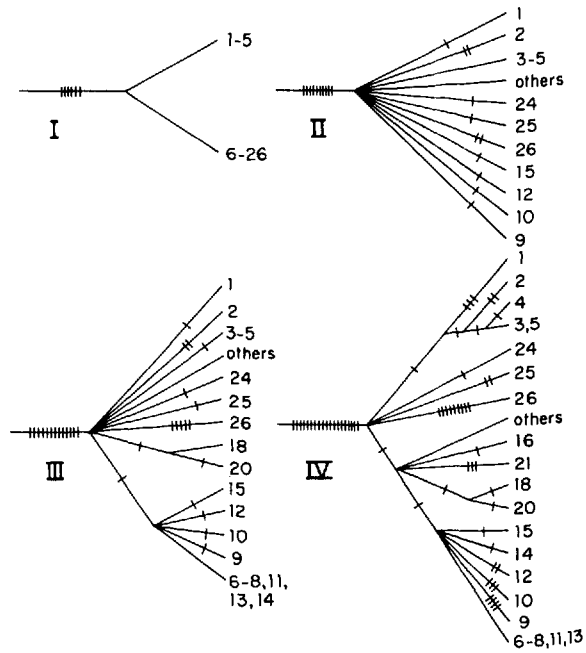

Fig. 1. Stages in the development of a qualitative Hennigian tree for waterfowl (see text), based on protein-electrophoretic data. Numbers refer to the species listed in Table 1

Results

Waterfowl

The phylogenetic tree in Fig. 2 represents one of many Hennigian trees which could be drawn for the waterfowl by altering the evolutionary interpretations of the "ambiguous" character states in Group IV, above. A second Hennigian tree for the waterfowl, which differs from the tree presented here only by placing *Anas acuta* between the nodes for *A. fulvigula* – *A. carolinensis* and *A. clypeata* – *A. cyanoptera*, is shown in Patton and Avise (submitted). We have constructed a total of 4 such alternative trees. We have also converted the genetic data for waterfowl into quantitative matrices of genetic distances among species. Nei's (1972) distance coefficients were subsequently employed to generate a phenetic tree by the UPGMA procedure (Sneath and Sokal 1973); Rogers' (1972) distance coefficients were used to generate a distance-Wagner tree (Farris 1972). These trees are pictured in Fig. 3. How well do results of these various methods of data analysis reflect the probable phylogenetic histories of waterfowl?

Fortunately, the Anatidae has been the subject of intense study by systematists for many years (Bellrose 1976; Delacour and Mayr 1945; Johnsgard 1968; Sibley and Ahlquist 1972). For those 26 species examined by

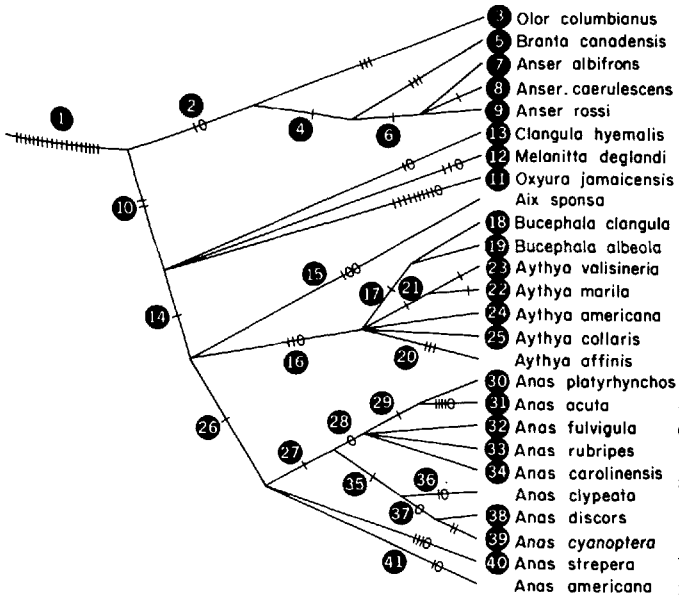

Fig. 2. Completed phylogenetic tree for waterfowl based on qualitative Hennigian analyses of 82 electromorphs encoded by 17–19 genetic loci. Slashes crossing branches of the tree represent plesiomorphs for Anatidae (branch 1), synapomorphs defining clades, or autapomorphs. Circles crossing branches represent electromorphs for which the ancestral states remain undetermined. Electromorphs postulated to have arisen along the numbered branches are listed in Table 6

us, waterfowl experts are in fairly close agreement on taxonomic (and evolutionary) relationships, and from their discussions we have distilled a summary phylogeny (Fig. 4). This summary provides a "model" against which the results of various methods of genetic data analysis may be evaluated.

