



The American Society of Naturalists

The Continuing Search for Order

Author(s): Robert R. Sokal

Source: *The American Naturalist*, Vol. 126, No. 6 (Dec., 1985), pp. 729-749

Published by: The University of Chicago Press for The American Society of Naturalists

Stable URL: <http://www.jstor.org/stable/2461253>

Accessed: 01/11/2008 18:35

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=ucpress>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press and The American Society of Naturalists are collaborating with JSTOR to digitize, preserve and extend access to *The American Naturalist*.

THE CONTINUING SEARCH FOR ORDER

ROBERT R. SOKAL

Department of Ecology and Evolution, State University of New York, Stony Brook,
New York 11794-5245*Submitted May 28, 1985*

ORDER AND DIVERSITY

The subject matter of this essay, taxonomy, was probably the most frequent topic of discussion during the early years of the American Society of Naturalists. Various aspects of taxonomy and discussions of particular taxonomies were regular features of the meetings of the Society since its founding in the nineteenth century and well into the present century. But in these days when molecular genetics provides us with new and exciting discoveries on a regular basis—findings that are of profound importance for an understanding of the evolution of organisms—one may well wonder why scientists should still bother with taxonomy. Is this not an outdated science, practiced by a few unreconstructed museum types, whose ideas are as dusty as some of the cases and specimens with which they surround themselves?

Any such attitude reveals a lack of understanding and appreciation of the fundamental role that taxonomy continues to play in modern biology. There are few papers, if any, published in *The American Naturalist* or presented at any meeting of the Society that are not founded on comparative biology. That this is more than a propagandistic claim can be shown by a detailed analysis of any series of articles in the journal. Even if the thrust of a given paper is along lines quite remote from taxonomy, this science is implicated whenever a comparative approach is taken in which a biological phenomenon is compared over different groups of organisms. Alternatively, a taxonomic hypothesis is implied when the generality of a phenomenon is assumed. When we study transposable elements in *Drosophila*, it is believed that these are not exclusively restricted to *Drosophila melanogaster*, but that other *Drosophila*, and presumably other flies (indeed, other organisms as well), will exhibit the phenomenon. It is this generality that justifies the effort invested into an elucidation of such complex biological processes whose investigation is so time-consuming and expensive as to make a comprehensive analysis in numerous organisms prohibitive. Of course, whenever evolutionary hypotheses are tested, the importance of taxonomy is demonstrated directly. For without a taxonomic framework in which these hypotheses are considered, evolutionary hypotheses such as punctuated-equilibrium evolution, constant-rate evolution, species selection, or vicariance biogeography could not

be tested. Thus, classifications are necessary if we are to engage in biology. But how are these classifications to be constructed?

The classificatory system set up by Linnaeus is hierarchical. At the lower-rank levels—species and genus—the Linnaean system was a direct continuation of folk taxonomy. Primitive peoples have been shown to recognize many of the species in their immediate surroundings, and they resemble the European tradition preceding Linnaeus in being aware of the similarity of some species and assembling these conceptually and linguistically into groups of similar species—the genera (see, e.g., Breedlove and Raven 1974). Linnaeus extended the method by establishing higher-ranking taxa and defining them on the basis of group characteristics that distinguished one taxon from another at the same rank. These differentiating characteristics were to all practical purposes Aristotelian, defining essences of the established groups (Cain 1958). It has been pointed out repeatedly that the nature of the classificatory system has changed little since Linnaeus' time, although, of course, the classifications themselves have undergone drastic changes and expansions. It is also remarkable to what extent the classificatory system remained unaffected by the advent of evolutionary theory in the latter half of the nineteenth century. The classificatory system was soundly established and successful well before Darwin's time. As numerous authors have shown, the early taxonomists attempted to establish a natural system, and whereas evolutionary theory furnished a new explanation for the natural system, the thought processes concerning its establishment and the actual taxonomic practices changed little (e.g., Remane 1956).

But what is a natural system? The common answer is that it is a system that reflects the state of nature. Early biologists and indeed the common people, as reflected in their various languages, seem to have recognized a roughly hierarchical arrangement of natural diversity. So, to the degree that nature is truly hierarchical, a system should be the more natural, the more it conforms to the true innate hierarchy of nature (if only we could know it). An alternative view is that a natural system is one that is natural to the human mind. Do we find it easier to pigeonhole objects into mutually exclusive classes that are subsumed under ever more inclusive nonoverlapping classes? Cognitive psychologists tend to think that we do, and that the human mind cannot easily perceive membership in overlapping classes and in continua (Smith and Medin 1981). It has even been suggested that because we are evolving organisms and because the human mind has to deal with an evolved (and hence hierarchical) universe, the structure of our thinking developed so as to favor hierarchical interpretations (Riedl 1976, p. 230). In actual work with arranging organisms in a natural system, it was eventually discovered that group characteristics were difficult to define categorically and that exceptional organisms could always be found. Such organisms would in most respects clearly be members of a given taxon but would lack one or another defining characteristic. Groups with such members were termed polythetic (Beckner 1959; Sneath 1962), and most taxonomists have recognized that natural taxa are largely polythetic.

We may visualize a polythetic taxon as arrayed in a space defined by the characters describing members. In figure 1 the horizontal and vertical dimensions

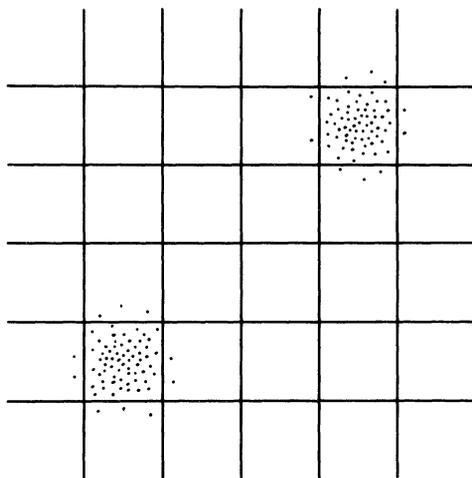


FIG. 1.—Two polythetic taxa defined by two characters. (For explanation, see the text.)

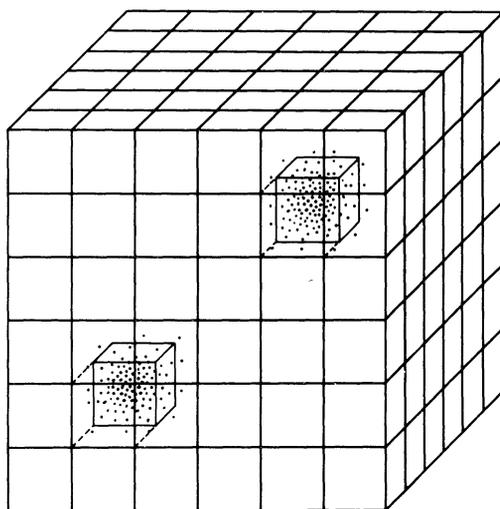


FIG. 2.—Two polythetic taxa defined by three characters. (For explanation, see the text.)

