

THE QUARTERLY REVIEW of BIOLOGY



PHENETIC CLUSTERING IN BIOLOGY: A CRITIQUE

KEVIN DE QUEIROZ

*Department of Vertebrate Zoology, National Museum of Natural History
Smithsonian Institution, Washington, DC 20560 USA*

DAVID A. GOOD

*Museum of Natural Science, Louisiana State University
Baton Rouge, Louisiana 70803 USA*

ABSTRACT

Phenetic clustering, the forming of hierarchical nonoverlapping groups strictly according to degree of similarity, has serious shortcomings as it is commonly used in biology. When used as a method for estimating phylogeny, phenetic clustering rests on a questionable assumption of correspondence between similarity and recency of common ancestry. This compromises its ability to reconstruct the correct branching sequence when rates of evolutionary divergence are unequal among lineages, as well as causing it to obscure rate differences even when the branching sequence is reconstructed correctly. When used as a method for analysing patterns of geographic variation and genetic continuity among populations, phenetic clustering rests on a questionable assumption of correspondence between similarity and degree of genetic continuity. This compromises its ability to identify genetically continuous units when their component populations are differentiated, and combined with its sensitivity to uneven geographic sampling, it can cause the method to yield misleading results if sampling patterns are not taken into consideration. Finally, even when used simply as a method for analysing patterns of similarity without regard to causal processes, phenetic clustering rests on a questionable assumption of nested hierarchical structure. This compromises its ability to represent similarity relationships accurately when those relationships exhibit a significant nonhierarchical component. For all of the common biological applications of phenetic clustering, there exist alternative analytical methods that do not suffer from the problems associated with phenetic clustering. The problems in question result not from the phenetic (similarity) data themselves, which often can be analysed in more appropriate ways, but from the phenetic clustering procedure. At least some of the limitations of phenetic clustering as well as the advantages of alternative methods have been known for many years. Advocacy of phenetic clustering at the expense of more appropriate methods can be explained as the result of constraints imposed by an implicit assumption of nested hierarchies that was part of the taxonomic context within which the methods were developed.

The Quarterly Review of Biology, March 1997, Vol. 72, No. 1

Copyright © 1997 by The University of Chicago. All rights reserved.

0033-5770/97/7201-0001\$1.00

INTRODUCTION

PHENETIC CLUSTERING is a class of methods used to form groups based on similarities among entities. Although the popularity of these methods has waned somewhat in recent years, they continue to be used for diverse purposes by biologists, including analyses of phylogenetic relationships among species and patterns of variation within them, as well as the general representation of similarity among diverse kinds of entities, from populations, species, and higher taxa to biogeographic areas and ecological communities (reviewed by Sneath and Sokal 1973; Clifford and Stephenson 1975; see Everitt 1993 for other applications). There are, however, problems that impose serious limitations on the use of phenetic clustering—both as applied to specific biological questions and in the general analysis of similarity patterns. Here we critically review the use of phenetic clustering in the specific biological applications of phylogeny reconstruction and analyses of variation among potentially conspecific populations, as well as in the general analysis of similarity. In addition to pointing out problems with phenetic clustering, we call attention to more appropriate methods of analysis that either already exist or currently are being developed. Finally, by examining the historical and disciplinary context within which phenetic clustering gained its popularity, we attempt to explain why this class of methods became popular despite an early awareness of its shortcomings.

BASIC TERMS AND PRINCIPLES

Cain and Harrison (1960) distinguished between *phenetic* taxonomic arrangements based on overall similarity (including both phenotypic and genetic components) and *phyletic* arrangements based on evolutionary relationships. They have been followed by numerous subsequent authors in equating phenetic affinity (relationship) with overall similarity. A multitude of methods have been developed for analysing phenetic relationships, most of which fit into one or the other of two major categories. The units of analysis, whether individual organisms or groups of organisms (at any hierarchical level), are termed *operational taxonomic units* or *OTUs* (Sokal and Sneath 1963). *Ordination* methods array OTUs in a

continuous hyperspace whose dimensions are defined by the characters of the OTUs; in contrast, *clustering* methods assign OTUs to groups (Sneath and Sokal 1973; Clifford and Stephenson 1975; Dunn and Everitt 1982). Methods based on graph theory (e.g., Kruskal 1956; Prim 1957) can be used to connect OTUs in character space, combining elements of both clustering and ordination.

Cluster analysis originally referred to a method that produced partly overlapping groups (Tryon 1939; cited by Michener and Sokal 1957; Sokal and Michener 1958), and sometimes the term still is applied to partly overlapping and nonhierarchical methods (Sneath and Sokal 1973). Nevertheless, in biology, clustering effectively has become synonymous with methods that produce *hierarchical non-overlapping groups*. Such groups are either *nested* (one included entirely within another) or *mutually exclusive* (have no members in common). Hierarchical nonoverlapping clustering methods include methods known as *single linkage*, *complete linkage*, and *average linkage* clustering. Among biologists, by far the most popular phenetic clustering method is a member of the class of average linkage clustering methods, the *unweighted pair-group method using arithmetic averages* or *UPGMA* (Sokal 1986).

Phenetic clustering is often described as a two-step process (reviewed by Jardine and Sibson 1971; Sneath and Sokal 1973; Clifford and Stephenson 1975; Hartigan 1975; Dunn and Everitt 1982; Romesburg 1984; Everitt 1993). After initial data collection, overall similarity is first estimated over a set of characters by means of a *coefficient of resemblance* or *similarity coefficient*, which assigns a value to each pair of OTUs (for example, the value of the coefficient known as the mean character difference is calculated by determining the absolute values of the differences between the two OTUs for each character, summing those values over all characters, and then dividing by the total number of characters). We will follow Sneath and Sokal (1973) in using these terms to refer both to coefficients that increase with increasing similarity and to those that increase with increasing dissimilarity or distance. After calculating similarity values for all pairs of OTUs, a matrix of such values is assembled and subjected to a clustering algorithm, which forms

clusters or groups of similar OTUs according to some mathematical criterion. Although raw data commonly are converted to similarity values, sometimes raw data are collected directly as similarities (e.g., quantitative immunological techniques and DNA-DNA hybridization). Furthermore, the problems discussed in the present article apply to the group-forming process rather than the data on which it is used. For these reasons, we distinguish between *phenetic (similarity) data* and *phenetic clustering methods*.

The results of phenetic clustering (hierarchical nonoverlapping groups) can be represented in a variety of ways (reviewed by Sneath and Sokal 1973), but they are most commonly represented as rooted, diverging trees. Most subsequent authors have followed the terminology of Mayr (1965) and Camin and Sokal (1965), who distinguished between *phenograms*, tree-like branching diagrams representing phenetic relationships, and *phylograms* (including *cladograms* and *phylogenetic trees*), which also take a tree-like form but represent evolutionary relationships. Although our review will focus on average linkage clustering methods, particularly the UPGMA, most of our conclusions apply to other hierarchical nonoverlapping phenetic clustering methods and the diagrams, branching or otherwise, derived from them.

PHENETIC CLUSTERING AND PHYLOGENY RECONSTRUCTION

Phenetic clustering is commonly used as a method for estimating or reconstructing phylogeny (e.g., Colless 1970; Prager and Wilson 1978; Tatenko et al. 1982; Nei et al. 1983; Nei 1987; Sourdis and Krimbas 1987; Saitou and Nei 1987; Felsenstein 1988; Swofford and Olsen 1990). Although this application of clustering has declined in recent years, it has nonetheless persisted, particularly in molecular systematics. Among the various problems with phenetic clustering, its limitations in estimating phylogeny are relatively well known. In particular, problems caused by rate variations among lineages have been discussed by several authors (Michener and Sokal 1957; Sokal and Sneath 1963; Jardine et al. 1969; Kirsch 1969; Colless 1970; Farris 1971, 1972; Moore 1971; Sneath and Sokal 1973; Felsenstein 1982a; So-

ber 1983a; Sober 1988), and can be summarized as follows.

In order for phenetic clustering to yield accurate reconstructions of phylogeny, the degree of similarity among OTUs must correspond with their recency of common ancestry. For contemporary OTUs, this condition implies that the similarity data being analysed must have been generated by roughly uniform rates of evolutionary divergence among their lineages (so that the similarities themselves are approximately ultrametric). If this implicit assumption is violated, then phenetic clustering may give erroneous results; in other words, the branching pattern of the phenogram may not correspond with the branching pattern of phylogenetic divergence.

The problem sometimes has been stated in terms of alternative interpretations of the remote clustering of an OTU in a phenogram, referred to as the "pregroup-exgroup problem" (Michener and Sokal 1957; Sokal and Sneath 1963). Attachment of an OTU outside of a cluster of OTUs can result from two very different evolutionary possibilities. On the one hand, the OTU in question may be a *pregroup* derivative, the terminus of a lineage that diverged from the lineages of the cluster before they diverged from one another. In this case, the remote position of the OTU in the phenogram accurately reflects its phylogenetic position. Alternatively, the OTU in question may be an *exgroup*, the terminus of a highly modified lineage descended from ancestors within the cluster. In this case, the remote position of the OTU in the phenogram misrepresents its phylogenetic position. A high degree of modification in one of the OTUs stemming from a particular common ancestor implies an increase in the rate of evolution in its lineage. Therefore, remote clustering of an exgroup OTU from its close phylogenetic relatives, and the erroneous conclusions that would be reached by interpreting the phenogram as a phylogenetic tree, are attributable to variation in rates of evolution among lineages.

Colless (1970) pointed out that absolute constancy of evolutionary rates is not necessary for phenetic clustering to give accurate estimates of phylogeny. Minor variations should not cause problems. Furthermore, even a substantial rate increase or decrease within a sin-

gle lineage can be offset by a later change in the opposite direction, and a rate increase or decrease in one member of a pair of sister lineages can be offset by a similar change in the other member of the pair. In general, rate variations should not lead to erroneous conclusions so long as the total amount of evolutionary change between OTUs stemming from a particular common ancestor does not exceed that between any of those OTUs and other OTUs sharing only more remote common ancestors with them.

Variation in evolutionary rates is a well-known phenomenon. It has been demonstrated repeatedly even in molecular data (see Moritz and Hillis 1990; Li 1993; and references therein), notwithstanding the existence of a controversial proposition that molecular sequence evolution is related directly to absolute time (e.g., Zuckerkandl and Pauling 1962; Wilson et al. 1977, 1987; Jukes 1987; Li 1993). The accuracy of phenetic clustering as a means of estimating phylogeny under different kinds and amounts of rate variation has been investigated in simulation studies (e.g., Tateno et al. 1982; Fiala and Sokal 1985; Sourdis and Krimbas 1987; Rohlf et al. 1990; Jin and Nei 1991; DeBry 1992; Kim et al. 1993; Huelsenbeck and Hillis 1993; Huelsenbeck 1995), which confirm the sensitivity of phenetic clustering to variation in evolutionary rates among lineages. Nevertheless, the questions remain as to whether rate variations that violate the required conditions exist in real evolutionary lineages and how they are to be detected in practice.

Potential problems often can be detected using information in similarity values. One method, the relative rate test (Sarich and Wilson 1967a,b; Wilson et al. 1977), uses these values in the context of a minor assumption about phylogeny to detect rate variation in the *ingroup* (the group of OTUs under study). One or more *outgroups* (taxa for which there is evidence of divergence prior to the most recent common ancestor of the ingroup OTUs) are first identified; similarity values are then determined for comparisons between the outgroup(s) and each of the ingroup OTUs. If rates of evolution have been more or less constant, then the values for all such comparisons should be roughly equivalent. On the other

hand, variations in evolutionary rates should be reflected in the comparisons: rapidly evolving lineages should exhibit relatively low similarity to the outgroup, and slowly evolving lineages should exhibit relatively high similarity. Although such comparisons will usually reveal some variation in rates, certain results indicate potential problems. If an OTU that clusters remotely in a phenetic analysis also exhibits relatively high rates of divergence, there is reason to suspect that phenetic clustering may be reconstructing its phylogenetic relationships incorrectly.

Michener and Sokal (1957) discussed a conceptually related method for detecting potentially problematic rate variations that evaluates the original similarity values in the context of the pregroup-exgroup problem. The pregroup versus exgroup status of an OTU bearing an isolated position relative to a cluster of OTUs may be evident in the original similarity values for comparisons between the isolated OTU and the various members of the cluster. Given that the OTUs within the cluster have evolved at roughly similar rates, these similarity values should exhibit different patterns depending on the phylogenetic relationships of the isolated OTU. If the isolated OTU represents a lineage that diverged early, then it should exhibit more or less equivalent amounts of similarity to all members of the cluster. On the other hand, if the isolated OTU represents a highly divergent lineage derived from within the cluster, then despite its relatively low similarity to all members of the cluster, it should be more similar to some of them than to others. Some variation in these comparisons is inevitable, and there is no straightforward test for establishing significance. Nevertheless, the results of alternative tree reconstruction methods (see below) can reveal potential problems. If the remotely clustering OTU exhibits its greatest similarity to the same OTU or OTUs to which alternative methods indicate that it is most closely related, there is reason to suspect that phenetic clustering may be reconstructing its phylogenetic relationships incorrectly.