Before the various genetic phylogenies are compared with the model phylogeny, an important caveat must be mentioned. As is commonly the case in conventional systematics, the exact philosophy of data analysis leading to the proposed waterfowl phylogeny (Fig. 4) is unclear. Waterfowl systematists have employed information from a wide variety of sources, including comparative morphology, development, physiology, behavior, and the fossil record. The final conclusions about phylogenetic relationships probably represent some (unspecified) combination of phenetic and cladistic reasoning. For our purposes, the important point is that we know of no a priori bias in the "model" phylogeny that would favor it to more closely resemble one or another genetic phylogeny. Furthermore, because we are comparing protein-based phylogenies against those derived from morphology, development, etc., we are in part providing a test of the non-specificity hypothesis (Rohlf and Sokal 1980), and in part providing a test of alternative philosophies of tree construction. With the data currently available to us, it is impossible to dissect the relative errors introduced by these two possibilities. Since this study is among the first to attempt strict Hennigian analyses of qualitative electrophoretic data, and hence is exploratory in design, it will be sufficient to demonstrate that protein-based Hennigian trees can provide reasonable approximations to probable evolutionary histories.

Branching sequences of the model phylogeny were preliminarily compared against those of the UPGMA, distance-Wagner, and various Hennigian trees by use of matrix correlations. Data points for the correlations were the number of nodes (or branches) separating all pairs of extant species in the model versus proposed phylogenies. Results are presented in Table 3. For the waterfowl, all correlations are positive and large, but the strongest correlations result from the Hennigian analyses. Even when we eliminate from consideration all "ambiguous" electromorphs in the Group IV loci, and hence compare Fig. 1 IV against the model phylogeny, the correlation remains 0.86.

There are several reasons for caution in interpreting these quantitative correlations. First and perhaps most important, the correlations do not take into account degrees of resolution in the phylogenies compared (Mickevich 1978). Two unresolved ("bush") trees could be highly congruent and yet contain little information about species relationships. For the waterfowl data, both the model and the Hennigian phylogenies exhibit unresolved regions while the distance-Wagner and UPGMA trees are almost fully resolved. Second, the above correlations compare only branching sequences. The express intent of Hennigian cladistics is to identify correct branching sequences, but this was never the sole goal of phenetic clustering procedures which summarize overall evolutionary divergence. To evaluate results of UPGMA against criteria that it was not solely intended to satisfy is probably unfair. Third, the correlations do not convey any information about the number of synapomorphic characters contributing to clade identification in the Hennigian trees. Hence, they say nothing about confidence that the pictured clades are reliably defined, nor about the possible effects of new or altered data in clade delineation. We will later discuss this point in greater detail.

An alternative and often more meaningful approach to comparing shapes of trees involves generation of

Table 1. Species of waterfowl included in the present study

Species	Common name
Subfamily Anserinae	
(1) *Olor columbianus*	Whistling Swan
(2) *Branta canadensis*	Canada Goose
(3) *Anser albifrons*	White-Fronted Goose
(4) *Anser caerulescens*	Snow and Blue Goose
(5) *Anser rossi*	Ross' Goose
Subfamily Anatinae	
(6) *Anas platyrhynchos*	Mallard
(7) *Anas fulvigula*	Mottled Duck
(8) *Anas rubripes*	Black Duck
(9) *Anas acuta*	Pintail
(10) *Anas strepera*	Gadwall
(11) *Anas discors*	Blue-winged Teal
(12) *Anas cyanoptera*	Cinnamon Teal
(13) *Anas carolinensis*	Green-winged Teal
(14) *Anas americana*	American Widgeon
(15) *Anas clypeata*	Northern Shoveler
(16) *Aix sponsa*	Wood Duck
(17) *Aythya americana*	Redhead
(18) *Aythya valisineria*	Canvasback
(19) *Aythya collaris*	Ring-necked Duck
(20) *Aythya marila*	Greater Scaup
(21) *Aythya affinis*	Lesser Scaup
(22) *Bucephala clangula*	Goldeneye
(23) *Bucephala albeola*	Bufflehead
(24) *Clangula hyemalis*	Oldsquaw
(25) *Melanitta deglandi*	Surf Scoter
(26) *Oxyura jamaicensis*	Ruddy Duck