define the space spanned by two characters. The parallel lines indicate different states of the two characters, and the two clusters of points are two taxa. Most of their members share a specific combination of character states for the two characters and are shown within one of the squares defined by the lattice. But some members possess an alternative state for one or the other character, and yet belong to the taxon (the cloud of points). Such members are shown as points in neighboring squares. When three characters are considered simultaneously, the character space can be represented as a cubic lattice (fig. 2); polythetic taxa largely reside inside cubes representing the combinations of states of three charac-

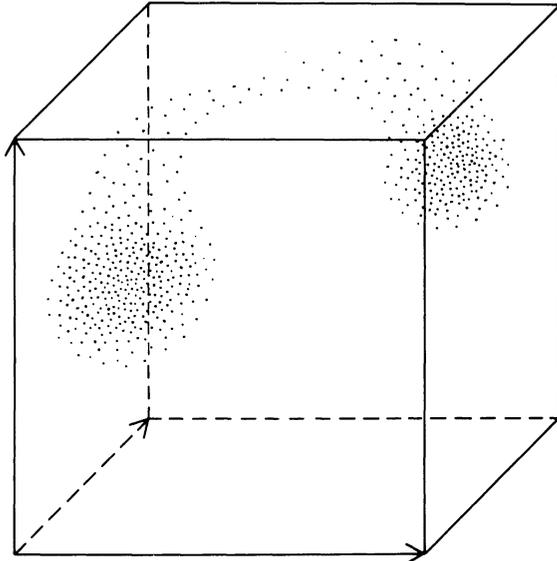


FIG. 3.—Two clusters (taxa) that show a narrow continuum in a three-dimensional character space.

ters within this lattice. Here again, some of the cluster members do not share all three “defining” character states.

In recent years, cognitive psychologists (Smith and Medin 1981) have recognized that the representation of a concept is a summary description of an entire class and cannot be restricted to a set of necessary and sufficient conditions; rather, it provides some measure of a central tendency of the patterns of its members. This probabilistic view of concepts corresponds to polythetic classes in biological taxonomy. If humans generally process information and form concepts on the basis of polythetic classes, it is not surprising that such classes have been developed for biological classification. Thus, it is also possible that taxonomists have been reinforced in the formation of hierarchical and polythetic classifications because they have an inborn tendency toward such arrangements.

Nevertheless, it 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 (fig. 3). This is most apparent at the population level, at which any hierarchical division can be shown to be inadequate. This can easily be demonstrated in human populations. Where should one draw the boundary between Europid and Mongolid populations in Siberia? Or between the Europid Weddids of India and the Southern Mongolids of Malaysia and Indonesia? Or between the Chinese and the Southern Mongolids of the Indochinese peninsula? Clearly all such boundaries are more or less arbitrary and may falsely suggest differences between equally ranked populations, even where such differences do not exist.

Do such continua also exist at higher taxonomic ranks? In numerous taxonomic

groups, not only genera, but families as well, grade into one another. Whereas the biological processes resulting in a phenetic continuum of populations within species are easily understood, one might expect the divergent dendritic nature of the phylogenetic process to result in discontinuities that would easily submit to a hierarchical arrangement. Yet parallelism, convergence, and hybridization readily produce results that will appear as continua in phenetic space.

There have been proposals from time to time to deal with this problem by abandoning the traditional Linnaean classificatory system in favor of one that is better adapted to reflect the actual taxonomic diversity (see, e.g., DuPraw 1964). All such proposals for non-Linnaean taxonomies, however, have foundered on the shoals of tradition, and it may indeed be true that humans, including taxonomists, have an innate bias toward hierarchical systems.

SCHOOLS OF TAXONOMY

In this century, biologists have attempted to provide a theoretical basis for taxonomy that goes beyond the considerations of the bare outline given above. There are three currently active schools of taxonomy. These are known as phenetic taxonomy, cladistics, and evolutionary systematics. Phenetic taxonomy (Sneath and Sokal 1973) is a system of classification based on the overall similarity of the organisms being classified. The similarity is expressed in terms of phenotypic characters (Cain and Harrison 1960). The goal of phenetic taxonomy is to arrange objects or operational taxonomic units (OTU's) in a stable and convenient classification. It is believed that basing classifications on similarity will result in such stability and convenience. The measurement of similarity is made on the basis of numerous, equally weighted characteristics. The degree of belonging to a class is based on its constituent properties. Following Gilmour's dictum (1937, 1940, 1951, 1961) that a system of classification is the more natural the more propositions there are that can be made regarding its constituent classes, affinity in a polythetic taxon is based on the greatest number of shared character states. No single state is either essential to group membership or sufficient to make an organism a member of the group.

In cladistic classification (Eldredge and Cracraft 1980; Wiley 1981), one establishes classes based on estimated cladograms, or branching trees, of phylogenetic relationships. Although there is currently considerable debate among cladists concerning the goals of cladistic classification (some recent workers—pattern cladists—appear to seek internal consistency of patterns divorced from the true genealogy of the organisms under study; for a discussion of this issue, see Beatty 1982; Brooks and Wiley 1985; Platnick 1985), I concur with Cracraft (1983) in considering the estimation of the phylogenetic relationship or genealogical affinities to be the ultimate purpose of establishing cladistic classifications. The cladograms are estimated by postulating monophyletic sister groups which, in turn, are based on putative synapomorphies, shared derived character states believed to have arisen in a common ancestor. Monophyly is defined by a strict criterion: all members of a taxon have a common ancestor, all of whose descendants are members of the taxon.

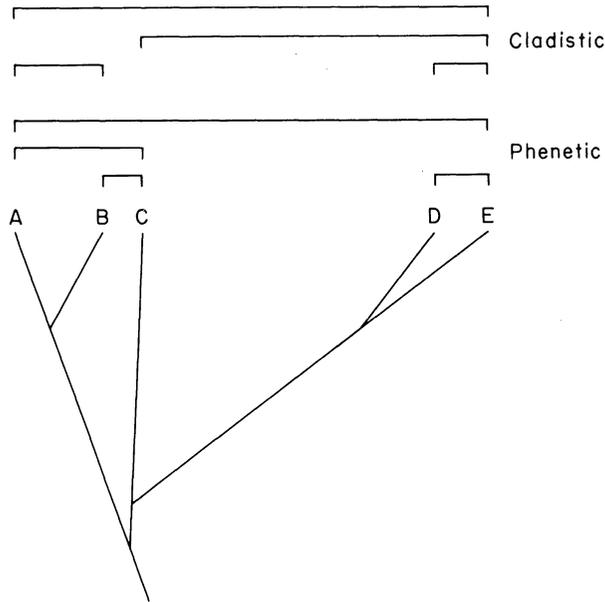


FIG. 4.—Dendrogram for five operational taxonomic units (OTU's) illustrating how the three schools of taxonomy would arrive at different classifications on the basis of the information furnished. (For further explanation, see the text.)

The evolutionary systematists (Mayr 1969, 1982 [p. 233]; Bock 1977; Ashlock 1980) also seek classes that share a common ancestor, but they permit some groups to be established in a classification that are not entirely monophyletic as defined by the cladists. As has been pointed out by pheneticists, as well as by cladists (Sneath and Sokal 1973; Cracraft 1983), evolutionary systematists have not attempted to establish a series of objective and quantitative procedures that result in classifications desirable from their point of view. In fact, Estabrook (1978) has provided a useful definition through his concepts of character state trees and convex taxa which, however, have not been adopted by other advocates of evolutionary systematics.