A combination of the two methods discussed above can be used to test alternative reconstructions of phylogeny, such as those resulting from the use of different analytical methods (including clustering). Alternative

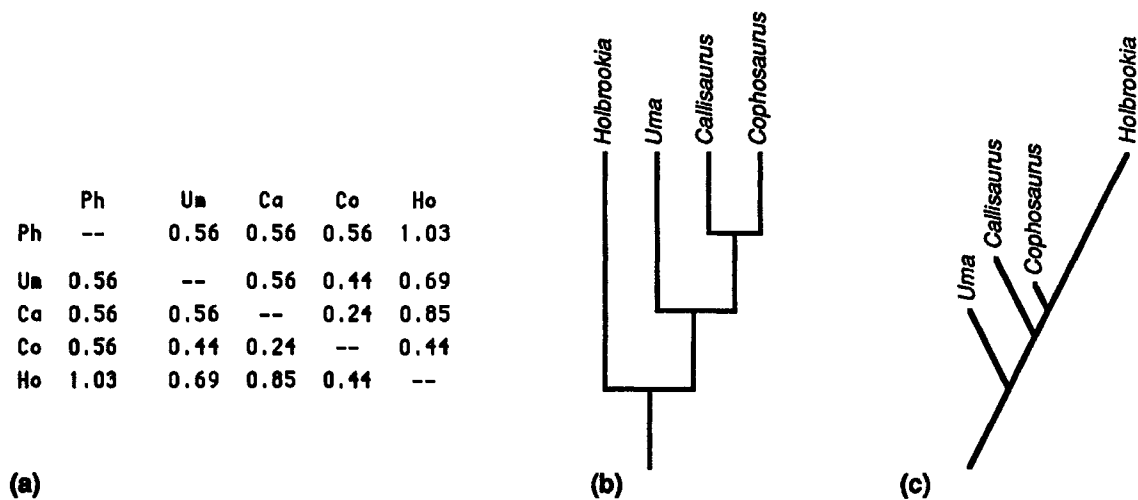


FIGURE 1. AN EXAMPLE IN WHICH VARIATION IN RATES OF EVOLUTION APPARENTLY CAUSES PHENETIC CLUSTERING TO YIELD AN INCORRECT PHYLOGENETIC TREE (AFTER DE QUEIROZ 1989, 1992).

(a) Genetic distances (coefficient of Nei 1972) between four ingroup OTUs (Um = *Uma*, Ca = *Callisaurus*, Co = *Cophosaurus*, Ho = *Holbrookia*) and one outgroup (Ph = *Phrynosoma*) based on data collected by Adest (1978) with the omission of 2 loci that were not scored in a second outgroup. (b) Ingroup phenogram based on UPGMA clustering of allozyme data converted to genetic distances. (c) Phylogeny suggested by neighbor-joining and parsimony analyses of allozyme data, parsimony analysis of morphological and behavioral characters, and patterns of genetic distance (relative rate tests and ingroup comparisons). The discrepancy is attributable to an increased rate of evolutionary divergence in the *Holbrookia* lineage. See text for details.

trees may yield different predictions concerning relative amounts of similarity for particular pairs of OTUs. These alternative hypotheses can then be evaluated by comparing the predictions derived from them with the observed similarity values. In short, the family of methods described above is useful not only for assessing variation in amounts of divergence among lineages but also for choosing among competing phylogenetic hypotheses. Although at least some of the methods have been known for a long time, they have not been used to full advantage. The following example serves to illustrate the practical use of the methods, as well as the existence of significant rate variation among lineages in real data, that is, variation of sufficient magnitude to cause problems for phenetic clustering as a method for estimating phylogeny.

Patterns of allozyme variation among species within a clade of lizards were studied by Adest (1978) and de Queiroz (1989, 1992). When converted to genetic distances (Figure 1a) and analysed using UPGMA clustering, the allele frequency data yielded a phenogram

in which one taxon, *Holbrookia*, was positioned outside of a cluster formed by all other members of the clade (Adest 1978; de Queiroz 1989) (Figure 1b). Comparisons with outgroup OTUs revealed a greater amount of divergence for the remotely positioned taxon than for the other ingroup taxa (de Queiroz 1992) (Figure 1a), suggesting that *Holbrookia* may be an exgroup. This conclusion is supported by the results of parsimony and neighbor-joining analyses of the same allozyme data as well as the results of parsimony analysis of morphological and behavioral characters (de Queiroz 1989, 1992), all of which indicate that *Holbrookia* is the highly divergent sister group of *Cophosaurus* (Figure 1c). Further comparisons involving the genetic distances between various pairs of ingroup OTUs support a close phylogenetic relationship between *Holbrookia* and *Cophosaurus*. Despite the relatively great distances between representatives of the divergent *Holbrookia* lineage and all other ingroup OTUs, the former exhibit smaller genetic distances to *Cophosaurus* than to the other ingroup OTUs (Figure 1a). In short, the genetic

similarity data themselves imply that *Holbrookia* represents a highly divergent lineage that shares a relatively recent common ancestor with *Cophosaurus* (Figure 1c). This conclusion contradicts the tree resulting from phenetic clustering and illustrates that variation in amounts of divergence among real lineages can be sufficiently great to violate the assumptions necessary for phenetic clustering to estimate phylogeny correctly.

Amounts of rate variation among lineages are expected to differ from one clade to another. If so, phenetic clustering may recover the correct phylogeny in some study groups even if it does not in others. Nevertheless, the existence of a single case demonstrating amounts of variation sufficient to cause problems for phenetic clustering implies that uncritical use of the method is not justified. At the very least, one should test for potentially problematical rate variations using one or more of the procedures described above. Alternatively, one can use methods of phylogeny reconstruction that do not rest on assumptions of rate uniformity.

Several such alternative methods have been developed (reviewed by Felsenstein 1988; Swofford and Olsen 1990). One of the simplest takes the phenogram as a starting point and then corrects for rate variations, which are detected by comparison of the original similarity values among pairs of OTUs in the context of the phenogram (Li 1981). The neighbor-joining method (Saitou and Nei 1987; Studier and Keppler 1988) also corrects the original similarity values for unequal amounts of divergence among lineages. It and several other distance matrix methods (e.g., Fitch and Margoliash 1967; Farris 1972) fit the original distances to the branches of an additive tree, and (unlike phenetic clustering) permit branches of equal temporal duration to differ in length. Some of these methods are algorithms that yield a single tree for a given distance matrix, while others choose among alternative trees based on an optimality criterion, the most common of which minimize the difference between the observed distances and those implied by the fitted branch lengths (e.g., Fitch and Margoliash 1967; Cavalli-Sforza and Edwards 1967; Farris 1972; Prager and Wilson 1978; Swofford 1981). Rate uniformity is not

assumed, but these methods assign length to terminal branches wherever possible (Swofford and Maddison 1987).

Other methods are based on parsimony, that is, minimizing the amount of evolutionary change required by the data (e.g., Edwards and Cavalli-Sforza 1964; Camin and Sokal 1965; Farris 1970, 1972; Farris et al. 1970; Swofford and Berlocher 1987), or compatibility, that is, maximizing the number of congruent characters (e.g., LeQuesne 1969, 1972, 1974; Estabrook et al. 1976a,b). Neither parsimony nor compatibility assume rate uniformity, but both can yield incorrect trees when long branches in a tree are separated by short ones, a situation that can result from certain patterns of rate inequality among lineages (Felsenstein 1978; Hendy and Penny 1989). Maximum likelihood methods (e.g., Edwards and Cavalli-Sforza 1964; Cavalli-Sforza and Edwards 1967; Felsenstein 1973, 1979) require an evolutionary model, which can be formulated specifically to include or to exclude assumptions of rate uniformity. Methods based on invariants (reviewed by Penny et al. 1992) use the data to estimate properties of the evolutionary process that generated them, including variation in rates. Some of these methods have been developed specifically to solve the problem of phylogeny reconstruction under unequal rates (Lake 1987a,b) and the attraction of long branches (Hendy and Penny 1989).

All of these methods have their own limitations (reviewed by Felsenstein 1982a; Swofford and Olson 1990; Penny et al. 1992), but the results of simulation studies indicate that most perform better than phenetic clustering, particularly when rates are variable among lineages (Tateno et al. 1982; Sourdis and Krimbas 1987; Rohlf et al. 1990; Jin and Nei 1991; DeBry 1992; Kim et al. 1993; Huelsenbeck and Hillis 1993; Huelsenbeck 1995). This fact is presumably related to an important respect in which most alternative phylogenetic methods differ from phenetic clustering. Whether they use character or similarity/distance data, most alternative methods analyse the data as changes along the branches of phylogenetic trees. For character data, transformations from one state to another are assigned to the various branches of the tree in order to ac-

count for the states present at the branch ends. For distance data, the distances between taxa at the branch ends are partitioned among the branches connecting them. In both cases, the branches of the trees have lengths that represent amounts of evolutionary change. Such methods thus conform with a general phylogenetic branching model, or what O'Hara (1988) calls "tree thinking." In contrast, phenetic clustering does not directly analyse similarity data as having been produced in the context of a branching phylogeny, but instead forms hierarchical nonoverlapping clusters strictly on the basis of similarity. These clusters need not be represented as trees, and when they are, the lengths of the branches indicate (strictly speaking) not amounts of evolutionary change but only the level of similarity at which certain OTUs form clusters. Because clusters are formed without regard for how the similarities may have been produced evolutionarily, interpretation of the branch lengths (clustering levels) as amounts of evolutionary change carries with it an implicit and dubious assumption of uniform rates. Given the existence of alternative methods based on a more appropriate theoretical context, and whose performance is generally less sensitive to rate variation among lineages, there seems to be little reason for the continued use of phenetic clustering as a method for estimating phylogenetic relationships.

PHENETIC CLUSTERING AND STUDIES OF INTRASPECIFIC VARIATION

Phenetic clustering is also used to analyse relationships among potentially conspecific populations, with frequent application to the issues of geographic variation and genetic continuity, and consequently, of species limits (e.g., Highton 1989). In contrast with its use in phylogeny reconstruction, clustering is still one of the most popular methods in such studies, but its limitations in this area are less well known. However, just as in the case of phylogeny reconstruction, there are serious limitations to the use of phenetic clustering for analysing geographic variation and genetic continuity. The two problems are closely analogous. As a method of phylogeny reconstruction, the accuracy of clustering by similarity is limited because of the potential noncorre-

spondence between similarity and recency of common ancestry. Likewise, as a method for analysing species limits, the usefulness of phenetic clustering is limited because of the potential noncorrespondence between similarity and degree of genetic continuity.

The problem has often been stated in terms of the nonconformity of relationships among populations to a general nested hierarchical pattern. For example, both the degree of genetic continuity in the form of gene flow among populations and the patterns of phenotypic and genetic similarity resulting from it are unlikely to form nested hierarchies except under unusual circumstances (e.g., Felsenstein 1982b). These kinds of relationships may not even exhibit discrete clusters, and individual phenotypic and genotypic characters (whether ancestral, derived, or both) may characterize overlapping groups of organisms or populations, rather than nested or mutually exclusive ones (e.g., Sokal 1983b; for an analogous conclusion concerning biogeographical analyses see Hengeveld 1990).

In the case of relationships that do not intrinsically possess a nested hierarchical structure, the use of phenetic clustering and other analytical methods (e.g., cladistic analysis) that are constrained to yield results in the form of nested hierarchies is problematical. Because these methods will almost inevitably produce a nested hierarchy of groups (Sneath and Sokal 1973:252; Dunn and Everitt 1982:94), applying them to most real data is likely to yield results in this form even if the relationships of interest are not intrinsically hierarchical. For this reason, several authors have attempted to analyse intraspecific relationships using ordination methods, including principal component and principal coordinate analysis (e.g., Menozzi et al. 1978; Rendine et al. 1986; Majumder 1988; Sanchez-Mazas and Langaney 1988) and multidimensional scaling (Baker and Moeed 1987; Sokal et al. 1987; Derish and Sokal 1988; Lessa 1990), all of which summarize similarity data—genetic or otherwise—without imposing a nested hierarchy.

Ordination methods may be preferable to clustering for summarizing nonhierarchical patterns of similarity among populations (e.g., Jardine 1969; Sneath and Sokal 1973:367–368; Sokal 1983b, 1985; Thorpe 1983; Lessa 1990),

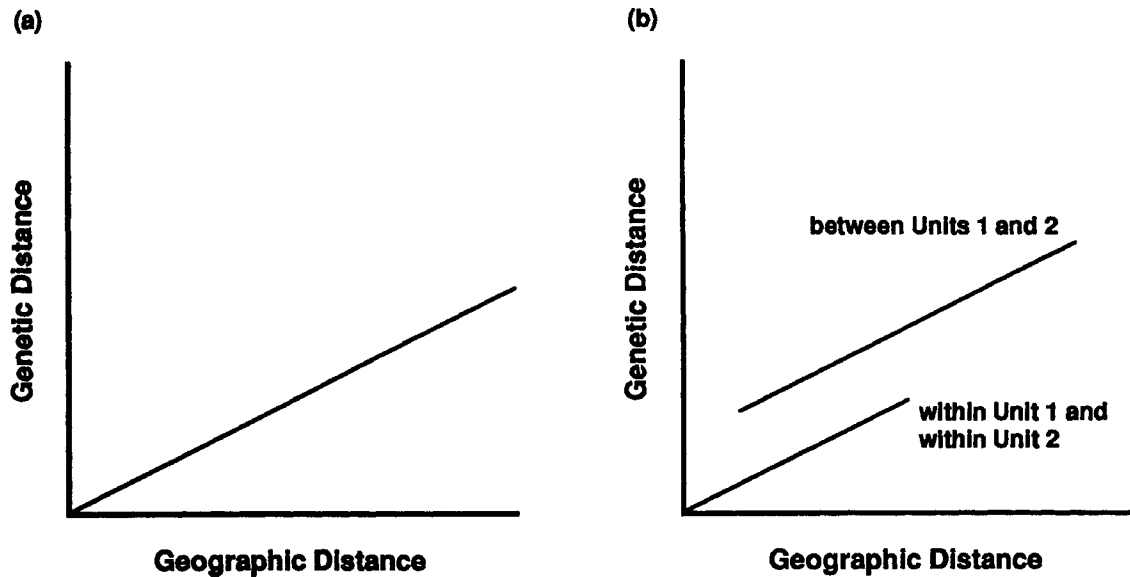


FIGURE 2. EXPECTED RELATIONSHIP BETWEEN GEOGRAPHIC DISTANCE AND GENETIC DISTANCE FOR GENETICALLY CONTINUOUS AND DISCONTINUOUS GROUPS OF POPULATIONS UNDER AN ISOLATION BY DISTANCE MODEL.