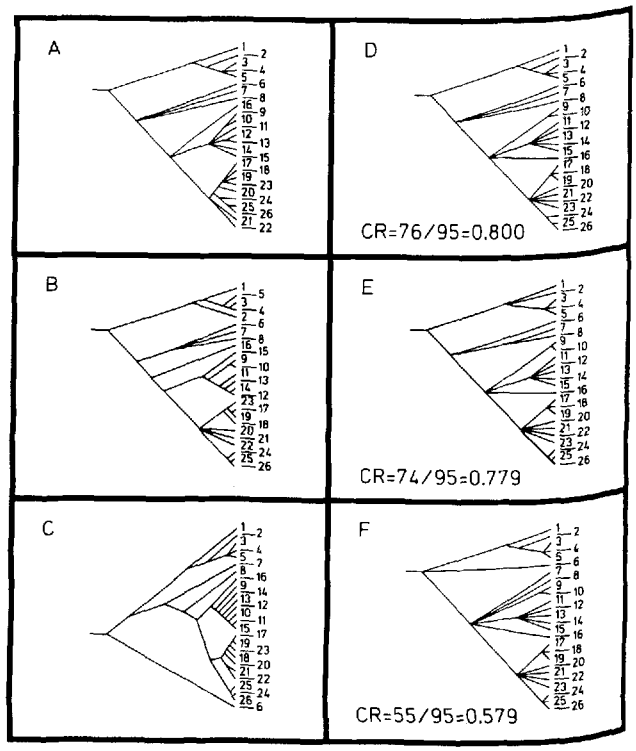

Fig. 3. Relationships of anatid taxa (numbered as in Table 1) based on electrophoretic data of Patton and Avise (submitted) when analysed by A) Hennigian B) distance-Wagner and C) UPGMA methods. The respective congruence trees, and their coefficients of resolution (CR) when compared to the model phylogeny, are shown in D, E, and F

"consensus" trees (Adams 1972). A consensus tree represents *only* the information that is shared by two (or more) rival trees under comparison. Any information not present in all rivals is absent from the consensus tree. Hence the degree of information content in a consensus tree is high only when the rival trees are very similar (Adams 1972). Portions D, E, and F of Fig. 3 illustrate three consensus trees for waterfowl, representing combinations of Hennigian and Model, Distance-Wagner and Model, and UPGMA and Model, respectively.

To quantify the amount of resolution of the consensus tree relative to the model phylogeny, we used a measure of resolution here termed the coefficient of resolution (CR) which is the sum of the number of nodes (branch points) from the root to each species of the consensus tree (excluding the node of the root) divided by the sum of the number of nodes from the root to each species of the model tree. The calculated value represents the percent resolution of the consensus relative to the model tree. Values of this formula can range from 1.00, where the consensus tree and model phylogeny are equivalent, to 0, where the resulting consensus tree is a totally unresolved bush phylogeny. Results are presented in Table 4 and Fig. 3. For the waterfowl, all consensus trees are largely resolved relative to the model but the greatest resolution results from the Hennigian analysis.

Overall, it is clear that the Hennigian analysis of the waterfowl data has identified a number of clades which correspond closely to the natural groups described by other systematists. The subfamily Anserinae is apparently defined by the synapomorph Pgd^{200}, and supported by Alb^{105}; Anatinae is defined by $Idh\text{-}1^{100}$ and $Idh\text{-}2^{100}$; the geese (*Anser* and *Branta*), by $Idh\text{-}1^{120}$; *Anser*, by $Ck\text{-}2^{140}$; *Anas*, by Pgm^{100}; and the "blue-winged" ducks (*Anas clypeata*, *A. discors*, and *A. cyanoptera*), by $Ldh\text{-}1^{95}$. It is primarily through such case-by-case analyses (described in greater detail in Patton and Avise, submitted) that we have gained some confidence in the utility of this qualitative approach for evaluating electrophoretic data.

Other Organisms

We have employed a similar protocol to evaluate electrophoretic data sets in several other groups of organisms for which it was also possible to derive "model" phylogenies from the literature. Figure 5 shows a "model" phylogeny for North American rodents (Cricetidae), and a Hennigian cladistic tree based on 103 electromorphs encoded by 15 loci; Figure 6 shows a "model" phylogeny for bats (primarily of the genus *Lasiurus*), and a Hennigian cladistic tree based on 71 electromorphs encoded by 21 loci. Detailed descriptions of these data

Table 2. Electromorphs observed in samples of the 26 species of Anatidae numbered in Table 1. The electromorphs were labeled according to gel mobility relative to the common electromorph at each locus in species 6. At each of six additional loci, all taxa appeared monomorphic for the same allelic product. Loci are listed by standard abbreviations which can be found in Avise et al. (1980) and references therein

Locus	*Species*																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(A) GOT-1	80	100 130	100	100	100	100	100	100	100 -50	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	300 200
(B) HB	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-90	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-120	-140	-100
(C) MDH-1	100	100 70	100	100	100	100	100	100	100	100	100	100	100	100	100 90	100	100	100	100	100	100	100	100	100	100	100
(D) MDH-2	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100 100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100	-100
(E) CK-2	–	100	140	140	140	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	200
(F) PEPT-2	–	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100 95	100	100 95	100	100	100	100	100	100
(G) PGM	50	50	50	50	50	100	100	100	100 50	100	100	100	100	100	100	50	50	50	50	50 10	50	50	50	50	50	140 120
(H) GPD	60	75	75	75 30	75	100	100	100	100	80 100 40	100	100	100	100 45	100	100 120	–	100	100	100	125 100 70	100	100 70	75	72	130 85
(I) IDH-1	125	120 90	120 90	120 90	120 90	200 100	100	200 100	160 200 100 50	200 100	200 100	200 100	200 100	200 100	200 100	200 100 90	200	200 100	200 100	200 100	200 100	180 100	180 100	200 100	100	200 100
(J) PGD	200	200	200 100	200	200	100	100	100	110 100 130	100	100	110 100 10	100	100	100	100	120 100	100	100	100	100	100	120 100	100	100	100
(K) PGI	100	100	50	50	50	200 100	100	100	200 100	100	100	100	100	100	100	100	100	100	100	100	210 100 180	100	100	100	50	90 45
(L) LDH-1	100	100	100	100	100	100	100	100	100	100	95	95	100	100	95	100	100	100 180	100	100	100	100	100	100	100	100
(M) *ALB*	105	105 95	105	105	105	100	100	100	98	97	102	102	100	93	101	101.5 104	103	103	103	103	103	103	103	110	106	107