The differing classifications of these schools of taxonomy are illustrated in figure 4. Phenetic and cladistic relations for five OTU's are approximated by the dendrogram. The horizontal axis represents phenetic dissimilarity; the presumed phylogenetic branching sequence is shown by the bifurcations; and the vertical axis indicates time. Because OTU's D and E have diverged greatly from the other three and are mutually close, they would be placed in their own taxon by pheneticists. OTU C has not diverged much from the ancestral stem, and some convergence in B has made B and C quite similar. The phenetic classification, therefore, erects a taxon BC, which at a higher-rank level adds A to form the taxon ABC, which ultimately joins DE to include all OTU's. This classification is shown by "roofs" over the OTU's. Cladists establish the taxa AB and DE because these pairs of OTU's share a most recent common ancestor that is not also shared by other OTU's. In the cladistic classification (again shown by roofs), OTU C joins

taxon DE because these three share a common ancestor before taxa AB and CDE are united. It is not possible to specify how a taxonomist employing evolutionary systematics would classify these taxa, since no explicit classificatory rules are given by that school. Depending on whether the divergence of DE and convergence of B and C are considered more striking or whether more importance is given to the common ancestry of C and DE, the phenetic or cladistic classification may be adopted. For this reason, no roofs are shown for evolutionary systematics.

What are the assumptions that are used by the various schools of taxonomy? The Linnaean system assumes a nonoverlapping hierarchical arrangement of nature. All three schools of classification meet the assumption of the Linnaean system in attempting to create such an arrangement. Phenetic 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 nonoverlapping hierarchy. Cladistic taxonomy postulates a bifurcating, divergent topology and further assumes that these bifurcations are characterized by synapomorphies and that these synapomorphies can be recognized. Evolutionary systematics has no clearly stated special assumptions. If the character-state tree model of Estabrook (1978) is adopted, then one could state that classifications acceptable by evolutionary systematics should be convex, but, as already stated, this is a technical, not generally accepted definition.

Some of the assumptions are shared by the schools of taxonomy. The assumption that a hierarchical arrangement of taxa is appropriate is common to all three. The assumptions that an overall similarity can be quantified and that there are discontinuities among taxa are essential to phenetic numerical taxonomy; they must be applied in evolutionary systematics as well, because the departures of evolutionary classifications from strictly cladistic classifications are based on the amount of phenetic differentiation of various subgroups of a taxonomic tree. Cladistic numerical taxonomic methods do not specifically require this assumption, but in fact, a cladogram with the character states superimposed on it can result in a quantification of similarity of the terminal OTU's, which may exhibit discontinuities if a given bifurcation is supported by numerous synapomorphies. The notion of the length of a tree, commonly employed when estimating cladograms by parsimony methods, also implies a measure of dissimilarity between OTU's and their ancestors. The bifurcation and synapomorphy assumptions in classification are specific to cladistics, including numerical cladistics, and are not an essential assumption of any other method. The convexity of character-state trees as defined by Estabrook is a requirement of evolutionary classifications only. It is not rigorous enough for numerical cladistics, and it is not applicable to phenetic classifications.

All these assumptions have been subjected to serious questioning. We have already seen that the hierarchical nonoverlapping arrangement of organized nature is not necessarily the best representation of the actual diversity. The measurement of similarity by pheneticists has been criticized as being ambiguous, subject to variations in character coding, scaling, and similarity coefficients. Even if these effects can be ignored, differences in the clustering algorithms produce

differences in classifications (Sneath and Sokal 1973, p. 427). The operational assumptions of cladists that cladograms should be entirely bifurcating trees can surely not be true in nature; nor will the presence of shared derived character states in the two sister species be inevitable in a given data set. There has additionally been considerable controversy concerning the methods for recognizing the direction or polarity of character-state change, a necessary step in cladistic analysis (Eldredge and Cracraft 1980, p. 54; Brooks and Wiley 1985; Kluge 1985).

The assumptions made by the various taxonomic schools are not appropriate at all systematic levels. Thus, the hierarchical assumption is clearly not applicable for populations within a species (and even less for individuals within a population). Except for a few specialized instances in which populations split as a result of a dendritic branching process, the notion of a hierarchical system of populations is not supportable. One of these rare exceptions is that of the villages of the Yanomama Indians in South America, which exhibit a true branching system. Because each newly branched population occupies a new habitat, the process continues with little gene flow and without the replacement of one terminal population by another (Neel 1978). There should be analogies of this phenomenon in the plant and animal world, perhaps when there are introductions of new populations into unoccupied habitats, but this situation is not typical of populations in an existing, reasonably stable environment.

The assumption that similarity can be quantified holds for all taxonomic levels, but that concerning discontinuities does not. Discontinuities possibly do not occur among populations within species, and perhaps not even among species within a genus, since gene flow will tend to diminish the distinctness of their phenetic boundaries. The cladistic assumptions of bifurcations and of synapomorphies marking these branches are not tenable for populations within a species, but they may apply to all higher levels. Similar relations hold for the evolutionary systematic assumption of the convexity of taxa. Phenetic taxonomy can handle all of the levels, although because of a lack of discontinuities at the intraspecific level, the techniques for populations within a species have to be tailored to the nature of the material and generally involve ordinations, rather than cluster analyses. The other schools have difficulties incorporating populations within a species into their systems.

The emergence of the modern schools of taxonomy was accompanied by the development of numerical methods for obtaining the classifications (Sokal and Sneath 1963; Camin and Sokal 1965; Fitch and Margoliash 1968; Farris et al. 1970; Sneath and Sokal 1973). Characters were precisely defined, expressed or coded numerically, and subjected to appropriate algorithms intended to achieve the goals of a given school. In practice, however, the methods developed until now only approximate these goals at best. These developments in zoology and botany, with parallel trends in anthropology, have been given the general name of numerical taxonomy with subdivisions of numerical phenetics, numerical cladistics, and numerical systematics, corresponding roughly to their applications to the respective taxonomic schools.

Regardless of one's classificatory philosophy, the trend toward precise definitions of characters and their states and toward an explicit presentation of

complete data sets, which was brought about by the development of numerical taxonomy, is undoubtedly a considerable improvement over the subjective, traditional methods still practiced by many taxonomists. The taxonomic process is fraught with considerable risk of subjective bias. Battles among individual taxonomists regarding the importance and suitability of certain characters are legendary. Paleoanthropology still reflects this trend today, each new find seemingly the cause of a new argument. Different taxonomists using subjective methods can obtain quite different classifications of the same taxa by stressing different characters (Sokal and Rohlf 1980). Yet, it can be shown that attempts by different persons to obtain explicit data matrices of the same group result in largely identical classifications, regardless of subjective divergences in character coding and terminology (Sokal and Rohlf 1970). Different populations, such as Europeans and Chinese, do not agree on the characteristics they employ to distinguish their groups (Sokal 1974): whereas Europeans are struck by "slanted" eyes, skin color, and prominent cheekbones (in that order), Chinese emphasize wavy and light-colored hair and the prominent noses of Europeans. Similar cultural and personal biases surely affect the work of taxonomists working with other species.