(a) Plot of genetic distance versus geographic distance for populations within a single genetically continuous unit with a linear geographic distribution. (b) Plot of genetic distance versus geographic distance for populations separated into two units by a genetic discontinuity under a vicariance model; each unit has a linear geographic distribution and together the two units are also distributed linearly; displacement of the line is proportional to the length of time since genetic isolation.

but populations presumably do form nonoverlapping groups delimited by the presence or absence of genetic continuity. Because ordination methods do not explicitly produce groups, it is still tempting to use clustering methods to identify groups. The problem is that groups of genetically similar populations do not necessarily correspond with groups of genetically continuous populations, and this can lead to misinterpretations if clustering methods are applied uncritically.

This problem is illustrated by considering the expected relationship between the genetic similarity of populations and their geographic separation in the presence versus absence of genetic continuity (Good and Wake 1992) (Figure 2). Within a genetically continuous group of populations, the closer two populations are in space, the more easily alleles can be exchanged between them. Therefore, provided that gene flow is sufficiently low to permit differentiation, and provided that mutation and migration rates are at equilibrium,

there should be a positive relationship between the genetic dissimilarity exhibited by any two populations and the geographic distance separating them (Nei 1972), that is, isolation-by-distance (Wright 1943, 1946, 1969). Within such a group, a regression line for all pairwise comparisons among the populations should have a positive slope and should intersect the origin (Figure 2a). Under such circumstances, relatively great genetic differences can develop between geographically distant populations within the same genetically continuous unit.

In contrast, populations that are not parts of the same genetically continuous unit exchange no alleles, regardless of their spatial relationships. Therefore, the genetic differentiation of populations in groups separated by a genetic discontinuity should be unrelated to the geographic distance between them. In other words, such populations should diverge at the same rate (all else being equal) regardless of whether they are close together or far

apart. Provided that there has been sufficient time for genetic differentiation since the separation of two such genetic units, a regression line for comparisons between populations in separate units should fail to pass through the origin (Figure 2b). Displacement of the line will increase as genetic differences between populations in separate units accumulate, but it will be slight in the initial stages of divergence. Under such circumstances, even relatively small genetic differences may be indicative of genetic discontinuity if the populations in question are sufficiently close geographically.

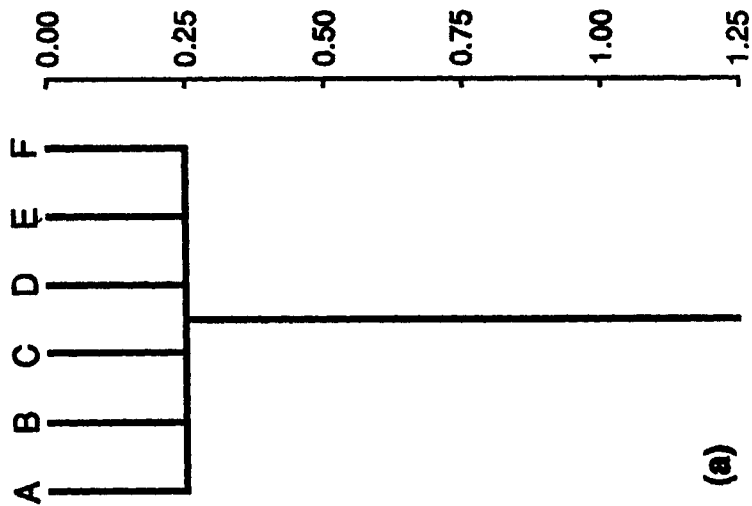
The absence of a simple relationship between similarity and continuity complicates the interpretation of results obtained using phenetic clustering, a situation that is further complicated by the inevitable discontinuous sampling of genetically continuous populations. Combined with the strong tendency of phenetic clustering to form groups, these facts can lead to the formation of groups that are artifacts of nonuniform sampling or minor sampling errors rather than reflections of intrinsic species organization. Consider a hypothetical situation (Figure 3) in which the correlation between geographic and genetic distance is perfect, sampling is absolutely regular, and estimates of genetic distance are without error. In such an ideal case, phenetic clustering will not produce any groups (Figure 3a). However, if populations are sampled in such a way that they happen to be located toward opposite ends of the geographic range of the species, phenetic clustering is likely to form two primary groups corresponding with the geographically separated samples (Figure 3b). The greater the geographic gap between the samples—which might result solely from the vagaries of field collecting—the more distinct the groups will appear (Figure 3c). In fact, no such division exists in the genetical organization of the species; it is an artifact of uneven geographic sampling and the inherent properties of phenetic clustering. Thus, in the case of geographically differentiated populations, the existence of phenetic clusters, even seemingly distinct ones, is not necessarily indicative of intrinsic species organization, that is, genetic discontinuity.

These problems are illustrated by a study of

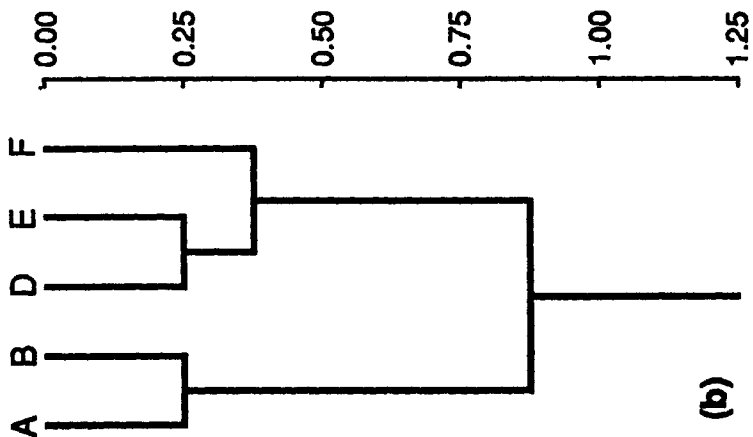
geographic variation in the salamander *Ambystoma rosaceum* in the Sierra Madre Occidental of northwestern Mexico (Figure 4a). Shaffer (1983) analysed allozyme data for populations of *A. rosaceum* using phenetic clustering (Figure 4b) and the method of Fitch and Margoliash (1967), another method constrained to produce results in the form of a tree. Based on the results of these analyses, he suggested that the populations of *A. rosaceum* formed northern and southern units, with little or no gene flow between them. However, the distribution of sample populations was uneven, and the division between the two putative forms identified on the basis of phenetic clustering (Figure 4b) corresponds with a large geographic gap in sampling (Figure 4a). Analysing these genetic distance data in the context of the geographic distances between the various pairs of populations reveals that the apparent existence of distinct "forms" may be attributable solely to the pattern of geographic sampling. Given the relationship between geographic and genetic distance seen among the populations *within* the northern and southern "forms," the level of genetic differentiation across the wide geographic gap *between* those "forms" corresponds closely with that predicted for populations exhibiting isolation-by-distance within a single genetically continuous unit (Figure 4c).

Several alternatives to clustering have been developed to examine the relationship between genetic (or any other kind of) variation and the geographic dispersion of populations. Mantel's (1967) statistical test for comparing matrices has been used to compare genetic and geographic distances for the same pairs of populations (Sokal 1979). Mantel's test might be used to distinguish between genetically cohesive sets of populations (among which there presumably would be a significant correlation between genetic and geographic distance) and sets of populations that do not exchange genes (among which such a correlation should not be significant). Spatial autocorrelation (Cliff and Ord 1973), the correlation between values of variables at locations different distances apart, has been used to measure nonrandomness in the spatial distributions of individual genotypes (e.g., Sokal and Oden 1978a,b; Sokal 1979; Sokal and Wartenberg

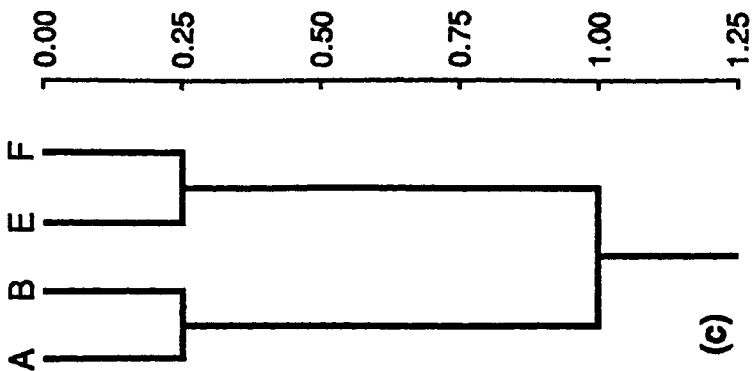
	A	B	C	D	E	F
A	--					
B	0.25	--				
C	0.50	0.25	--			
D	0.75	0.50	0.25	--		
E	1.00	0.75	0.50	0.25	--	
F	1.25	1.00	0.75	0.50	0.25	--



	A	B	D	E	F
A	--				
B	0.25	--			
D	0.75	0.50	--		
E	1.00	0.75	0.25	--	
F	1.25	1.00	0.50	0.25	--



	A	B	E	F
A	--			
B	0.25	--		
E	1.00	0.75	--	
F	1.25	1.00	0.25	--



1983; Slatkin and Arter 1991; Epperson 1993). This method should be able to distinguish between genetically cohesive units (a gradual decline with increasing distance in spatial autocorrelation values) and noncohesive units (no such trend in correlations). In both Mantel's test and spatial autocorrelation, the units (sets of populations) would have to be identified by some other criterion prior to analysis.

The method of Gabriel and Sokal (1969) attempts to partition a network of connected populations into units, which may be either mutually exclusive or overlapping. This is a statistical method for categorizing sets of populations on the basis of (genetic or phenotypic) similarity, but it does not take into consideration causal processes. Populations that exchange genes should be tightly "connected," but connections will by no means be limited to such populations. An approach termed "phylogeography" (Avice et al. 1987) examines the phylogeny of alleles or haplotypes in the context of their geographic distribution. Although the sharing of alleles from separate branches in a gene tree by the members of two or more geographically delimited sets of populations could result either from gene flow or from lack of sufficient time for random extinction of gene lineages, geographic separation of such alleles would seem to indicate genetic discontinuity, especially if a similar geographic pattern occurred in several independently assorting genes (Avice and Ball 1990; Baum and Shaw 1995).

Explicit methods have also been developed

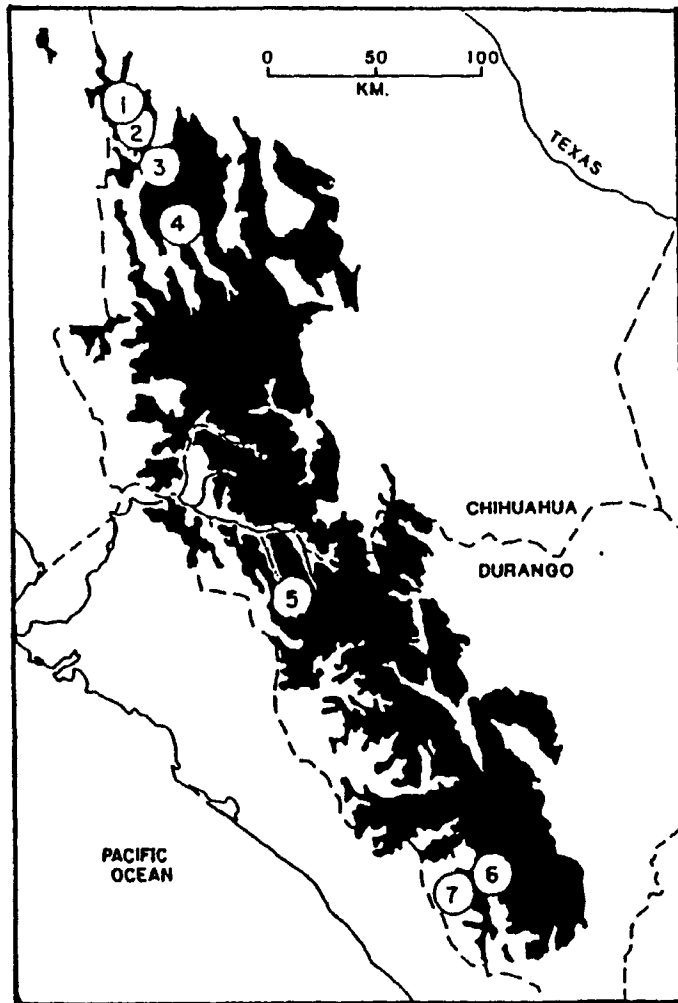
to measure gene flow, most of which assume rather than test genetic continuity. Early attempts to measure gene flow indirectly were based on the frequency of allelic lethals (reviewed by Crow and Temin 1964; Wallace 1966). Several other methods based on Wright's (1951) coefficient of the component of genetic variation due to population subdivision (F_{ST}) have been discussed at length (e.g., Wright 1978; Weir and Cockerham 1984; Slatkin 1985a, 1987; Nei 1987; Slatkin and Barton 1989). Other proposed procedures involve maximum likelihood methods (Barton et al. 1983; Wehrhahn and Powell 1987), the analysis of the distribution of rare alleles (Slatkin 1981, 1985b, 1987; Barton and Slatkin 1986), and the analysis of coalescence in gene trees (Slatkin and Maddison 1989, 1990).