Table 3. Coefficients of correlation between "model" phylogenies and trees produced by various methods of analysis of protein electrophoresis information. Data for the correlations were the numbers of nodes separating pairs of extant species [a]

Group	# Pairwise comparisons	Analysis			
		UPGMA	Distance-Wagner	Mean Hennigian	Hennigian (range; # of trees tested)
Waterfowl	325	0.62	0.78	0.87	(0.86–0.90;4)
Rodents	91	0.84	0.82	0.90	(0.88–0.92;3)
Bats	55	0.87	0.84	0.83	(0.80–0.86;2)
	$\bar{r}$	0.78	0.81	0.87	

a See text for necessary reservations in interpreting these correlations

Table 4. Coefficients of Resolution between "model" phylogenies and consensus trees produced by comparing "model" phylogenies with trees produced by various methods of analysis of protein electrophoresis information (see text)

Group	# of nodes of model phylogeny	Analysis			
		UPGMA	Distance-Wagner	Mean Hennigian	Hennigian (range; # of trees tested)
Waterfowl	95	0.579	0.779	0.800	(0.800;4)
Rodents	66	0.606	0.606	0.737	(0.682–0.818;3)
Bats	39	0.615	0.615	0.718	(0.718;2)
	$\overline{CR}$	0.600	0.667	0.752	

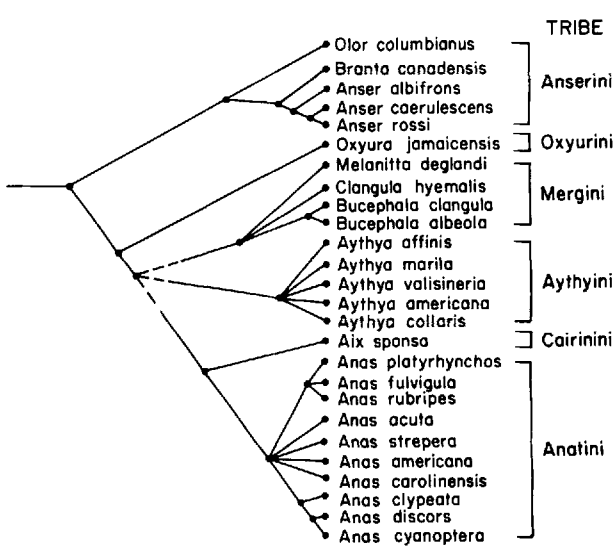

Fig. 4. "Model" phylogeny of waterfowl based on classical systematic criteria (Delacour and Mayr 1945; Johnsgard 1968)

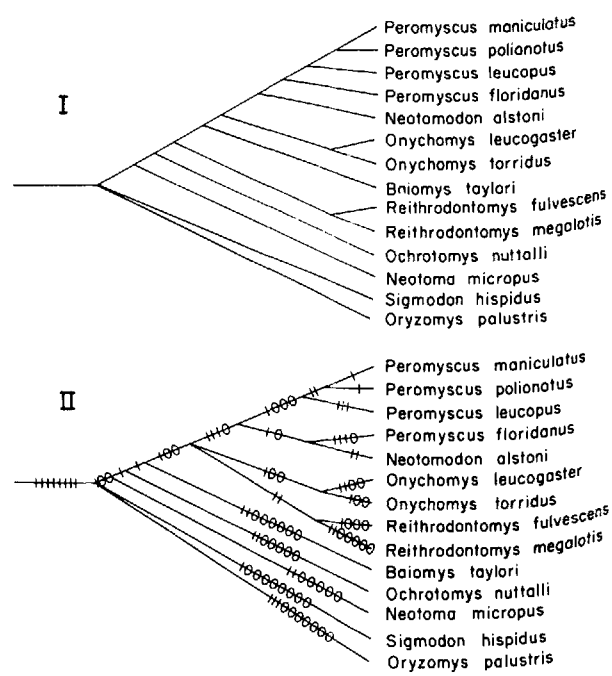

Fig. 5. Above, "model" phylogeny for cricetine rodents based on non-molecular data (Patton et al. 1980). Below, phylogenetic tree for rodents based on qualitative Hennigian analyses of protein-electrophoretic data (see legend to Fig. 2)