In lectures to my classes I have for years cited the example of the intergradation between the Mongolid and Indian populations of Southeast Asia to point out the inadequacy of a hierarchical arrangement of these groups. Lately I have recognized that this view is also biased, from the perspective of the larger population groups at the ends of this spectrum: the Chinese and the Indians. A Khmer, Thai, or Burmese doubtlessly does not consider himself or herself as an intergrade, but rather as a clearly defined central type with the individuals in these more numerous but peripheral populations being extreme departures from one's own norm. In a similar way, greater familiarity with one of several related groups, or a greater number of species in some taxa than in others, may distort the taxonomic judgment. A related phenomenon is the greater perception of diversity in one's own group or in a familiar group with respect to a less familiar one. Europeans tend to stress the great diversity of Europid types even when these are restricted to those found in Europe alone. There is no reason to doubt that a comparable diversity exists in the population of China; yet this is rarely perceived by European scientists. Although even the intelligent layman is aware of the great diversity of the African population, differences are nevertheless deemphasized by comparison with one's own group. Similar biases exist in the taxonomy of all organisms and can be minimized only by requiring explicit and clearly defined data matrices, such as those used in numerical taxonomy.

Let us next examine the purposes of taxonomies. There is, first of all, the need to provide a "system of nature," a classification for laymen as well as other scientists. This is the classificatory system that exists today. It has been established generally on an ill-defined basis, typically without quantitative or other objective criteria. The importance of a consistent and stable classificatory system for reference purposes cannot be overemphasized. The public at large and scientists who are not systematists need a system that assigns an organism to a taxon, gives it a name, associates it with other related taxa (the meaning of related here is, of course, the catchword being defined variously by different schools of

taxonomy), and enables statements and inferences to be made about the characteristics of a given group. Biologists of all kinds need such a system of classification to test the generality of phenomena they observe; they need to know what other organisms exist that are closely related, so that a phenomenon observed in one member of the taxon might be looked for in other members.

A second purpose is to generate hypotheses about evolutionary relationships. Are there laws about the number of species to be found in a genus? About the density with which species fill niche spaces; about the conformation of such space; about the diversity in shape and width of niches; and so forth? To this end, more than a formal classification is required. A phenetic classification quantifying the amount of variation of OTU's within and among taxa is necessary for studying these phenomena.

A third purpose of a classification is to serve as a model or benchmark against which to test hypotheses about evolutionary phenomena, such as phyletic evolution, gradualism versus punctuated-equilibrium evolution, convergence, reversed evolution, and the like. For this purpose, one requires an estimated cladogram of the group and a mapping of the character-state changes that are believed to have occurred along its edges. If we wish to estimate the true phylogenetic tree of a taxon to test various evolutionary hypotheses, cladistic methods would seem indicated. But, as we shall see, methods taken from other schools of taxonomy may frequently yield better results.

The relations of the various taxonomic methods to these taxonomic purposes can be summarized as follows: all methods aim at representing a "natural" system. Phenetics and evolutionary systematics may serve to test hypotheses about patterns of diversity or analyses of niche hypervolume. Evolutionary systematics can furnish evolutionary rate, as well as serve as a benchmark for testing hypotheses about evolutionary phenomena as described above. Cladistic taxonomy can only provide such a benchmark. These purposes are met, of course, only to the degree to which the methods of each school succeed in meeting their goals.

OPTIMALITY CRITERIA

In recent controversies among contending schools and methods, three criteria of the quality of classifications have figured prominently: stability, predictive value, and fit to the true cladistic relationship (Rohlf and Sokal 1981). Taxonomic stability includes the robustness of classifications to the addition of new characters. Methods exhibit greater *character stability* when classifications based on different subsets of a suite of characters, or on different kinds of characters, are more congruent. Robustness of classifications to the addition (or deletion) of OTU's is known as *OTU stability*. A third type of stability is robustness of classifications to methods of coding characters and their states and to the numerical algorithms employed. *Predictive value* is a measure of how similar OTU's are with respect to their character states to the other members of their taxon (at various rank levels). *Fit to the true cladistic relationship* is an optimality criterion only for cladistic classifications. We shall take up this criterion first.

The accuracy of an estimated cladogram can be tested in only those few cases in

which the true phylogeny is known. Although a few studies (e.g., Baum and Estabrook 1978; Baum 1983) have been reported as comparisons of true phylogenies with estimated ones, these comparisons are more accurately described as being of better-documented estimates with less well documented ones. In general, phylogenies of real organisms are unknown. For this reason, tests of the accuracy of cladogram estimation have to be carried out with artificial data sets.

One such data set, which has received considerable attention by systematists, is the Caminalcules (Sokal 1983a). This group of organisms has the great advantage of a known phylogeny. Its disadvantage, however, is that the group is artificial, owing its existence to the fantasy and inventiveness of the late Professor Joseph Camin. For this reason, it has been claimed that inferences made on the basis of the Caminalcules are not valid for the systematics of living organisms. Doubtlessly, any group of real organisms can be differentiated in innumerable ways from the Caminalcules. The comparisons to which I shall refer, however, have been made only with respect to a subset of properties relevant to the analysis of classifications. In a study of such measurable properties of the Caminalcules, I found that for none of the properties examined (homoplasy, symmetry, adequacy of the character states for resolving the cladogram, evolutionary rates, species longevities, and the ratio of speciation to extinction) do the characters of the Caminalcules differ from those of living organisms (Sokal 1983a). In fact, none of the statistics calculated for the Caminalcules are beyond the range of those observed for 19 zoological data sets. In view of these findings, I maintain that, with respect to the properties important for classification, the Caminalcules behave similarly to real organisms, and the burden of proof falls on those critics who wish to ignore the Caminalcules as relevant for biological systematics. Recently, E. W. Holman (pers. comm.) has shown that the Caminalcules deviate from real organisms in not following a hollow curve distribution with respect to the number of Recent species per genus. He concluded that this is due to a variation in evolutionary rates among the lineages of these creatures.

Subjective taxonomists, ranging from distinguished professors to grade school students, have divided the Caminalcules into five major groups, so-called genera, and this classification has been supported by numerical phenetic analysis of the data (Sokal and Rohlf 1980). Figure 5 shows one representative species of each genus. The characters of the Caminalcules were described by persons unaware of the true phylogeny of these organisms. There are 29 Recent species of Caminalcules and 48 fossil species, making 77 species in all for this group. Eighty-five characters were described on the basis of the 29 Recent forms, but 106 characters are necessary to describe all of the species including the fossils (Sokal 1983a).