Although many methods may have difficulty detecting recent genetic discontinuities (Larson et al. 1984), this is a minor problem compared with those resulting from the simple comparison of populations in terms of their genetic similarity, as in phenetic clustering. The preceding analysis suggests that such comparisons may reveal little about the presence or absence and extent of genetic continuity. Without considering the geographic context of the comparisons, large differences cannot necessarily be taken as evidence for genetic discontinuity, nor can small ones be taken as evidence for genetic continuity. It is even possible for differences between geographically distant populations within a single genetically continuous unit to be greater than

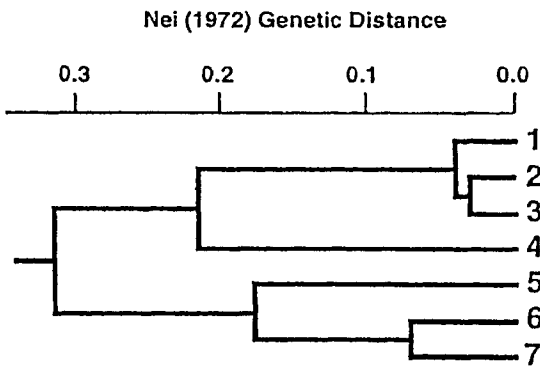
FIGURE 3. CLUSTERS THAT ARE ARTIFACTS OF UNEVEN GEOGRAPHIC SAMPLING RATHER THAN INTRINSIC SPECIES ORGANIZATION.

(a) Genetic distance matrix (above) and UVGMA (V = Variable) phenogram (below) for populations sampled at regular intervals from a genetically continuous series of populations in which the correlation between geographic and genetic distance is perfect, sampling is absolutely regular, and genetic distances are estimated without error; there are seven alternative UPGMA phenograms for the same data, the consensus of which has the same topology as the UVGMA phenogram. In this ideal case, the phenogram is unresolved. (b) Genetic distance matrix (above) and one of two equivalent UPGMA phenograms (below) for the same series of populations with a sampling gap in the middle of the range (population C not sampled). (c) Genetic distance matrix (above) and UPGMA phenogram (below) for the same series of populations with a larger sampling gap (populations C and D not sampled). Although there is no genetic discontinuity within the species, the sampling gap results in an apparent discontinuity in the form of distinct clusters in the phenogram. The apparent distinctness of the clusters, indicated by the clustering level on the phenogram, is exaggerated by an increase in the size of the sampling gap. Note that even with absolutely regular sampling, a less than perfect correlation between geographic and genetic distance or errors in the estimation of genetic distance can lead to the production of clusters.

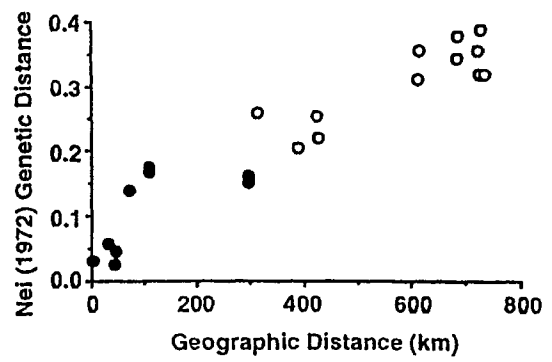
(a)



(b)



(c)



those between populations in separate genetic units. Consequently, there is little to be gained by analysing the data with phenetic clustering. In addition to imposing an inappropriate hierarchical structure on the results, clustering is sensitive to uneven sampling and—because it does not incorporate geographic information (but see Legendre 1987)—has no direct bearing on the question of genetic continuity. Other methods, particularly those incorporating geographic information, use the information inherent in the pattern of geographic sampling (whether even or uneven) and yield results that bear directly on the issue of genetic continuity.

PHENETIC CLUSTERING AND THE GENERAL REPRESENTATION OF SIMILARITY

The limitations of phenetic clustering both as a method for analysing phylogenetic relationships and as a method for assessing genetic continuity among populations might have been anticipated; the procedure was borrowed—rather than being specifically designed—for these applications. Phenetic clustering was designed, and consequently its results are to be strictly interpreted, simply as a method for analysing or summarizing the similarity relationships among OTUs under a nested hierarchical model (e.g., Michener and Sokal 1957; Sokal and Michener 1958; Sneath and Sokal 1962, 1973; Sokal 1963; Sokal and Sneath 1963). In many cases—including applications in taxonomy, ecology, biogeography, and psychology—phenetic clustering is used only in this general way, that is, as a method for analysing or summarizing the similarities and differences among entities regardless of the specific biological processes that may have produced them. We will now argue that even in this very

general application, phenetic clustering has serious limitations.

The examples used to illustrate problems with the use of phenetic clustering in studies of geographic variation (Figures 3 and 4) demonstrate that clustering can give misleading indications of nested hierarchical organization. Nevertheless, in those examples the similarity data themselves exhibit a nested hierarchical component. For example, the phenograms in Figures 3b and 3c are not misleading in indicating that OTUs A and B are more similar to one another than either is to OTUs E and F, and vice versa. But even if one is concerned only with the representation of similarity relationships as they exist in the data—that is, without concern for their cause or interpretation—phenetic clustering is a poor choice as an analytical method. The reason is that, despite containing some nested hierarchical components, similarity relationships rarely exhibit a strictly nested hierarchical structure. Consequently, these relationships are inevitably distorted, or at least oversimplified, by methods that are constrained to yield results in the form of nested hierarchies.

Relationships exhibiting a strictly nested hierarchical structure must satisfy two conditions for statements taking the general form *A is more closely related to B than it is to C*. First, the relation must be symmetric with respect to the terms A and B, that is, interchanging the terms must not alter the truth of the statement. Second, the converse statement concerning the relationships of C must specify an equal relationship to A and B. If these conditions hold, then the relationships can be accurately and entirely represented by a hierarchy of nested groups.

Phenetic relationships rarely satisfy either

FIGURE 4. AN EXAMPLE IN WHICH UNEVEN GEOGRAPHIC SAMPLING APPARENTLY CAUSES PHENETIC CLUSTERING TO YIELD MISLEADING RESULTS.

(a) Distribution of *Ambystoma rosaceum* populations sampled by Shaffer (1983) in a study of allozyme variation (modified from his Figure 1); shaded area represents land above 2,000 m elevation. (b) Phenogram resulting from UPGMA analysis of Nei (1972) genetic distances, upon which basis the populations were divided into northern (OTUs 1–4) and southern (OTUs 5–7) groups. Notice that the division between these two “groups” corresponds with the largest geographic gap in sampling. (c) Plot of Nei genetic distance against geographic distance for all pairs of populations; solid circles represent pairwise comparisons between populations within each of the two “groups”; open circles represent pairwise comparisons between populations in northern versus southern “groups.” The data conform well to the expectations for a single genetically continuous unit within which there is isolation-by-distance.

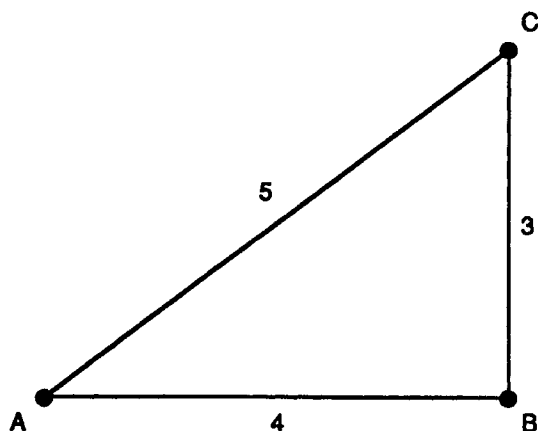


FIGURE 5. PHENETIC (SIMILARITY) RELATIONSHIPS, ILLUSTRATED AS DISTANCES IN TWO DIMENSIONAL CHARACTER SPACE, DO NOT CONFORM TO THE PATTERN OF A STRICT NESTED HIERARCHY.

See text for explanation.

of these conditions. Given that A is more similar to B than to C, it does not necessarily follow, first, that B is more similar to A than to C, or second, that A and B are equally similar to C. These conclusions should be evident from the following example in which dissimilarity is conceptualized as a Euclidean distance in a character space of two dimensions. Suppose that three OTUs are situated in this character space such that the shortest paths connecting them form a right triangle, with B at the right angle and A at the end of the longer leg (Figure 5). Although the distance from A to B is smaller than that from A to C, the distance from B to A is larger than that from B to C. Furthermore, although the distance between B and C is smaller than that from either B or C to A, the distances from B to A and C to A are not of equal magnitude.

That similarity does not exhibit a strictly nested hierarchical structure has important consequences for its analysis using phenetic clustering. Consider once again OTUs distributed in a multidimensional character space with dissimilarity measured as the distance between them. If OTUs are distributed uniformly in this phenetic space, no nested hierarchy exists. Although groups can be delimited artificially to form a nested hierarchy, no such

structure is inherent in the data. Of course, many clustering methods will not yield nested clusters when applied to a uniform distribution of OTUs. Only one large unresolved cluster will be recognized, or there will be multiple alternative trees (e.g., Hart 1983), the consensus of which is unresolved. Nevertheless, even the slightest departure from uniformity will change the situation, and even random distributions of OTUs will result in the formation of clusters (Sneath and Sokal 1973:252; Dunn and Everitt 1982:94).

Given that the application of phenetic clustering methods will almost inevitably yield clusters, it is instructive to consider the meaning of those clusters and their relationship to the intrinsic structure of the data. Although similarities rarely conform to the pattern of a simple nested hierarchy, they often contain nested, hierarchical components. That is to say, it is often possible to identify some groups of OTUs that are nested and mutually exclusive in terms of similarity. Continuing with the previous example (Figure 5), OTUs B and C are more similar to one another than either is to A, which means that B and C form an exclusive group in terms of similarity. And if all other OTUs exhibit distances to A, B, and C that are greater than 5 units, then group BC is nested within another exclusive similarity group, ABC. The groups produced by phenetic clustering, however, do not necessarily correspond with such exclusive similarity groups. Here it is important to distinguish between mutually exclusive groups that are intrinsic to the data and those that result from applying a particular analytical method. Just because groups are mutually exclusive in terms of membership as determined by some clustering procedure does not mean that they are also mutually exclusive in terms of phenetic relationships. As it turns out, several of the most popular phenetic clustering methods can produce groups that are not mutually exclusive in terms of the phenetic relationships of their member OTUs.

The production of such groups is not entirely unexpected, for the clustering procedures exhibiting this property were not designed to find exclusive similarity groups. Instead, each method forms groups according to a specific mathematical criterion and pro-

TABLE 1
Group formation criteria and group properties for three of the most popular classes of clustering methods¹

Method	Clustering Criterion	Similarity Properties of Clusters
Single linkage (nearest neighbor, minimum method)	Minimum distance between most similar members of different clusters.	The degree of similarity between any member of a given cluster and the most similar member of that cluster (nearest neighbor) is greater than or equal to that between any member of the cluster and any nonmember OTU.
Complete linkage (furthest neighbor, maximum method)	Minimum distance between least similar members of different clusters.	The degree of similarity between any member of a given cluster and the least similar member of that cluster (furthest neighbor) is greater than or equal to that between any member of the cluster and the least similar OTU within any other cluster.
Average linkage (including UPGMA, WPGMA, UPGMC, and WPGMC) ²	Minimum average distance ³ between members of different clusters.	The average degree of similarity ³ between the (two) primary subclusters of a given cluster is greater than or equal to the average degree of similarity between either of those subclusters and any other cluster.

¹ An isolated OTU is considered to be the lone member of a cluster.

² UPGMA = unweighted pair-group method using arithmetic averages; WPGMA = weighted pair-group method using arithmetic averages; UPGMC = unweighted pair-group method using centroids; and WPGMC = weighted pair-group method using centroids.