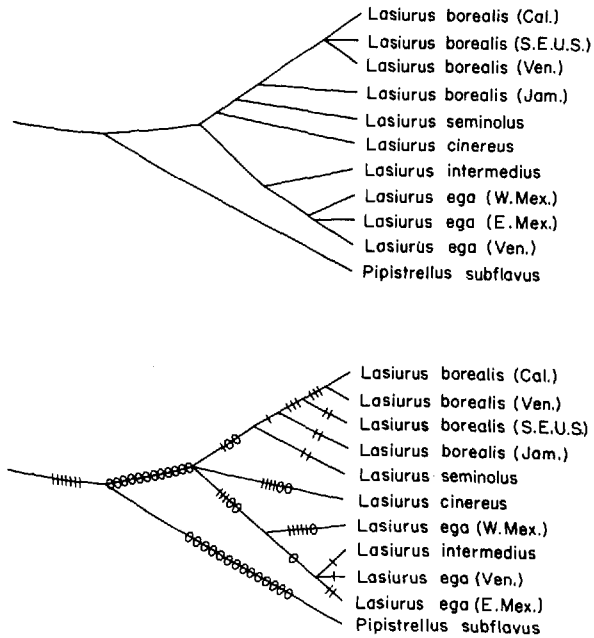

Fig. 6. Above, "model" phylogeny for lasiurine bats based on non-molecular data (Baker et al. submitted). Below, phylogenetic tree for bats based on qualitative Hennigian analyses of protein-electrophoretic data (see legend to Fig. 2)

sets and the systematic conclusions to be drawn from them can be found in Patton et al. (1981) and Baker et al. (submitted).

Results of comparisons of various genetic trees against the respective model phylogenies are given in Table 4. Overall, the results are similar to those for the waterfowl. All trees are fairly highly resolved, and the congruences of the Hennigian trees to the model phylogenies are superior in each case to those of the UPGMA and distance-Wagner trees. As above, reservations apply to these interpretations, but it is evident that the Hennigian analyses are providing useful information. In fact, as detailed by Patton et al. (1981) and Baker et al. (submitted), in a few instances the Hennigian analysis appears to have helped resolve some specific systematic controversies.

We have also successfully appied Hennigian analyses to electrophoretic data on North American thrushes, Muscicapidae (Avise et al. 1980a). In this case, a model phylogeny was unavailable for comparison, so we have not included results in this summary. In each of two other families of birds, Emberizidae and Parulidae, we were unable to provide Hennigian trees of electrophoretic data (Avise et al. 1980b,c). In the Emberizidae, an appropriate "outgroup" could not be determined, and in the Parulidae we were simply unable to identify enough synapomorphs for meaningful clade description. Thus in our experience involving routine electrophoretic data sets of about 15–30 loci, we have sometimes, though not always, felt it appropriate to include qualitative Hennigian approaches in the data summary.

Discussion

The above attempts to apply Hennigian cladistics to qualitative molecular genetic data have disclosed both weaknesses and strengths of the approach.

Weaknesses

A serious weakness is the likelihood of producing a large number of alternative Hennigian trees for a given data set. In situations where the distributions of apparent derived electromorphs do not coincide, alternative and sometimes equally-attractive explanations for the inconsistencies will result in alternate trees. For example, where synapomorphic character states at two or more loci identify different but partially overlapping clades, it is possible that (1) all states were present in the ancestral taxon, distributed to the species in a cladistically uninformative manner, or (2) one or another derived state arose independently in some extant species by convergence. Several specific examples of these and similar situations have been discussed in detail for the ambiguous "Group IV" genetic characters in the waterfowl.

In the conventional electrophoretic data sets we have examined (involving about 15–30 loci), only a small proportion of total electromorphs appear to define syna-

Table 5. Relative numbers of ancestral, unique, and derived electromorphs observed in various phylads, as determined by Hennigian analyses

Phylad	# taxa	# loci	Electromorphs	# (and percent of total) electromorphs			
				Plesiomorphic to entire group	Synapomorphic	Autapomorphic	Other[a]
Waterfowl	26	17–19	82	19 (23.2)	14 (17.1)	45 (54.9)	4 (4.9)
Rodents	14	15	103	8 (7.8)	14 (13.6)	70 (68.0)	11 (10.7)
Bats	11	21	71	7 (9.8)	11 (15.5)	36 (50.7)	17 (23.9)

[a] Non-autapomorphic characters at loci for which the plesiomorph remains undetermined

pomorphs (Table 5). Hence clade identification rests entirely on a few characters, and several regions of the tree remain cladistically unresolved. For example, while a total of 82 electromorphs were identified in waterfowl, only 14 of these (17%) are synapomorphs which actively contribute to clade definition (Table 5). This is an empirical rather than theoretical objection to the Hennigian approach, but it will likely limit the general applicability of qualitative Hennigian analyses to electrophoretic data. (On the other hand, by cladistic reasoning these 14 waterfowl characters are the *only* ones in the available data which do in fact contain any information about branching sequences in the tree. Other approaches (such as UPGMA) which include data from retained plesiomorphic and autapomorphic characters are likely to produce incorrect branching patterns.)