To discover how well various numerical methods estimate the true phylogeny, we can try numerical phenetic classifications, as well as numerical cladistic estimates of the true cladogram. Phenetic classifications are achieved by means of the usual procedures. Characters are standardized, correlations and taxonomic distances are computed between all pairs of OTU's, and the resemblance matrices are clustered by the UPGMA algorithm (for these methods, see Sneath and Sokal 1973). Cladistic estimates are obtained by common numerical cladistic procedures, which either are based on a hypothesis of parsimony, resulting in the

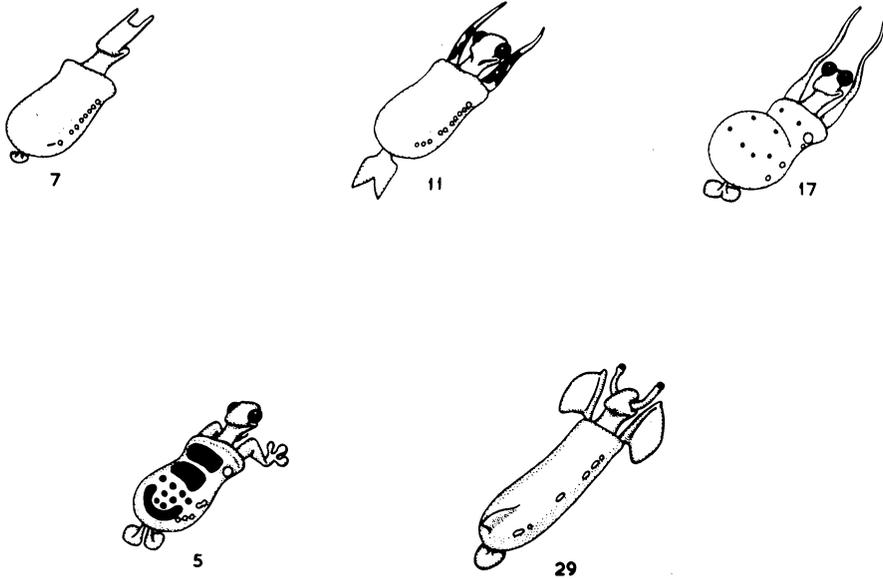


FIG. 5.—Representative species from the five genera of the Caminalcules. The five genera A, B, C, DE, and F are represented, respectively, by the following five numbered species: 7, 11, 17, 5, 29.

shortest obtainable trees, or are based on tree structures that permit the greatest compatibility among character states. I discuss only Wagner parsimony (Farris 1970; Felsenstein 1982), having found this to be the relatively best of the cladistic methods available.

Both the phenograms and the estimated cladograms are compared with the true cladogram. The latter is featured here only in a reduced form in figure 7 below, but can be inspected in Sokal (1983*a,b*, or 1984). To compare the estimated cladograms with the true cladogram, one needs a measure of consensus between the two. This subject is currently an active area of research, and there is a multiplicity of proposed indexes (Rohlf 1982; Shao 1983; Stinebrickner 1984). Although I have applied several of these to the Caminalcules, I present here only the simplest and perhaps also the most conservative index, the strict consensus index CI_C (Colless 1980; Rohlf 1982). When all characters of the Caminalcules are used, one finds that cladistic methods (Wagner trees) estimate the true cladogeny better than phenetic methods (UPGMA trees). With fewer characters, however, phenetic methods give closer estimates to the true cladogram than cladistic methods (Sokal 1983*b*), and this corresponds to the findings of other authors (Colless 1970; Sokal and Rohlf 1981; Tateno et al. 1982; Fiala and Sokal 1985), who have shown that phenetic classifications in many cases are as good as cladistic classifications for estimating the true cladogeny and in some cases provide superior estimates. We shall return to the relation between character number and goodness of estimate presently.

A second important finding is that none of the numerical methods correctly

estimates the entire phylogenetic tree (Sokal 1983*b*). If such work were done on real organisms, one could not expect even the relatively weak consensus that was found, since I assisted the estimate in various ways that could not be repeated in the absence of knowledge of the true cladogeny. An attempt to obtain an estimate of the true cladogeny by conventional, nonnumerical cladistics resulted in a relatively poor approximation of the true tree (Sokal 1983*b*).

Numerical cladistic estimates of the true cladogram are seriously affected by the input order of the OTU's to the computer program; many different combinations yield an array of different tree lengths for any one data set. Only rarely does one know that the shortest length obtained is indeed a tree of minimum length. In the Caminalcules one notes in general that shorter trees tend to give better estimates of the true cladogeny, but this is only a trend. Some of the shortest trees deviate appreciably from the best estimates, and conversely, some of the best estimates of trees are longer than other, poorer estimates. This should not surprise us greatly. The true length of the cladogram of the Caminalcules is 321 evolutionary units, but when the computer optimizes the distribution of character states on the true tree so as to obtain the shortest possible length, one obtains a tree with only 217 units, considerably shorter than the true length. A Wagner-tree program, constructing hypothetical taxonomic units (HTU's) as ancestors in order to minimize the total length of the tree, will, however, produce trees as short as 211 units. Such trees cannot possibly be topologically correct.

It is possible to analyze in detail the differences in results between the true cladogram and various estimates of it obtained by numerical cladistic and numerical phenetic methods. One can examine these relationships by means of a standard three-taxon cladogram in which B and C join before their common stem joins A (see fig. 6). The results of such an endeavor can be summarized as follows. In the absence of homoplasy and additional divergence, phenetic relationships fully correspond to cladistic ones. Phenograms also correspond to the true cladogram in a three-taxon case whenever there is divergence in the outgroup species, divergence in the stems subtending the ingroup, or parallelism between the two ingroup species. Phenetic-cladistic agreement diminishes in response to divergence in either or both ingroup species, parallelism between outgroup and one or both ingroup species, or reversals in one or both ingroup species.

One finds empirically that estimated cladograms obtained by various numerical cladistic methods are also affected by homoplasy and unusual divergence, although not to an equally great extent. When one compares the actual phenetic classifications and estimated cladograms of the Caminalcules with the true cladogram, one notices good correspondence in all such cases in which there is considerable divergence in the stem, setting off the taxon from other taxonomic units that might compete for phenetic affiliation. Thus, in figure 6, whenever length d is substantial by comparison with lengths b and c , we would expect good fits by either phenograms or estimated cladograms to the true tree. Discrepancies from the true cladogeny arise from one of two kinds of situations: parallelisms in the cladogram affect the phenetic similarities of the OTU's; and divergence of cladistically closer relatives increases the relative phenetic similarity of cladistically more distant relatives. These two phenomena occur in more or less equal

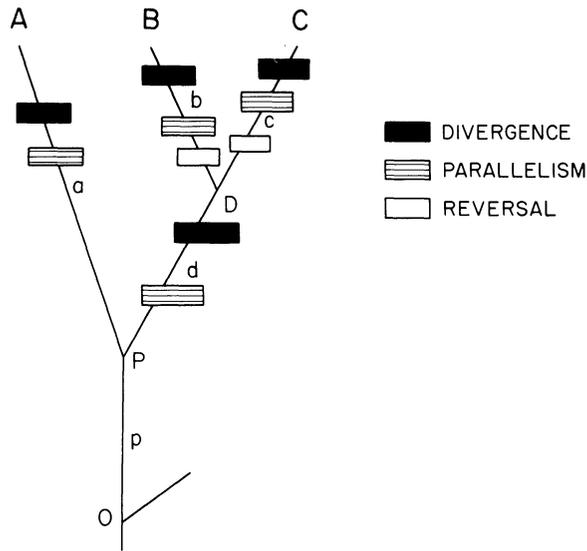


FIG. 6.—A three-taxon cladogram to illustrate the effects of divergence and homoplasy on differences obtained by numerical phenetic and numerical cladistic methods. Capital letters indicate Recent OTU's and ancestors; lowercase letters, the lengths of the corresponding internodes measured in units of evolutionary change. The bars across the internodes show the potential for evolutionary factors that could enhance or diminish the congruence between classifications obtained by phenetic and cladistic methods.

frequency in the Caminalcules, and they affect both the numerical phenetic and the numerical cladistic estimates. These two situations can be illustrated quite easily with examples from the Caminalcules (see fig. 7).