³ Determined either by weighted or unweighted averaging of the similarity values between either the OTUs in the different subclusters or their centroids.

duces groups whose properties reflect its particular criterion. The grouping criteria and similarity properties of the groups formed by several of the best known phenetic clustering methods are given in Table 1. Notice that none of the group-forming criteria used in these methods rules out the possibility of non-exclusive clusters, that is, the possibility that an OTU within a given cluster is more similar to certain OTUs outside of that cluster than to other OTUs within the cluster. In single linkage clustering, an OTU can be more similar to members of another cluster than to all but the most similar member of its own cluster. In complete linkage clustering, an OTU can be more similar to some members of another cluster than to the least similar member of its own cluster, as long as it is even less similar to the least similar member of the other cluster. And in average linkage clustering, an OTU can be more similar to some members of another cluster than to some members of its own cluster, as long as its average similarity to the members of its own cluster is greater than its average similarity to the members of the other cluster. Thus, all of the methods can produce mutually exclusive groups that do not exist in the raw similarity data.

Examples of this phenomenon are given in

Figures 6 and 7. Ten OTUs were assigned randomly to cells in a 10×10 matrix in two-character space (Figure 6a), and the dissimilarities for all pairs were calculated using the Euclidean distance coefficient (Figure 6b). When these data were analysed using methods from each of the three major classes of clustering procedures in Table 1 (Figure 7a-c), in every case groups resulted that are not mutually exclusive in terms of the observed similarities (Figure 6). Consider the seven clusters resulting from application of the UPGMA (Figure 7c). Of those seven clusters, four (ABCD, ABC, EFGHIJ, and FGHIJ) are not exclusive in terms of similarity. For example, OTU C is no more similar to OTUs A and D, with which it forms cluster ABCD, than it is to OTU G, which is not part of that cluster ($D = 3.0$ for all comparisons). OTU E is more similar to B ($D = 5.7$), with which it clusters only at the level of all 10 OTUs, than it is to G ($D = 6.0$), with which it forms cluster EFGHIJ. And OTU H is less similar to F ($D = 5.0$) and G ($D = 6.0$), with which it forms cluster FGHIJ, than to nonmember E ($D = 2.8$). Distortion of the original similarity relationships by clustering procedures is a well-known phenomenon, and various methods have been developed to summarize or illustrate the magnitude of distor-

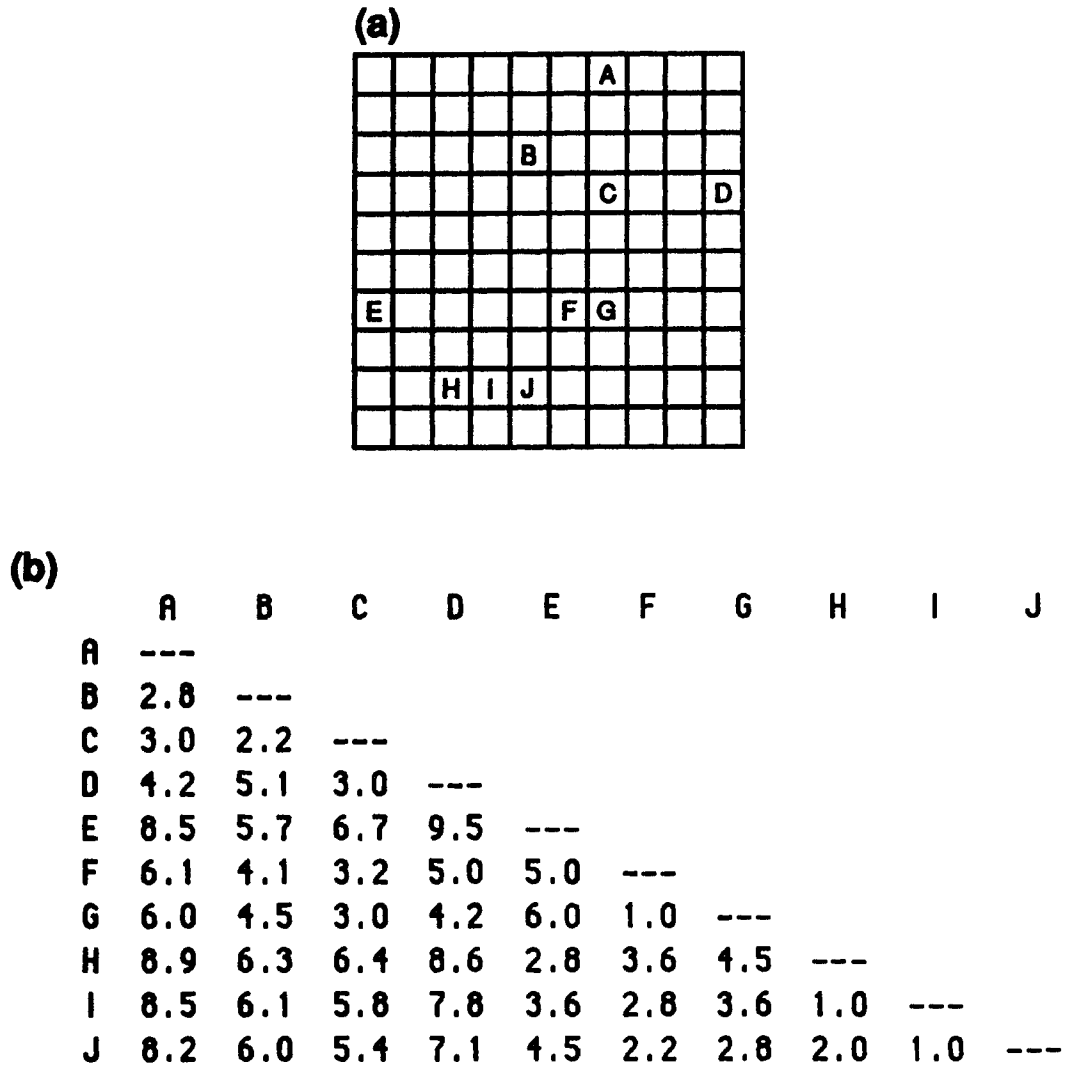


FIGURE 6. RANDOM DATA USED TO ILLUSTRATE THE DISTORTION OF PHENETIC RELATIONSHIPS BY PHENETIC CLUSTERING.

(a) Ten OTUs (A-J) assigned randomly to cells in two-character space. (b) Euclidean distance matrix for OTUs illustrated in (a).

tion (e.g., Jardine and Sibson 1968b; Hartigan 1967; Rohlf 1970). Perhaps the best known method is that of cophenetic correlations (Sokal and Rohlf 1962; Farris 1969).

Despite these problems, most real data—and even randomly generated artificial data—contain some mutually exclusive similarity groups (e.g., groups BC, FG, and HIJ in Figure 6), and these groups will often be recognized in the results of phenetic clustering (Figure 7a-c). Furthermore, methods have been devised that use mutual exclusivity with respect

to the degree of similarity as their criterion for forming clusters (e.g., McQuitty 1963, 1965, 1967; Jardine et al. 1969) so that groups are formed *if and only if* all of their members are more similar to one another than to all non-member OTUs (Figure 7d). Such methods represent only those mutually exclusive similarity groups intrinsic to the data; in other words, they do not impose clusters where clusters do not exist.

Nevertheless, even clustering methods that form only mutually exclusive similarity groups

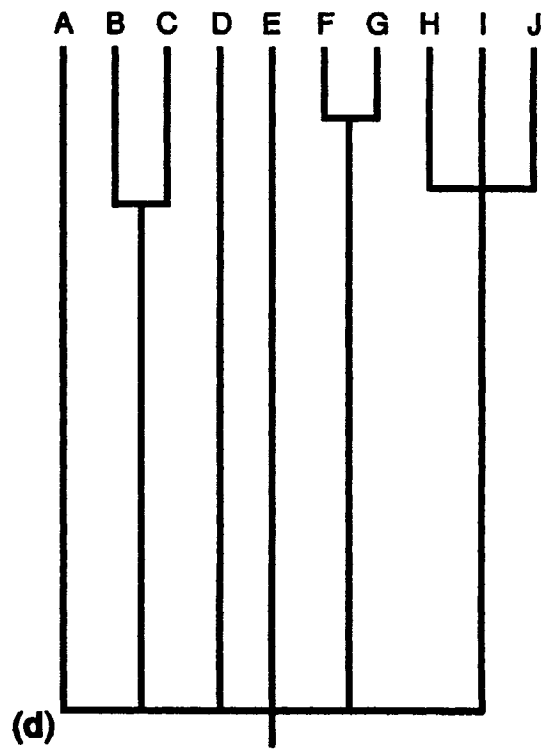
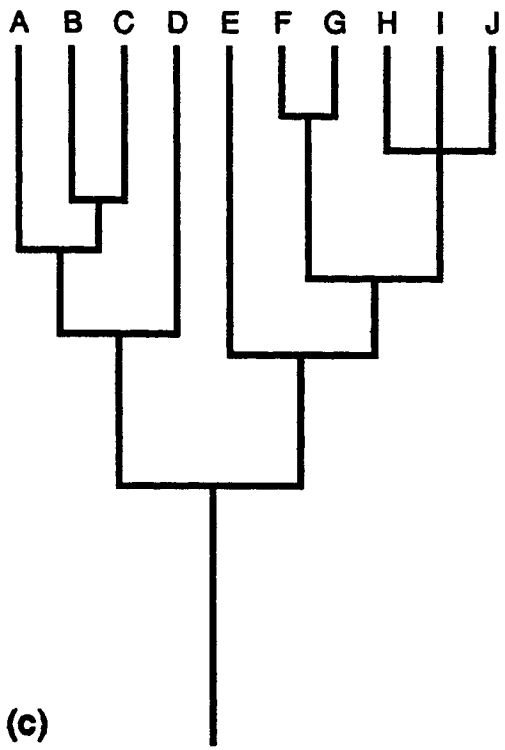
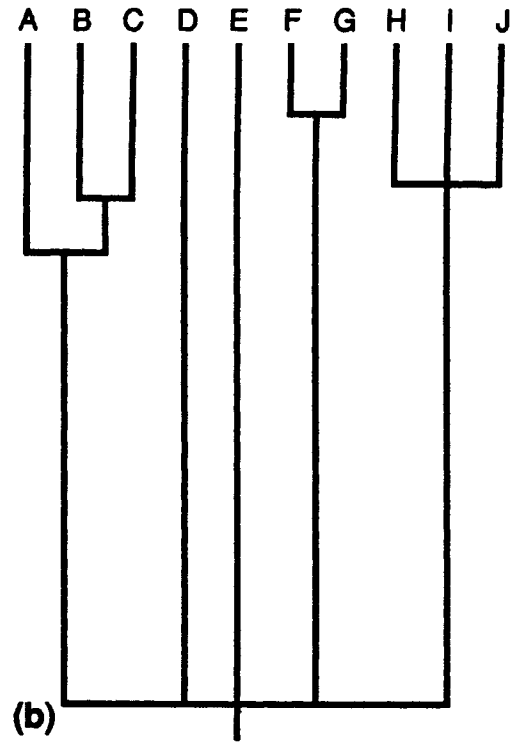
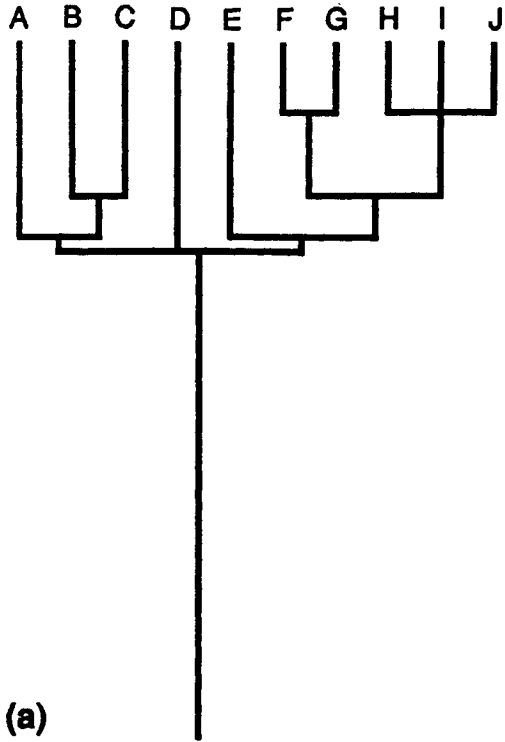
do not necessarily represent similarity relationships well. Various aspects of phenetic relationships often cannot be captured by a set of nested mutually exclusive groups, because the members of exclusive similarity groups are not necessarily equally similar either to one another or to nonmember OTUs. For example, OTUs H, I, and J (Figure 6) form an exclusive group, but the dendrograms (Figure 7a-d) do not capture either the greater similarity between J and I than between J and H, or the greater similarity between H and E than between either I or J and E. Therefore, although hierarchical nonoverlapping clusters can be used to represent mutually exclusive similarity groups, that is, the nested hierarchical component of phenetic relationships, they will often fail to represent other aspects of the more complex totality of phenetic relationships.

Thus, in addition to specific problems with the use of phenetic clustering in analyses of phylogenetic relationships and variation among populations, this class of methods is not particularly appropriate for analysing phenetic (similarity) relationships themselves. Phenetic relationships are too complex to be adequately represented by methods limited to producing results in the form of hierarchical nonoverlapping clusters. These complex relationships are more appropriately analysed using: clustering methods that permit overlapping groups (e.g., Jardine and Sibson 1968a,b; Cole and Wishart 1970), undirected graphs—such as minimum spanning trees (e.g., Kruskal 1956; Prim 1957)—which connect OTUs without imposing a hierarchy, or ordination methods such as principal components analysis (e.g., Hotelling 1933; Harman 1976), multiple factor analysis (e.g., Harman 1960; Rohlf and Sokal 1962), and multidimensional scaling (Sheppard 1962a,b, 1966; Kruskal 1964a,b), which array OTUs continuously in multidimensional character space rather than grouping them into discrete clusters. Graphs and ordinations are particularly effective in combination (Gower and Ross 1969; Rohlf 1970). The existence and sophistication of alternative methods leaves little reason for continuing to use hierarchical nonoverlapping clustering methods for analysing relationships that conform poorly to this kind of structure.