One consequence of this empirical problem is that clades identified in the final Hennigian tree are likely to be highly dependent upon the particular loci examined, and furthermore, that branching sequences in the tree may be sensitive to any misscoring of electromorphs. Presumably these problems could be overcome if much larger data bases were available for analysis. In such situations, it might be useful to count the numbers of synapomorphs contributing to the definitions of various clades, and to rest the strongest conclusions on clades identified independently by many derived characters (Throckmorton 1978).

Another potential problem is that very rare electromorphs carry as much weight as abundant electromorphs in the above Hennigian analyses. For example, in the waterfowl, the important Anserinae clade is defined only by Pgd^{200} (this is subsequently supported by Alb^{105}). Pgd^{200} is identified as a synapomorph because Pgd^{100} is the apparent plesiomorph for Anatidae, shared by Anserinae and Anatinae. Yet Pgd^{100} was observed only in one Anserinae species (*Anser albifrons*), and only in 5 percent frequency in that species. It could easily have gone unsampled in an electrophoretic survey. A low frequency presence versus absence of Pgd^{100} in *Anser* would have virtually no effect on the contribution of PGD to the UPGMA clustering of Anserinae species, but it makes a great difference to the Hennigian identification of the Anserinae clade.

In the waterfowl, we sampled an average of eight individuals per species. It might be argued that had our samples been much larger (in theory, limited only by the sizes of the species) eventually most or all electromorphs in the entire family would be found at some finite frequency within each species. If true, by strict Hennigian criteria few synapomorphs or clades could be identified. In such cases where large sample sizes are employed and many rare electromorphs are found, a modified procedure would be to disregard any electromorphs present in a species below some arbitrary frequency, say one percent. This procedure might be justifiable. It seems likely that such electromorphs shared by species might often have evolved independently, rather than have been retained at low frequency for long periods of time from a common ancestor.

In applying Hennigian cladistics to electrophoretic data, proper choice of an outgroup or sister-group clade is critical. The outgroup should be unequivocally distinct from the clade being investigated, yet closely enough related to maximize the number of plesiomorphic characters for the group analyzed. If the outgroup is too distantly related to the group under study, few plesiomorphs (and hence few known synapomorphs) can be identified for the entire group. If the form(s) chosen as

Table 6. Electromorphs defining branches of the Hennigian tree in Fig. 2. Line 1 lists electromorphs presumed plesiomorphic to Anatidae, and successive lines list apomorphic characters. Electromorphs in brackets are those whose presumed ancestral states remain undetermined

(1) $Mdh\text{-}1^{100}$, $Mdh\text{-}2^{100}$, $Ldh\text{-}1^{100}$, $Ldh\text{-}2^{100}$, $Idh\text{-}1^{90}$, $Idh\text{-}2^{100}$, Pgd^{100}, Gpd^{75}, Pgi^{100}, Pgi^{50}, $Ck\text{-}2^{100}$, $Ck\text{-}3^{100}$, $Got\text{-}1^{100}$, $Got\text{-}2^{100}$, Ipo^{100}, $Pept\text{-}1^{100}$, $Pept\text{-}2^{100}$, Hb^{100}
(2) Pgd^{200}, [Alb^{105}]
(3) $Idh\text{-}1^{125}$, Gpd^{60}, $Got\text{-}1^{80}$
(4) $Idh\text{-}1^{120}$
(5) $Mdh\text{-}1^{70}$, $Got\text{-}1^{130}$, [Alb^{95}]
(6) $Ck\text{-}2^{140}$
(7) (speciation)
(8) Gpd^{30}
(9) (speciation)
(10) $Idh\text{-}1^{100}$, $Idh\text{-}1^{200}$
(11) Gpd^{130}, Gpd^{85}, Pgi^{90}, Pgi^{45}, Pgm^{140}, Pgm^{120}, $Ck\text{-}2^{200}$, $Got\text{-}1^{300}$, $Got\text{-}1^{200}$, [Alb^{107}]
(12) Gpd^{72}, Hb^{140}, [Alb^{106}]
(13) Hb^{120}, [Alb^{110}]
(14) Gpd^{100}
(15) Gpd^{120}, [Alb^{104}, $Alb^{101.5}$]
(16) Pgd^{120}, Gpd^{70}, [Alb^{103}]
(17) $Idh\text{-}1^{180}$
(18) (speciation)
(19) (speciation)
(20) Gpd^{125}, Pgi^{210}, Pgi^{180}
(21) $Pept\text{-}2^{95}$
(22) Pgm^{10}
(23) $Ldh\text{-}1^{180}$
(24) (speciation)
(25) (speciation)
(26) Pgm^{100}
(27) Pgd^{110}
(28) [Alb^{100}]
(29) Pgi^{200}
(30) (speciation)
(31) $Idh\text{-}1^{160}$, $Idh\text{-}1^{50}$, $Got\text{-}1^{50}$, Pgd^{130}, [Alb^{98}]
(32) (speciation)
(33) (speciation)
(34) (speciation)
(35) $Ldh\text{-}1^{95}$
(36) $Mdh\text{-}1^{90}$, [Alb^{101}]
(37) [Alb^{102}]
(38) (speciation)
(39) Pgd^{10}, Hb^{90}
(40) $Mdh\text{-}2^{100}$, Gpd^{80}, Gpd^{40}, [Alb^{97}]
(41) Gpd^{45}, [Alb^{93}]