In genus B, OTU's 11 and 21 are cladistically closest, but 21 diverges from 11 by 5 evolutionary units, whereas 11 is only 2 units distant from OTU 10, although the latter branches off earlier. That is why 10 and 11 are closer in a phenogram than 11 is to its sister species, 21. This type of argument can be continued for OTU's 6 and 9, which in a phenogram would next join the nuclear cluster 10–11 in that order. All are closer to 11 and to each other in terms of evolutionary units than they are to OTU 21. That is why they join the phenetic cluster before OTU 21 does. In contrast to these results, those for genus C show complete agreement among the true cladogram (fig. 7), a phenogram, and an estimated cladogram (Sokal 1983*b*). The evolutionary changes in the stems leading to the bifurcations in this genus are so great that they ensure the phenetic similarity of sister species.

An additional recent finding in our laboratory (Fiala and Sokal 1985) supports these conclusions. In a series of simulation studies in which different tree topologies were given different patterns of character-state evolution based on various evolutionary assumptions, we found that, on the average, phenograms gave estimates of the true cladogeny that were as accurate as numerically estimated cladograms. The relative successes of the methods have to do with the topology of the trees. Those trees that have relatively long basal stems vis-à-vis short issuing branches are estimated much better than those with short joint stems and long

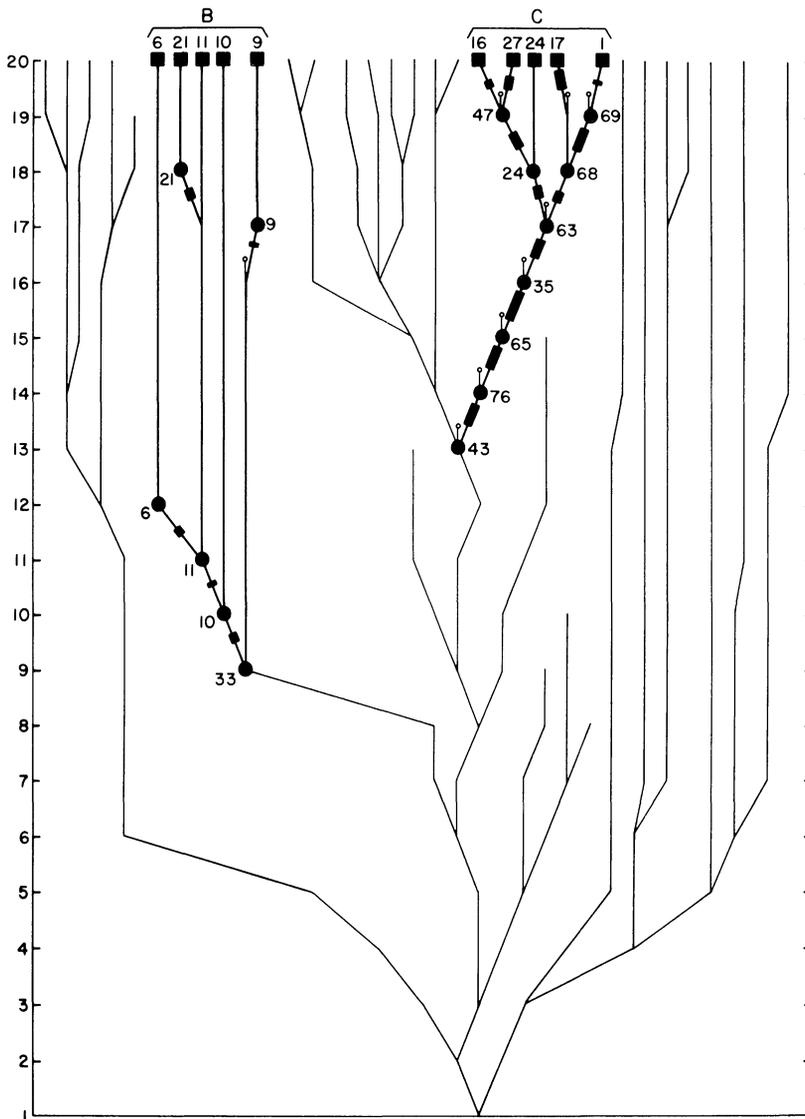


FIG. 7.—Subtrees for genera B and C from the true cladogram of the Caminalcules. Morphological change (slanted lines) in these genera occurred from time period 8 (B) or 13 (C) onward; vertical lines indicate periods without such change. The amount of change (path length of the internode) based on 85 characters is shown by the lengths of the thickened bars along the slanted lines. Species are identified by numbers: squares identify Recent species; circles, fossil species. Small hollow circles represent extinct species whose lineages continue with evolutionary change.

issuing branches. This "stemminess" has a marked effect in our simulation work. It therefore appears that the results of cladistic estimation methods greatly depend on the unknown topology of the true evolutionary tree. Additional problems with estimating cladograms by parsimony methods have recently been described by Rohlf (1984). Peculiarities in character-state codes result in dramatic instabilities of estimated cladograms by means of minimum-length Wagner trees.

It appears common wisdom among systematists that if only fossils were available, it would be easy to reconstruct phylogenetic trees and to establish classifications. When we include all 48 fossil species with the 29 Recent ones in the Caminalcules, the results are surprisingly no better than the analysis of Recent OTU's alone (Sokal 1983c). Even the best estimated cladogram has a strict consensus index with the true cladogram of only 0.667. The phenetic classification of these OTU's introduces some nonconvex taxa at higher phenetic levels, but unites phenetically homogeneous groups of mixed Recent and fossil composition. There is good correspondence of phenetics with phylogenetic sequences. All but one of the mutually closest pairs in the phenogram are ancestor-descendant pairs. Yet, it would not be possible to piece together a true phylogenetic tree of the Caminalcules by using these short sequences, even if their polarity were known, which is not always the case.

The first of the purposes of taxonomy, enumerated earlier, is the establishment of a natural system, available for reference to the scientific public and to laymen. Stability is clearly desirable for such a natural system. Actual tests of taxonomic stability are not usually carried out by adding either characters or OTU's to a data matrix. Rather, they are accomplished by taking the most complete available data matrix for the group and randomly subsampling characters or OTU's from it. From each subsample one computes a phenogram and also an estimated cladogram, and compares each with other such subsamples or with a standard dendrogram, that is, with the phenogram or the cladogram based on the entire data set.