PHENETIC CLUSTERING IN HISTORICAL CONTEXT

Several of the limitations of phenetic clustering have been known for a long time. For example, problems concerning the use of clustering as a method of phylogeny reconstruction were discussed in the late 1950s (e.g., Michener and Sokal 1957), the distortion of similarity relationships was addressed in the early 1960s (e.g., Sokal and Rohlf 1962), and problems related to the study of intraspecific variation were noted by the late 1960s (e.g., Jardine 1969). By the early 1970s, pheneticists were clearly aware both that clustering tended to yield clusters regardless of the data and that other methods often represented similarity relationships more accurately than did clustering (Sneath and Sokal 1973; see below). In this light, the widespread use of phenetic clustering in biology appears incongruous. On further consideration, however, the situation is understandable as the result of a constraint imposed by the historical and disciplinary context within which the methods were developed.

In the biological sciences, phenetic clustering methods were developed largely within the context of taxonomy—specifically, within the context of a taxonomic reform movement known as *numerical taxonomy* or *phenetics* (for early papers see Michener and Sokal 1957; Sneath 1957a,b; Cain and Harrison 1958; Sokal and Michener 1958; for reviews see Sneath 1962; Sneath and Sokal 1962, 1973; Sokal and Sneath 1963; Sokal 1963, 1986; for historical reviews see Hull 1988; Vernon 1988). Although the participants in this movement were interested in the application of quantitative methods to diverse biological problems, including phylogeny reconstruction (e.g., Camin and Sokal 1965; Sokal 1983a), shape analysis (e.g., Sneath 1967; Rohlf 1990; Rohlf and Bookstein 1990), and biostatistics (e.g., Sokal and Rohlf 1969, 1981), their interest in clustering methods was tied strongly to taxonomy. Phenetics was a reform movement that arose in response to a dissatisfaction with the subjective and nonquantitative nature of taxonomy as it existed in the late 1950s (see above references), and phenetic clustering methods were developed by the participants in that movement as part of the solution to the per-



ceived problem. Because these scientists held certain tenets that seem to account for their advocacy of phenetic clustering, we will classify them as "pheneticists" on that basis. This label is used strictly for the purpose of explaining a particular phenomenon in the history of taxonomy, and is not intended to represent the whole character of any scientist's contributions or beliefs.

As a taxonomic movement, phenetics was constrained by the traditions of biological taxonomy, in particular, the Linnean conventions that had been accepted as the basis of that discipline for nearly two centuries. Along with binomial species names, the most fundamental convention of the Linnean taxonomic system is its hierarchy of taxonomic categories—from kingdom to species—used to convey the relative positions (ranks) of taxa in nested hierarchies. By and large, pheneticists accepted the taxonomic convention of nested hierarchies:

In order to evaluate the role in biological taxonomy of the methods [for measuring dissimilarity and for clustering] we must first examine the kind of classificatory system which biologists use, and the aims and methods of orthodox taxonomy.

The taxonomic or Linnaean hierarchy is an ordinal stratified hierarchic clustering (Jardine and Sibson 1971:127).

The classical tendency in biological systematics has been hierarchic classification . . . we are so obedient to the Linnean system, which requires mutually exclusive and hierarchically ordered classes, that the process of classification has become synonymous . . . with a mapping of the diversity of nature into the Linnean system (Sneath and Sokal 1973:200).

Concerning the underlying basis for the nested hierarchy of taxa, the phenetic movement was characterized by a particular philosophical position. Despite an interest in phy-

logeny, pheneticists argued that biological taxonomy should not be based on inferred phylogenetic relationships but on observed similarity, without regard to its cause:

A basic attitude of numerical taxonomists is the strict separation of phylogenetic speculation from taxonomic procedure. Taxonomic relationships are evaluated purely on the basis of the resemblances existing *now* in the material at hand. These phenetic relationships do not take into account the origin of the resemblance found nor the rate at which resemblances may have increased or decreased in the past (Sneath and Sokal 1973:9; emphasis in original).

Similar statements are widespread in the phenetic literature (e.g., Cain and Harrison 1958; Sneath and Sokal 1962; Sokal and Sneath 1963; Heywood 1964; Sokal 1985, 1986). Phenetic relationships were conceptualized as overall similarity expressed as a continuous, quantitative variable (Sokal and Sneath 1963; Jardine and Sibson 1971; Sneath and Sokal 1973). Thus,

[p]henetic taxonomy assumes that similarity can be measured and that nature is not continuous, so that mutually most similar taxa can be defined and placed into the non-overlapping hierarchy (Sokal 1985:735).

Phenetic clustering methods were largely developed within this particular taxonomic context. Specifically, they were developed to produce nested hierarchies from similarity data. Thus, "[t]he most common and convenient representation of the results of numerical taxonomy is by dendrograms" (Sokal and Sneath 1963:198), "which have the advantage that they are readily interpretable as conventional taxonomic hierarchies" (Sneath and Sokal 1973:260). Similarly:

For the taxonomist in particular, hierarchical classifications are attractive. . . . Hierarchical nonoverlapping classification pro-

FIGURE 7. DISTORTION OF PHENETIC RELATIONSHIPS BY PHENETIC CLUSTERING (BASED ON DATA IN FIGURE 6).

(a) Single linkage phenogram. (b) Strict consensus of four equivalent complete linkage phenograms. (c) Strict consensus of two equivalent average linkage phenograms produced by the UPGMA. (d) Phenogram representing only those clusters that are exclusive (nonoverlapping) in terms of similarity relationships. Notice that each of the first three phenograms contains some groups (e.g., ABC) that are not exclusive in terms of similarity. All of the phenograms distort the similarity relationships by failing to indicate, for example, that C is more similar to G ($D = 3.0$) than is B ($D = 4.5$).

duces groups, hereinafter termed *clusters*, whose relationships to one another are readily expressed in two dimensions, generally in the form of a dendrogram (Clifford and Stephenson 1975:28; emphasis in original).

In order to summarize and make sense of the diversity of organisms the taxonomist customarily constructs a taxonomic hierarchy in which a taxon occupies a position in a nested scheme . . .

Taxonomy as a quantitative science is concerned with the problems of constructing such (usually) hierarchical structures . . . (Dunn and Everitt 1982:2-3).

Because the Linnean system requires that taxonomic entities be arranged in a hierarchic nonoverlapping manner, most clustering methods used in biological taxonomy have been hierarchic, nonoverlapping techniques (Sokal 1986:430).

Pheneticists were aware of the problems of representing similarities with nested hierarchical taxonomies:

Hierarchical clustering techniques impose a hierarchical structure on data and we need to consider whether this is merited or whether it introduces unacceptable distortions of the original relationships between the OTUs . . . (Dunn and Everitt 1982:96).

[I]t is not necessarily obvious that a hierarchical system is the most faithful representation of organic diversity. Continua may exist in character space, which would make it difficult and rather arbitrary to decide how to arrange the taxa hierarchically (Sokal 1985:732).

For other examples see Cain and Harrison (1958) and Sneath (1962). They were also aware of the advantages of alternative methods:

The classificatory procedures described earlier would all result in the production of one or more two-dimensional graphs (dendrograms, minimum spanning trees, etc.) from a given set of data. In order to achieve such simplicity there has been of necessity a considerable loss of information. Much of this loss would be avoided if the data could be looked at in space of several dimensions (Clifford and Stephenson 1975:169).

Such measures as the cophenetic correlation coefficient (Sokal and Rohlf 1962) and other measures of stress have led to the real-

ization that hierarchic classifications often are poor representations of actual phenetic relationships found in nature. Far better representations are often obtained by summarizing the data in an ordination of as few as three dimensions (Sneath and Sokal 1973:201) [See also Rohlf 1967, 1968].

Nevertheless, the constraints imposed by the traditions of the taxonomic context were strong. Consequently, pheneticists continued to advocate methods that yielded nested, hierarchical results.

The relative merits of hierarchic versus nonhierarchic classifications are difficult to evaluate. For traditional biological taxonomy, hierarchic classifications are required. . . . Nonhierarchic representation is preferred when emphasis is placed on a faithful representation of the relationships among the OTU's rather than on a summarization of these relationships (Sneath and Sokal 1973:206).

[O]rdination or scaling techniques . . . may be very useful for indicating the taxonomic structure in a collection of organisms. However, they do not lead to an *explicit* separation of the organisms into groups and so neither do they produce classifications *per se*. For this the numerical taxonomist turns to one of the many available methods of *cluster analysis* (Dunn and Everitt 1982:77; emphasis in original).

Even when experimenting with greater taxonomic flexibility, pheneticists for the most part remained faithful to tradition:

This conventional arrangement [nonoverlapping taxa] has been built into customary taxonomic practice . . . when that concept is combined with a hierarchic classification it gives the familiar nested classifications. However . . . these quite frequently distort the phenetic relationships among OTUs. For this reason some workers have preferred to relax the criterion of mutual exclusiveness in taxonomy, and prefer to permit overlapping in membership at a given rank rather than to resort to ordination, the antithesis of a nested hierarchic classification (Sneath and Sokal 1973:207-208).

For example, Jardine and Sibson (1968a,b; see also Michener 1963) experimented with methods that permitted overlapping clusters. Only a few workers (e.g., DuPraw 1964, 1965),

went so far as to advocate taxonomies based on ordinations, but even these “non-Linnean classifications” were seen as supplemental to, rather than replacements for, traditional hierarchical taxonomies.

Advocacy of phenetic clustering can thus be understood as resulting from a constraint imposed by the traditions of the disciplinary context within which those methods were developed. By the time the phenetic movement started, the assumption of a nested hierarchical system had become thoroughly entrenched in biological taxonomy. As a consequence, no taxonomic movement was likely to succeed unless it could be accommodated within such a framework. According to Jardine (1969:49), “[i]t is improbable that non-hierarchical systems of classification will be found acceptable by biologists,” and Sokal (1985:733) noted that “proposals for non-Linnaean taxonomies . . . have foundered on the shoals of tradition.” It is not surprising, therefore, that methods producing nested hierarchical results were favored by phenetic taxonomists. Although nested nonoverlapping clusters were not the best means of representing phenetic relationships, they were critical to the viability of phenetics as a taxonomic movement.

The perspective developed above not only accounts for the advocacy of phenetic clustering despite known shortcomings, it also bears on the contribution of phenetics to biology in general and its fate as a taxonomic movement. Phenetics contributed many important analytical methods to biology. According to Hull (1988:233), “During the past two decades, numerical techniques have become increasingly prominent in systematics. . . . In this sense ‘numerical taxonomy’ has been extremely successful.” For example, Sokal, a leading pheneticist, was among the first to develop quantitative phylogenetic methods based on parsimony (e.g., Camin and Sokal 1965). He developed important methods for the study of geographic variation (e.g., Sokal 1979), and he pioneered the biological application of ordination techniques (e.g., Rohlf and Sokal 1962), which are now used widely in studies where the construction of taxonomies is not the primary focus and thus associated constraining assumptions are effectively removed. Quantitative methods developed by pheneti-

cists are also used widely in other biological disciplines. Morphometrics, a burgeoning field dealing with the analysis of shapes, has at least part of its origins in phenetics (Strauss 1991), and pheneticists made significant contributions to the field of biostatistics, where Sokal and Rohlf’s (1969, 1981) *Biometry* has become a standard reference.

But despite the success of phenetics in the areas noted above, as well as its popularity within taxonomy during the 1960s and 1970s, in the long run, phenetics has not been successful as a taxonomic movement. Hull (1988:233) described phenetic taxonomy as having “degenerated precipitously,” and according to Donoghue (1990:468), “it’s been dead in the water for some time” (see also Ghiselin 1984). Specifically, the phenetic approach has not come to predominate as the basis for the comprehensive taxonomic system. Sneath and Sokal’s (1973) *Numerical Taxonomy*, the standard of the phenetic movement, is devoted mostly to discussions about the workings of various analytical methods rather than the construction of working taxonomies. The authors themselves noted that much work had been done on methods for representing the relationships implied by similarity matrices, less on using those representations to establish taxa, and still less on the ranking and naming of taxa (Sneath and Sokal 1973:259). More importantly, the taxonomic philosophy advanced by pheneticists, with its advocacy of raw similarity and antipathy toward inferred evolutionary relationships as the basis for taxonomy, has been largely rejected (Hull 1988).

Hull (1988) attempted to explain the failure of phenetics in sociological and psychological terms: “[N]umerical taxonomists branched out too quickly. Before they had succeeded in establishing their methods for classification in biological systematics, they dissipated their energies in applying quantitative techniques in too many areas” (p 519). They also offered “a plethora of techniques, each with its own strengths, each with its own weaknesses . . . for the practicing systematist, it was immobilizing” (pp 519–520). In contrast, Donoghue (1990) argued for the importance of ideas in determining the success or failure of a movement. In his view, phenetics failed because sys-

tematists really wanted a taxonomy based on phylogeny rather than overall similarity.