the outgroup resides phylogenetically inside the group in question, the result will be either unresolved clades with respect to the outgroup, or groupings defined on the basis of symplesiomorphs incorrectly treated as synapomorphs.

Finally, electromorphs at a given locus represent qualitative multistate traits whose relative evolutionary relationships cannot be inferred by migration position on gels. Thus there is no evolutionary directionality apparent from examination of the electromorphs themselves; such directionality must be inferred indirectly from Hennigian (for example) analyses of electromorph distributions across taxa.

Strengths

Despite these numerous reservations, in each of the major electrophoretic data sets summarized in this paper, a strict application of qualitative Hennigian principles has resulted in phylogenies which correspond very well to "model" classifications previously defined by systematists employing non-molecular data. The branching sequences of the Hennigian trees fit those of the model phylogenies as well as or better than do the branching sequences implied by UPGMA or distance-Wagner analyses.

We would not expect UPGMA and Hennigian trees to be identical, in part because they are based on philosophically different grounds. It is not our intent here to recapitulate the merits of philosophical issues underlying phenetic versus cladistic analyses; these are debated at length in the literature. But the various trees generated above do permit an examination of the pragmatic consequences of Hennigian versus phenetic procedures. One final example from the waterfowl will suffice. The Ruddy duck (*Oxyura jamaicensis*) falls well outside the phenetic cluster of all other waterfowl (including both Anserinae and Anatinae) in the UPGMA analysis of the electrophoretic data (Patton and Avise submitted). In the Hennigian analysis of these same data, *Oxyura* belongs to the Anatinae clade by shared possession of the derived electromorphs $Idh\text{-}1^{100}$ and $Idh\text{-}1^{200}$ (Fig. 2). The traditional phylogeny of waterfowl also places *Oxyura* within Anatinae, so the Hennigian and "model" classifications agree. The phenetic distinctness of *Oxyura* from all other waterfowl results from its possession of a large number (ten) of autapomorphic character states which provide no basis for cladistic inference.

Nonetheless, we feel that the major strengths of the Hennigian approach result from the explicit treatment of individual, qualitative characters. Whether or not the final distributions of electromorphs along a particular Hennigian tree are correct, they are at least expressly defined. The distribution of electromorphs on the Hennigian tree for waterfowl, for example, is given in Table 6. Such summaries of information are not possible in phenetic clustering or distance-Wagner analyses, because the particular characters ultimately contributing to tree structure are first submerged in a distance matrix.

There are several advantages to the manipulation of individual characters and to the qualitative output of Hennigian trees: (1) for electrophoretic data sets of conventional size, the analysis can be performed quickly and by hand without need for a computer (the "Wagner 78" program of James Farris, which apparently employs a similar approach, could be used for larger data sets); (2) as a consequence, the practitioner gains a strong "feel" for the data, and for the relative contributions of particular loci to the final phylogenetic conclusions; (3) additional data, either from new loci or from new species, can quickly be added to earlier Hennigian trees (in contrast, for phenetic or distance-Wagner trees, any new information must first be used to recalculate all genetic distances before a new tree can be generated); (4) the qualitative output of Hennigian analysis provides a degree of testability not apparent in distance-manipulating algorithms.

This last point warrants elaboration. Because discrete character states are defined for all branches of the Hennigian tree, points of ambiguity in the tree may be specifically identified and advocated as important areas for further study. Several such cases were described for the waterfowl. Suppose for example that two cladistically distant species (belonging to separate clades as provisionally defined by several loci in a survey) nonetheless appear to share a derived electromorph as determined by comparison with the outgroup. One then might question the identity of the electromorph. An error may have arisen in scoring or in data transfer. Or it might be the case that additional laboratory study would reveal that distinct allelic products had been masquerading as a common electrophoretic band. If the electromorphs did indeed prove to represent identical proteins, they could (1) have evolved convergently (an interesting molecular phenomenon), (2) have been retained in common from a polymorphic ancestor, or (3) actually represent synapomorphs defining a tree very different from that provisionally proposed. At least the problem area is identified and alternative hypotheses advanced to account for the dilemma.