Tests of stability carried out in this manner in the Caminalcules are of considerable interest. When we analyze the entire number of characters, estimated cladograms are more stable than phenograms when OTU numbers are small; phenograms are more stable than cladograms when OTU numbers become large (Sokal 1983d). These results can be inspected in figure 8, where each of the points is computed on the basis of at least 100 random samples. The ordinate indicates the ratio $n'/(2t - 3)$, where n' is the number of characters (in binary equivalents) and t , the number of OTU's, also defines the abscissa. For some selected sample sizes of characters (n , not binary), curves show the ratio as a function of OTU number t . The circles indicate the results of the comparisons. White denotes a superiority of estimated cladograms; black, a superiority of phenograms. The upper semicircle furnishes results on stability against a standard; the lower semicircle shows the results of comparisons of classifications with the true cladogram. For the combination 29 OTU's and 85 characters, the upper semicircle is missing because a stability test is meaningless for the full data set. When the number of characters in the sample is small, or the OTU number is large, phenograms are more stable and provide better estimates of the true cladogeny. But as the number of characters increases and/or the number of OTU's decreases, estimated cladograms improve

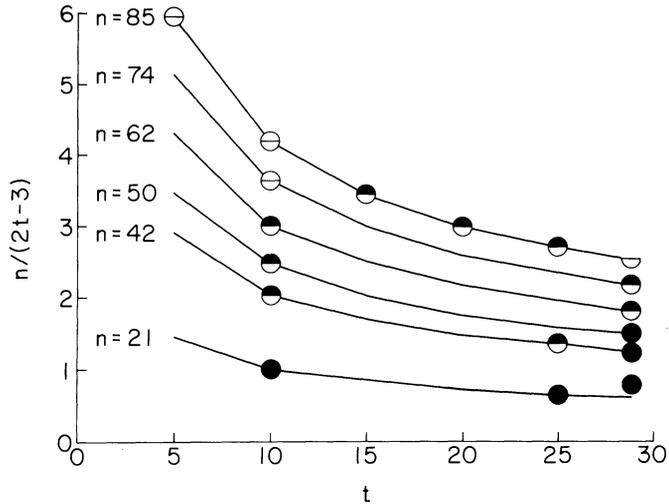


FIG. 8.—Phenograms and estimated cladograms for the Caminalcules compared in two ways: their stability tested against a standard, and their congruence compared with the true cladogeny. (For explanation, see the text.)

and ultimately overtake phenograms in terms of stability and as estimators of the true cladogeny.

It was noted eventually that the stability is a function of the ratio of the number of characters and OTU's (Sokal et al. 1984). If t is the number of OTU's and n the number of characters, cladograms have $2t - 3$ edges, which must be estimated, whereas phenograms need only $t - 1$ parameters to determine the junction levels. The ratio $n/(2t - 3)$ is a coarse indicator of the adequacy of the data matrix for cladistic estimates. When this value is above 1, stability is higher for classifications based on estimated cladograms; below 1, classifications based on phenograms are more stable. These results parallel our experiences with estimates of the true cladogram. For all 29 Recent OTU's and 50 or fewer characters, one can estimate the true cladogram better by means of a phenogram; for more than 50 characters the relation is reversed. Similar experiences were obtained on a group of real organisms, bugs of the genus *Leptopodomorpha* (Sokal et al. 1984). Here, of course, one can measure only stabilities, since the true cladogram is unknown. Similar results are arrived at for various other data sets: a study of proteins in mice (W. M. Fitch and W. R. Atchley, pers. comm.), two simulation studies (Fiala and Sokal 1985; J. Sourdis and C. Krimbas, MS), and an assemblage of 39 random and real data sets (Sokal and Shao 1985).

Taxonomic congruence is a special case of character stability. It is a measure of the similarity of classifications based on different kinds or classes of characters, such as external versus internal characters, morphological versus behavioral characters, and so forth. Here, too, the results of critical tests show varying outcomes in favor of phenetic or cladistic approaches (Rohlf et al. 1983). The outcome of congruence tests is undoubtedly also affected by the ratio of the numbers of characters and OTU's and by stemminess.

The measurement of predictive value in classifications has not so far been uniformly successful. It is difficult to decide whether predictive value should be computed separately for each taxon or globally for the entire classification. The hierarchical nature of biological classifications has also complicated the development of measures of predictive value. A good measure must allow for the several taxonomic levels of a given study, since any one character might be highly predictive at one categoric level but of little value at another level. A second important consideration is whether the predictive value should measure only errors of inclusion or both errors of inclusion and of exclusion. Errors of inclusion imply inhomogeneity of character states within a taxon. Errors of exclusion mean that character states found within a given taxon are also found in other taxa and, to this degree, decrease the predictive value of the given character state. Unanimity on this issue has been difficult to achieve, and the debate has been clouded by the controversy between pheneticists and cladists. In a recent study, Archie (1984) has clarified some of the concepts related to predictive value and developed new indexes for measuring this property of classifications. It is probably fair to say that at this time it is not clear which school of taxonomy produces classifications with higher predictive value.

CONCLUSIONS

How can all these results be summarized? Numerically estimated cladograms are not good estimates of the true phylogeny of a group of organisms. The shortest trees are not necessarily closest to the true tree. Differences between true cladograms and phenograms or between phenograms and estimated cladograms can be explained as the results of homoplasy or divergence. Estimated cladograms are affected almost as much by homoplasy as are phenograms. As the number of characters decreases or the number of OTU's increases, phenograms become better estimates of the true cladogeny than estimated cladograms. These same relations exist for taxonomic stability based on either characters or OTU's. Even the inclusion of fossils in the data matrix does not substantially increase the quality of the estimate of the phylogeny. The topology of the true tree is a critical factor in determining the quality of its estimate. Such results are not causes for optimism for those who wish to estimate phylogenies.

One should be cautious about establishing classifications on the basis of few characters, as is often done by conventional (i.e., nonnumerical) cladists. Under less than optimal conditions (i.e., with few characters or many OTU's), it would appear that a phenetic classification would continue to be the most desirable system for establishing general classifications because it permits greater stability and possibly greater predictive value than a cladistic classification. For those whose major purpose is to estimate phylogenies, phenograms would in many cases give estimates of the true cladogeny that are no worse, and possibly better, than those obtained by current cladistic methods.

Finally, it is now entirely obvious that almost all scientists who so desire have access to a computer. Data matrices in systematics can therefore be analyzed in a great variety of ways, with little effort and at relatively low cost. Alternative

systems of classification are clearly feasible. Scientists can analyze their data phenetically as well as cladistically, and then synthesize the two to arrive at evolutionary classifications. Nevertheless, a general system is still desirable; and for the reasons stated above, it would seem that such a system should be phenetic.

SUMMARY

In this paper I review the principles for forming biological classifications and summarize recent findings concerning optimality criteria for classifications. Natural taxa are recognized as polythetic and related to concept formation in cognitive psychology. The three currently advocated schools of taxonomy are reviewed and their assumptions and purposes compared. Three criteria of optimality—predictive value, stability, and fit to the true cladistic relationship—are discussed, and evidence from recent numerical taxonomic studies of these criteria is reviewed. Numerical classifications based on phenetic and cladistic computer programs differ in their taxonomic stability and fit to the true cladogram. There is no universally superior approach, but the relative advantage of phenetic versus cladistic algorithms is a function of the ratio of characters to OTU's in the data. Accuracy of cladogram estimation is also affected by tree topology.

ACKNOWLEDGMENTS

This article is an expanded version of the presidential address to the American Society of Naturalists presented in June 1984 at Crested Butte, Colorado. The assistance of K. Fiala and K. T. Shao, G. Hart, and B. Thomson in the research leading up to this work is gratefully acknowledged. F. J. Rohlf read the manuscript and furnished useful suggestions for improvements. R. Chapey prepared the manuscript and J. Schirmer drew the figures. The research upon which this article is based was supported by National Science Foundation grant BSR 83-06004. Contribution no. 556 in Ecology and Evolution from the State University of New York at Stony Brook.