Both Hull's and Donoghue's explanations have their merits, and they are not necessarily mutually exclusive. Our analysis suggests yet another explanation—or more properly, another part of the explanation—for the failure of phenetics as a taxonomic movement. Because similarity relationships do not strictly conform to the pattern of a nested hierarchy, attempts to base nested hierarchical taxonomies on similarity will never be entirely satisfactory. In short, there is a fundamental incompatibility between a basic premise of phenetics and a basic convention of taxonomy. This incompatibility is fundamental to explaining the failure of phenetics as a taxonomic movement in that it ties together the explanations of both Hull and Donoghue as parts of a single, unified explanation.

The incompatibility between similarity and hierarchical taxonomy explains why phenetics generated so many methods without being able to offer a clear preference for one of them, that is, Hull's (1988) primary explanation for the failure of phenetics. Ordination methods represented similarities more accurately than did clustering, but they did not give hierarchical results. Standard clustering methods gave hierarchical results, but each captured a somewhat different aspect of the complex pattern of phenetic relationships (Rohlf 1970). Moreover, many of the methods produced clusters that were not exclusive similarity groups, but methods that produced only exclusive similarity groups left an "unclassified residue" (Sneath and Sokal 1973:223) (see Figure 7d). Because of the nonhierarchical component of similarity, no method for converting similarities into traditional taxonomies was entirely satisfactory, and in retrospect, the development of such a method would seem to have been impossible.

The difficulty of developing hierarchical taxonomic methods based on similarity may in turn explain why pheneticists put their energies into the development of methods in different areas, that is, Hull's secondary explanation for the failure of phenetics. Pheneticists presumably did not realize that their taxonomic efforts were unlikely to succeed, but they were clearly aware of problems with the

methods they had already developed in this area (see above), and this may explain why they continued to develop new ones. In any case, given that pheneticists were experiencing greater success in other areas, this situation presumably would have favored a redirection of their efforts.

Finally, the poor fit of similarity to a nested hierarchical model may at least partly explain why systematists came to prefer common ancestry as the basis for taxonomy, that is, Donoghue's (1990) explanation for the failure of phenetics. In contrast with similarities (including patristic ones), common ancestry relationships conform to the model of a strict nested hierarchy, and consequently, they are easily accommodated with nested hierarchical taxonomies. Ironically, although pheneticists rejected phylogeny as the basis for taxonomy, they sometimes justified their assumption of nested hierarchical taxonomies by appealing to phylogeny (e.g., Sokal and Sneath 1963:171; Sneath and Sokal 1973:200). In any case, compatibility with the tradition of nested hierarchical taxonomies may be part of the reason for the current success of taxonomic approaches based on common ancestry (e.g., Hennig 1966; Eldredge and Cracraft 1980; Wiley 1981; Ax 1987).

In reaching these conclusions, we do not wish to advocate the primacy of traditions over the goals or functions of taxonomy. Specifically, we do not wish to advocate an approach based on common ancestry simply because it is compatible with a tradition of nested hierarchical taxonomies. Instead, we want to emphasize the constraints that traditions exert on taxonomy—a discipline steeped in traditions. Explicit consideration of the goals and functions of taxonomy may help to reveal assumptions that have previously been taken for granted, along with their accompanying constraints. But regardless of how the assumptions and constraints are revealed, an awareness of them can provide insights into both the usefulness of particular taxonomic methods and developments in the history of taxonomy. We hope that we have provided examples of such insights in the present article, in the first case, with regard to the limitations of phenetic clustering, and in the second, with regard to the endorsement of this family of methods de-

spite an awareness of its limitations, which is itself a key to understanding both the successes and the failures of phenetic taxonomy.

CONCLUSIONS

Phenetic clustering exhibits serious drawbacks not only in its specific applications to phylogeny reconstruction and studies of variation among populations but also as a general method for analysing patterns of similarity. Similarities do not exhibit the strictly nested hierarchical structure that phenetic clustering is constrained to yield. Consequently, if one is interested in analysing patterns of similarity, then phenetic clustering does not seem to be the most appropriate method for analysing those patterns, and if one is concerned with the construction of nested hierarchical taxonomies, then similarity does not seem to be the most appropriate property upon which to base those taxonomies.

Despite the problems with phenetic clustering, similarity data themselves can be very useful, for there are more appropriate ways of analysing such data. In phylogenetic studies, methods that explicitly analyse similarities as the result of evolutionary changes along the branches of trees can reconstruct relationships accurately in cases where phenetic clustering cannot, and they can reveal asymmetries

in branch lengths that constitute evidence for variation in rates of evolution among lineages, a phenomenon that can cause clustering to reconstruct the wrong tree. In studies of intraspecific variation, methods that analyse similarities in the context of the geographic relationships among populations can reveal evidence for the presence or absence of genetic continuity in situations involving geographic differentiation and uneven geographic sampling where clustering gives misleading results. And in studies of similarity for its own sake, methods that are unconstrained by the assumption of a nested hierarchy are able to analyse patterns of similarity more faithfully than is phenetic clustering. In all of these areas, direct inspection of a similarity or distance matrix often reveals aspects of complex phenetic relationships that are obscured by clustering. Phenetic clustering is not the most effective way to analyse similarity data.

ACKNOWLEDGMENTS

We thank J Felsenstein, D Hull, A Larson, J Losos, M Morris, D Swofford, D Wake, K Warheit, and an anonymous reviewer for discussions or comments on earlier versions of the paper. G Adest provided unpublished data on lizard allozymes. K Warheit and R Chapman provided access to copies of NTSYS-pc. This work was supported in part by a Tilton Postdoctoral Fellowship from the California Academy of Sciences to K de Queiroz.

REFERENCES

- Ax P. 1987. *The Phylogenetic System: The Systematization of Organisms on the Basis of their Phylogenesis*. Chichester: John Wiley.
- Adest G A. 1978. The relations of the sand lizards *Uma*, *Callisaurus* and *Holbrookia* (Sauria: Iguanidae): an electrophoretic study. [PhD dissertation]. University of California, Los Angeles.
- Avise J C, Arnold J, Ball R M, Bermingham E, Lamb T, Neigel J E, Reeb C A, Saunders N C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between populations genetics and systematics. *Annual Review of Ecology and Systematics* 18:489–522.
- Avise J C, Ball R M, Jr. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surveys in Evolutionary Biology* 7:45–67.
- Baker A J, Moeed A. 1987. Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). *Evolution* 41:525–538.
- Barton N H, Halliday R B, Hewitt G M. 1983. Rare electrophoretic variants in a hybrid zone. *Heredity* 50:139–146.
- Barton N H, Slatkin M. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56:409–415.
- Baum D A, Shaw K L. 1995. Genealogical perspectives on the species problem. In P C Hoch, A G Stephenson, editors. *Experimental and Molecular Approaches to Plant Biosystematics*. St Louis: Missouri Botanical Garden. pp 1–15.
- Cain A J, Harrison G A. 1958. An analysis of the taxonomist's judgement of affinity. *Proceedings of the Zoological Society of London* 131:85–98.
- Cain A J, Harrison G A. 1960. Phyletic weighting. *Proceedings of the Zoological Society of London* 135:1–31.
- Camin J H, Sokal R R. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19:311–326.

- Cavalli-Sforza L L, Edwards A W F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 21:550–570; *American Journal of Human Genetics* 19:233–257.
- Cliff A D, Ord J K. 1973. *Spatial Autocorrelation*. London: Pion Press.
- Clifford H T, Stephenson W. 1975. *An Introduction to Numerical Classification*. New York: Academic Press.
- Cole A J, Wishart D. 1970. An improved algorithm for the Jardine-Sibson method of generating overlapping clusters. *Computer Journal* 13:156–163.
- Colless D H. 1970. The phenogram as an estimate of phylogeny. *Systematic Zoology* 19:352–362.
- Crow J F, Temin R G. 1964. Evidence for partial dominance of recessive lethal genes in natural populations of *Drosophila*. *American Naturalist* 98:21–33.
- DeBry R W. 1992. The consistency of several phylogeny-inference methods under varying evolutionary rates. *Molecular Biology and Evolution* 9:537–551.
- de Queiroz K. 1989. Morphological and biochemical evolution in the sand lizards. [PhD dissertation]. University of California, Berkeley.
- de Queiroz K. 1992. Phylogenetic relationships and rates of allozyme evolution among the lineages of sceloporine sand lizards. *Biological Journal of the Linnean Society* 45:333–362.
- Derish P A, Sokal R R. 1988. A classification of European populations based on gene frequencies and cranial measurements: a map-quadrat approach. *Human Biology* 60:801–824.
- Donoghue M J. 1990. Sociology, selection, and success: a critique of David Hull's analysis of science and systematics. *Biology & Philosophy* 5:459–472.
- Dunn G, Everitt B S. 1982. *An Introduction to Mathematical Taxonomy*. Cambridge: Cambridge University Press.
- DuPraw E J. 1964. Non-Linnean taxonomy. *Nature* 202:849–852.
- DuPraw E J. 1965. Non-Linnean taxonomy and the systematics of honeybees. *Systematic Zoology* 14: 1–24.
- Edwards A W F, Cavalli-Sforza L L. 1964. Reconstruction of evolutionary trees. In V H Heywood, J McNeill, editors. *Phenetic and Phylogenetic Classification*. London: Systematics Association Publication Number 6. pp 67–76.
- Eldredge N, Cracraft J. 1980. *Phylogenetic Patterns and the Evolutionary Process*. New York: Columbia University Press.
- Epperson B K. 1993. Recent advances in correlation studies of spatial patterns of genetic variation. *Evolutionary Biology* 27:95–155.
- Estabrook G F, Johnson C S, Jr, McMorris F R. 1976a. An algebraic analysis of cladistic characters. *Discrete Mathematics* 16:141–147.
- Estabrook G F, Johnson C S, Jr, McMorris F R. 1976b. A mathematical foundation for the analysis of cladistic character compatibility. *Mathematical Biosciences* 29:181–187.
- Everitt B S. 1993. *Cluster Analysis*. Third Edition. London: Edward Arnold.
- Farris J S. 1969. On the cophenetic correlation coefficient. *Systematic Zoology* 18:279–285.
- Farris J S. 1970. Methods for computing Wagner trees. *Systematic Zoology* 19:83–92.
- Farris J S. 1971. The hypothesis of nonspecificity and taxonomic congruence. *Annual Review of Ecology and Systematics* 2:277–302.
- Farris J S. 1972. Estimating phylogenetic trees from distance matrices. *American Naturalist* 106:645–668.
- Farris J S, Kluge A G, Eckardt M J. 1970. A numerical approach to phylogenetic systematics. *Systematic Zoology* 19:172–189.
- Felsenstein J. 1973. Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22:240–249.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27:401–410.
- Felsenstein J. 1979. Alternative methods of phylogenetic inference and their interrelationship. *Systematic Zoology* 28:49–62.
- Felsenstein J. 1982a. Numerical methods for inferring evolutionary trees. *Quarterly Review of Biology* 57:379–404.
- Felsenstein J. 1982b. How can we infer geography and history from gene frequencies. *Journal of Theoretical Biology* 96:9–20.
- Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22:521–565.
- Fiala K L, Sokal R R. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. *Evolution* 39:609–622.
- Fitch W M, Margoliash E. 1967. Construction of phylogenetic trees. *Science* 155:279–284.
- Gabriel K R, Sokal R R. 1969. A new statistical approach to geographic variation analysis. *Systematic Zoology* 18:259–278.
- Ghiselin M T. 1984. Narrow approaches to phylogeny: a review of nine books on cladism. *Oxford Surveys in Evolutionary Biology* 1:209–222.
- Good D A, Wake D B. 1992. Geographic variation and speciation in the torrent salamanders (genus *Rhyacotriton*). *University of California Publications in Zoology* 126:1–91.