Our goal has not been to advocate Hennigian evaluations of electrophoretic data, but rather to objectively evaluate the logical consequences of qualitative data analyses when applied to real sets of electrophoretic data. A number of strengths characterize this qualitative approach, but these are counterbalanced by several disadvantages. In summary, we can propose that a more valuable and comprehensive program of genetic data summary would routinely involve both phenetic and qualitative cladistic approaches.

Surprisingly, cladistic analyses have seldom been applied to electrophoretic or other qualitative molecular data (Mickevich and Johnson 1976). Perhaps one reason for this neglect was the early discovery that the decay

of overall genetic similarity among species is progressive and correlated with period of time since separation of gene pools (Langley and Fitch 1974; Nei 1975; Wilson et al. 1977; Zuckerkandl and Pauling 1962). While a preoccupation with the development and calibration of "molecular clocks" has had a tremendous positive impact on the field of molecular evolution, it has also served to divert attention from analyses of the character states themselves. Perhaps the approach employed in this study, when applied to additional data, may help to redress this imbalance.

Acknowledgments. J.C.P was supported in part by an NIH traineeship. Work was also supported by NSF grants DEB-7814195 and DEB-8022135. We thank Niles Eldredge and one anonymous reviewer for constructive reviews of the manuscript. We also thank James Farris and Mary Mickevich for insightful comments, although this does not imply total agreement between the arguments presented herein and their positions.

References

Adams EN III (1972) Consensus techniques and the comparison of taxonomic trees. Syst Zool 21:390–397

Avise JC, Patton JC, Aquadro CF (1980a) Evolutionary genetics of birds. I. Relationships among North American thrushes and allies. The Auk 97:135–147

Avise JC, Patton JC, Aquadro CF (1980b) Evolutionary genetics of birds. II. Conservative protein evolution in North American sparrows and relatives. Syst Zool 29:323–335

Avise JC, Patton JC, Aquadro CF (1980c) Comparative molecular evolution in New World warblers (Parulidae) and rodents (Cricetinae). J Heredity 71:303–310

Bellrose FC (1976) Ducks, Geese and Swans of North America. Stockpole Books, Morrison, PA

Delacour J, Mayr E (1945) The family Anatidae. Wilson Bull 57: 3–35

Farris JS (1972) Estimating phylogenetic trees from distance matrices. Am Nat 106:645–668

Farris JS (1974) A comment on evolution in the *Drosophila obscura* species group. Evolution 28:158–160

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Gorman CG, Wilson AC, Nakanishi M (1971) A biochemical approach towards the study of reptilian phylogeny: evolution of serum albumin and lactic dehydrogenase. Syst Zool 20: 167–185

Hennig W (1966) Phylogenetic Systematics. Univ Illinois Press, Chicago

Howard H (1964) Fossil Anseriformes. In: Delacour J (ed) The Waterfowl of the World, Vol 4, Country Life Ltd, London, pp 233–326

Johnsgard PA (1968) Waterfowl. Nebraska Press, Lincoln, Nebraska

Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. J Mol Evol 3:161–177

Mickevich MF (1978) Taxonomic congruence. Syst Zool 27: 143–158

Mickevich MF, Johnson MS (1976) Congruence between morphological and allozyme data in evolutionary inference and character evolution. Syst Zool 25:260–270

Nei M (1972) Genetic distance between populations. Amer Nat 106:283–292

Nei M (1975) Molecular Population Genetics and Evolution. North-Holland, Amsterdam

Patton JC, Baker RJ, Avise JC (1981) Biochemical evolution in Peromyscine rodents. In: Smith MH, Joule J (eds) Mammalian Population Genetics, Univ Georgia Press, pp 288–308

Prager EM, Wilson AC (1978) Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. J Mol Evol 11:129–142

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in Genetics VII, Univ Texas Publ 7213: 145–153

Rohlf FJ, Sokal RR (1980) Comments on taxonomic congruence. Syst Zool 29:97–101

Romer AS (1966) Vertebrate Paleontology. Univ of Chicago Press, Chicago

Sibley CG, Ahlquist JE (1972) A comparative study of the egg-white proteins of non-passerine birds. Peabody Mus Nat Hist Bull 39, Yale Univ, New Haven, Connecticut

Sneath PHA, Sokal RR (1973) Numerical Taxonomy. WH Freeman & Co, San Francisco

Throckmorton LH (1978) Molecular phylogenetics. In: Foote RH, Knutson L, Lentz PL (eds) Biosystematics in Agriculture, John Wiley & Sons, New York, pp 221–239

Wiley EO (1981) Phylogenetics. John Wiley & Sons, New York

Wilson AC, Carlson SC, White TJ (1977) Biochemical Evolution. Ann Rev Biochem 46:573–639

Zuckerkandl E, Pauling L (1962) Molecular disease, evolution and genic heterogeneity. In: Kasha M, Pullman N (eds) Academic Press, New York

Received October 15, 1981/Revised January 7, 1983