- Gower J C, Ross G J S. 1969. Minimum spanning trees and single linkage cluster analysis. *Applied Statistics* 18:54–64.
- Harman H H. 1960. *Modern Factor Analysis*. Chicago: University of Chicago Press. [Third Edition 1976].
- Hart G. 1983. The occurrence of multiple UPGMA phenograms. In J Felsenstein, editor. *Numerical Taxonomy*. Berlin: Springer-Verlag. pp 254–258.
- Hartigan J A. 1967. Representation of similarity matrices by trees. *Journal of the American Statistical Association* 62:1140–1158.
- Hartigan J A. 1975. *Clustering Algorithms*. New York: John Wiley.
- Hendy M D, Penny D. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38:297–309.
- Hengeveld R. 1990. *Dynamic Biogeography*. Cambridge: Cambridge University Press.
- Hennig W. 1966. *Phylogenetic Systematics*. Urbana: University of Illinois Press.
- Heywood V H. 1964. Introduction. In V H Heywood, J McNeill, editors. *Phenetic and Phylogenetic Classification*. London: Systematics Association Publication Number 6. pp 1–4.
- Highton R. 1989. Biochemical evolution in the slimy salamanders of the *Plethodon glutinosus* complex in the eastern United States. Part 1. Geographic protein variation. *Illinois Biological Monographs* 57:1–153.
- Hotelling H. 1933. Analysis of complex statistical variables into principal components. *Journal of Educational Psychology* 24:417–441, 498–520.
- Huelsenbeck J P. 1995. Performance of phylogenetic methods in simulation. *Systematic Biology* 44:17–48.
- Huelsenbeck J P, Hillis D M. 1993. Success of phylogenetic methods in the four-taxon case. *Systematic Biology* 42:247–264.
- Hull D L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.
- Jardine N. 1969. A logical basis for biological classification. *Systematic Zoology* 18:37–52.
- Jardine N, Sibson R. 1968a. A model for taxonomy. *Mathematical Bioscience* 2:465–482.
- Jardine N, Sibson R. 1968b. The construction of hierarchic and nonhierarchic classifications. *Computer Journal* 11:177–184.
- Jardine N, Sibson R. 1971. *Mathematical Taxonomy*. London: John Wiley.
- Jardine N, van Rijnsbergen C J, Jardine C J. 1969. Evolutionary rates and the inference of evolutionary tree forms. *Nature* 224:185.
- Jin L, Nei M. 1991. Relative efficiencies of the maximum-parsimony and distance-matrix methods of phylogeny construction for restriction data. *Molecular Biology and Evolution* 8:356–365.
- Jukes T H, editor. 1987. Special issue—molecular evolutionary clock. *Journal of Molecular Evolution* 26:1–171.
- Kim J, Rohlf F J, Sokal R R. 1993. The accuracy of phylogenetic estimation using the neighbor-joining method. *Evolution* 47:471–486.
- Kirsch J A W. 1969. Serological data and phylogenetic inference, the problem of rates of change. *Systematic Zoology* 18:296–311.
- Kruskal J B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society* 7:48–50.
- Kruskal J B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.
- Kruskal J B. 1964b. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29:115–129.
- Lake J A. 1987a. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Molecular Biology and Evolution* 4:167–191.
- Lake J A. 1987b. Determining evolutionary distances from highly diverged nucleic acid sequences: operator metrics. *Journal of Molecular Evolution* 26:59–73.
- Larson A, Wake D B, Yanev K P. 1984. Measuring gene flow among populations having high levels of genetic fragmentation. *Genetics* 106:293–308.
- Legendre P. 1987. Constrained clustering. In P Legendre, L Legendre, editors. *Developments in Numerical Ecology*. NATO ASI Series, Volume G14. Berlin: Springer-Verlag. pp 289–307.
- Le Quesne W J. 1969. A method of selection of characters in numerical taxonomy. *Systematic Zoology* 18:201–205.
- Le Quesne W J. 1972. Further studies based on the uniquely derived character concept. *Systematic Zoology* 21:281–288.
- Le Quesne W J. 1974. The uniquely evolved character concept and its cladistic application. *Systematic Zoology* 23:513–517.
- Lessa E P. 1990. Multidimensional analysis of geographic genetic structure. *Systematic Zoology* 39:242–252.
- Li W-H. 1981. Simple method for constructing phylogenetic trees from distance matrices. *Proceedings of the National Academy of Science, USA* 78:1085–1089.
- Li W-H. 1993. So, what about the molecular clock hypothesis? *Current Opinions in Genetics and Development* 3:896–901.
- Majumder P P. 1988. Principal axis analysis of gene frequency data. *American Journal of Physical Anthropology* 76:313–320.

- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27:209–220.
- Mayr E. 1965. Numerical phenetics and taxonomic theory. *Systematic Zoology* 14:73–97.
- McQuitty L L. 1963. Rank order typal analysis. *Education and Psychological Measurement* 23:55–61.
- McQuitty L L. 1965. A conjunction of rank order typal analysis and item selection. *Education and Psychological Measurement* 25:949–961.
- McQuitty L L. 1967. A mutual development of some typological theories and pattern-analytic methods. *Education and Psychological Measurement* 27:21–46.
- Menozzi P, Piazza A, Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. *Science* 201:786–792.
- Michener C D. 1963. Some future developments in taxonomy. *Systematic Zoology* 12:151–172.
- Michener C D, Sokal R R. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130–162.
- Moore G W. 1971. A mathematical model for the construction of cladograms. [PhD thesis]. North Carolina State University, Raleigh.
- Moritz C, Hillis D M. 1990. Molecular systematics: context and controversies. In D M Hillis, C Moritz, editors. *Molecular Systematics*. Sunderland (MA): Sinauer Associates. pp 1–10.
- Nei M. 1972. Genetic distance between populations. *American Naturalist* 106:283–292.
- Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 19:153–170.
- O'Hara R J. 1988. Homage to Clio, or, toward an historical philosophy for evolutionary biology. *Systematic Zoology* 37:142–155.
- Penny D, Hendy M D, Steel M A. 1992. Progress with methods for constructing evolutionary trees. *Trends in Ecology & Evolution* 7:73–79.
- Prager E M, Wilson A C. 1978. Construction of phylogenetic trees for proteins and nucleic acids: empirical evaluation of alternative matrix methods. *Journal of Molecular Evolution* 11:129–142.
- Prim R C. 1957. Shortest connection networks and some generalizations. *Bell System Technical Journal* 36:1389–1401.
- Rendine S, Piazza A, Cavalli-Sforza L L. 1986. Simulation and separation by principal components of multiple demic expansions in Europe. *American Naturalist* 128:681–706.
- Rohlf F J. 1967. Correlated characters in numerical taxonomy. *Systematic Zoology* 16:109–126.
- Rohlf F J. 1968. Stereograms in numerical taxonomy. *Systematic Zoology* 17:246–255.
- Rohlf F J. 1970. Adaptive hierarchical clustering schemes. *Systematic Zoology* 18:58–82.
- Rohlf F J. 1990. Morphometrics. *Annual Review of Ecology and Systematics* 21:299–316.
- Rohlf F J, Bookstein F L, editors. 1990. *Proceedings of the Michigan Morphometrics Workshop*. Michigan: University of Michigan Museum of Zoology, Special Publication Number 2.
- Rohlf F J, Chang W S, Sokal R R, Kim J. 1990. Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. *Evolution* 44:1671–1684.
- Rohlf F J, Sokal R R. 1962. The description of taxonomic relationships by factor analysis. *Systematic Zoology* 11:1–16.
- Romesburg H C. 1984. *Cluster Analysis for Researchers*. Belmont (CA): Lifetime Learning.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406–425.
- Sanchez-Mazas A, Langaney A. 1988. Common genetic pools between human populations. *Human Genetics* 78:161–166.
- Sarich V M, Wilson A C. 1967a. Rates of albumin evolution in primates. *Proceedings of the National Academy of Science, USA* 58:142–148.
- Sarich V M, Wilson A C. 1967b. Immunological time scale for hominid evolution. *Science* 158:1200–1203.
- Shaffer H B. 1983. Biosystematics of *Ambystoma rosaceum* and *A. tigrinum* in northwestern Mexico. *Copeia* 1983:67–78.
- Shepard R N. 1962a. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* 27:125–140.
- Shepard R N. 1962b. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* 27:219–246.
- Shepard R N. 1966. Metric structures in ordinal data. *Journal of Mathematical Psychology* 3:287–315.
- Slatkin M. 1981. Estimating levels of gene flow in natural populations. *Genetics* 99:323–335.
- Slatkin M. 1985a. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393–430.
- Slatkin M. 1985b. Rare alleles as indicators of gene flow. *Evolution* 39:53–65.
- Slatkin M. 1987. Gene flow and geographic structure of natural populations. *Science* 236:787–792.
- Slatkin M, Arter H E. 1991. Spatial autocorrelation methods in population genetics. *American Naturalist* 138:499–517.
- Slatkin M, Barton N. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–1368.

- Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow from the phylogenies of alleles. *Genetics* 123:603–613.
- Slatkin M, Maddison WP. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* 126:249–266.
- Sneath P H A. 1957a. Some thoughts on bacterial classification. *Journal of General Microbiology* 17:184–200.
- Sneath P H A. 1957b. The application of computers to taxonomy. *Journal of General Microbiology* 17:201–226.
- Sneath P H A. 1962. The construction of taxonomic groups. In G C Ainsworth, P H A Sneath, editors. *Microbial Classification*. Cambridge: Cambridge University Press. pp 289–332.
- Sneath P H A. 1967. Trend-surface analysis of transformation grids. *Journal of Zoology, London* 151:65–122.
- Sneath P H A, Sokal R R. 1962. Numerical taxonomy. *Nature* 193:855–860.
- Sneath P H A, Sokal R R. 1973. *Numerical Taxonomy*. San Francisco: W H Freeman.
- Sober E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge (MA): MIT Press.
- Sokal R R. 1963. The principles and practice of numerical taxonomy. *Taxon* 12:190–199.
- Sokal R R. 1979. Testing statistical significance of geographic variation patterns. *Systematic Zoology* 28:227–232.
- Sokal R R. 1983a. A phylogenetic analysis of the caminalcules. II. Estimating the true cladogram. *Systematic Zoology* 32:185–201.
- Sokal R R. 1983b. Analysing character variation in geographic space. In J Felsenstein, editor. *Numerical Taxonomy*. Berlin: Springer-Verlag. pp 384–403.
- Sokal R R. 1985. The continuing search for order. *American Naturalist* 126:729–749.
- Sokal R R. 1986. Phenetic taxonomy: theory and methods. *Annual Review of Ecology and Systematics* 17:423–442.
- Sokal R R, Michener C D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* 38:1409–1438.
- Sokal R R, Oden N L. 1978a. Spatial autocorrelation in biology. 1. Methodology. *Biological Journal of the Linnean Society* 10:199–228.
- Sokal R R, Oden N L. 1978b. Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. *Biological Journal of the Linnean Society* 10:229–249.
- Sokal R R, Rohlf R J. 1962. The comparison of dendrograms by objective methods. *Taxon* 11:33–40.
- Sokal R R, Rohlf F J. 1969. *Biometry: The Principles and Practice of Statistics in Biological Research*. San Francisco: W H Freeman. [Second Edition 1981].
- Sokal R R, Sneath P H A. 1963. *Principles of Numerical Taxonomy*. San Francisco: W H Freeman.
- Sokal R R, Uytterschaut H, Rosing F W, Schwidetzky I. 1987. A classification of European skulls from three time periods. *American Journal of Physical Anthropology* 74:1–20.
- Sokal R R, Wartenberg D E. 1983. A test of spatial autocorrelation analysis using an isolation by distance model. *Genetics* 105:219–237.
- Sourdis J, Krimbas C. 1987. Accuracy of phylogenetic trees estimated from DNA sequence data. *Molecular Biology and Evolution* 4:159–166.
- Strauss R E. 1991. Size and shape [Review of: *Proceedings of the Michigan Morphometrics Workshop*, edited by F J Rohlf and F L Bookstein, 1990]. *Science* 253:1152.
- Studier J A, Keppler K J. 1988. A note on the neighbor-joining algorithm of Saitou and Nei. *Molecular Biology and Evolution* 5:729–731.
- Swofford D L. 1981. On the utility of the distance Wagner procedure. In V A Funk, D R Brooks, editors. *Advances in Cladistics*. Bronx: New York Botanical Garden. pp 25–43.
- Swofford D L, Berlocher S H. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Systematic Zoology* 36:293–325.
- Swofford D L, Maddison W P. 1987. Reconstructing ancestral character states under Wagner parsimony. *Mathematical Biosciences* 87:199–229.
- Swofford D L, Olsen G J. 1990. Phylogeny reconstruction. In D M Hillis, C Moritz, editors. *Molecular Systematics*. Sunderland (MA): Sinauer Associates. pp 411–501.
- Tateno Y, Nei M, Tajima F. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *Journal of Molecular Evolution* 18:387–404.
- Thorpe R S. 1983. A review of numerical methods for recognizing and analysing racial differentiation. In J Felsenstein, editor. *Numerical Taxonomy*. Berlin: Springer-Verlag. pp 404–423.
- Tyron R C. 1939. *Cluster Analysis*. Ann Arbor (MI): Edwards Brothers.
- Vernon K. 1988. The founding of numerical taxonomy. *British Journal for the History of Science* 21:143–159.
- Wallace B. 1966. On the dispersal of *Drosophila*. *American Naturalist* 100:551–563.
- Wehrhahn C F, Powell R. 1987. Electrophoretic variation, regional differences, and gene flow in the coho salmon (*Onchorhynchus kisutch*) of southern British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* 44:822–831.
- Weir B S, Cochran C C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.

- Wiley E O. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York: John Wiley.
- Wilson A C, Carlson S S, White T J. 1977. Biochemical evolution. *Annual Review of Biochemistry* 46:573–639.
- Wilson A C, Ochman H, Prager E M. 1987. Molecular time scale for evolution. *Trends in Genetics* 3:241–247.
- Wright S. 1943. Isolation by distance. *Genetics* 28:114–138.
- Wright S. 1946. Isolation by distance under diverse systems of mating. *Genetics* 31:39–59.
- Wright S. 1951. The genetical structure of populations. *Annals of Eugenics* 15:323–354.
- Wright S. 1969. *Evolution and the Genetics of Populations*. Volume II: The theory of gene frequencies. Chicago: University of Chicago Press.
- Wright S. 1978. *Evolution and the Genetics of Populations*. Volume IV: Variability within and among natural populations. Chicago: University of Chicago Press.
- Zuckerkandl E, Pauling L. 1962. Molecular disease, evolution, and genetic heterogeneity. In M Kasha, B Pullman, editors. *Horizons in Biochemistry*. New York: Academic Press. pp 189–225.