LITERATURE CITED

- Archie, J. W. 1984. A new look at the predictive value of numerical classifications. *Syst. Zool.* 33: 30–51.
- Ashlock, P. H. 1980. An evolutionary taxonomist's view of classification. *Syst. Zool.* 28:441–450.
- Baum, B. R. 1983. Relationships between transformation series and some numerical cladistic methods at the infraspecific level, when genealogies are known. Pages 340–345 in J. Felsenstein, ed. *Numerical taxonomy*. Springer-Verlag, Berlin.
- Baum, B. R., and G. F. Estabrook. 1978. Application of compatibility analysis in numerical cladistics at the infraspecific level. *Can. J. Bot.* 56:1130–1135.
- Beatty, J. 1982. Classes and cladists. *Syst. Zool.* 31:25–34.
- Beckner, M. 1959. *The biological way of thought*. Columbia University Press, New York.
- Bock, W. J. 1977. Foundations and methods of evolutionary classification. Pages 851–895 in M. K. Hecht, P. C. Goody, and B. M. Hecht, eds. *Major patterns of vertebrate evolution*. Plenum, New York.
- Breedlove, D. E., and P. H. Raven. 1974. *Principles of Tzeltal plant classification: an introduction to*

- the botanical ethnography of a Mayan speaking people of highland Chiapas. Academic Press, New York.
- Brooks, D. R., and E. O. Wiley. 1985. Theories and methods in different approaches to phylogenetic systematics. *Cladistics* 1:1–11.
- Cain, A. J. 1958. Logic and memory in Linnaeus's system of taxonomy. *Proc. Linn. Soc. Lond.* 169:144–163.
- Cain, A. J., and G. A. Harrison. 1960. Phyletic weighting. *Proc. Zool. Soc. Lond.* 135:1–31.
- Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19:311–326.
- Colless, D. H. 1970. The phenogram as an estimate of phylogeny. *Syst. Zool.* 19:352–362.
- . 1980. Congruence between morphometric and allozyme data for the *Menidia* species: a reappraisal. *Syst. Zool.* 29:288–299.
- Cracraft, J. 1983. The significance of phylogenetic classifications for systematic and evolutionary biology. Pages 1–17 in J. Felsenstein, ed. *Numerical taxonomy*. Springer-Verlag, Berlin.
- DuPraw, E. J. 1964. Non-Linnaean taxonomy. *Nature (Lond.)* 202:849–852.
- Eldredge, N., and J. Cracraft. 1980. *Phylogenetic patterns and the evolutionary process*. Columbia University Press, New York.
- Estabrook, G. F. 1978. Some concepts for the estimation of evolutionary relationships in systematic botany. *Syst. Bot.* 3:146–158.
- Farris, J. S. 1970. Methods for computing Wagner trees. *Syst. Zool.* 19:83–92.
- Farris, J. S., A. G. Kluge, and M. J. Eckardt. 1970. A numerical approach to phylogenetic systematics. *Syst. Zool.* 19:172–189.
- Felsenstein, J. 1982. Numerical methods for inferring evolutionary trees. *Q. Rev. Biol.* 57:379–404.
- Fiala, K. L., and R. R. Sokal. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. *Evolution* 39:609–622.
- Fitch, W. M., and E. Margoliash. 1968. The construction of phylogenetic trees. II. How well do they reflect past history? *Brookhaven Symp. Biol.* 21:217–242.
- Gilmour, J. S. L. 1937. A taxonomic problem. *Nature (Lond.)* 139:1040–1042.
- . 1940. Taxonomy and philosophy. Pages 461–474 in J. Huxley, ed. *The new systematics*. Clarendon, Oxford.
- . 1951. The development of taxonomic theory since 1851. *Nature (Lond.)* 168:400–402.
- . 1961. Taxonomy. Pages 27–45 in A. M. MacLeod and L. S. Copley, eds. *Contemporary botanical thought*. Oliver & Boyd, Edinburgh.
- Kluge, A. G. 1985. Ontogeny and phylogenetic systematics. *Cladistics* 1:13–27.
- Mayr, E. 1969. *Principles of systematic zoology*. McGraw-Hill, New York.
- . 1982. *The growth of biological thought*. Belknap, Cambridge, Mass.
- Neel, J. V. 1978. The population structure of an Amerindian tribe, the Yanomama. *Annu. Rev. Genet.* 12:365–413.
- Platnick, N. I. 1985. Philosophy and the transformation of cladistics revisited. *Cladistics* 1:87–94.
- Remane, A. 1956. *Die Grundlagen des natuerlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Theoretische Morphologie und Systematik. Vol. I.* 2d ed. Akademische Verlagsgesellschaft Geest & Portig, Leipzig.
- Riedl, R. 1976. *Die Strategie der Genesis*. Piper, Munich.
- Rohlf, F. J. 1982. Consensus indices for comparing classifications. *Math. Biosci.* 59:131–144.
- . 1984. A note on minimum length trees. *Syst. Zool.* 33:341–343.
- Rohlf, F. J., and R. R. Sokal. 1981. Comparing numerical taxonomic studies. *Syst. Zool.* 30:459–490.
- Rohlf, F. J., D. H. Colless, and G. Hart. 1983. Taxonomic congruence—reexamined. *Syst. Zool.* 32:144–158.
- Shao, K. T. 1983. *Consensus methods in numerical taxonomy*. Ph.D. diss. State University of New York, Stony Brook.
- Smith, E. E., and D. L. Medin. 1981. *Categories and concepts*. Harvard University Press, Cambridge, Mass.
- Sneath, P. H. A. 1962. The construction of taxonomic groups. *Symp. Soc. Gen. Microbiol.* 12:289–332.
- Sneath, P. H. A., and R. R. Sokal. 1973. *Numerical taxonomy*. Freeman, San Francisco.

- Sokal, R. R. 1974. Classification: purposes, principles, progress, prospects. *Science* (Wash., D.C.) 185:1115–1123.
- . 1983*a*. A phylogenetic analysis of the Caminalcules. I. The data base. *Syst. Zool.* 32:159–184.
- . 1983*b*. A phylogenetic analysis of the Caminalcules. II. Estimating the true cladogram. *Syst. Zool.* 32:185–201.
- . 1983*c*. A phylogenetic analysis of the Caminalcules. III. Fossils and classification. *Syst. Zool.* 32:248–258.
- . 1983*d*. A phylogenetic analysis of the Caminalcules. IV. Congruence and character stability. *Syst. Zool.* 32:259–275.
- . 1984. Die Caminalcules als taxonomische Lehrmeister. Pages 15–31 in H. H. Bock, ed. *Anwendungen der Klassifikation: Datenanalyse und numerische Klassifikation*. Indeks Verlag, Frankfurt.
- Sokal, R. R., and F. J. Rohlf. 1970. The intelligent ignoramus, an experiment in numerical taxonomy. *Taxon* 19:305–319.
- . 1980. An experiment in taxonomic judgment. *Syst. Bot.* 5:341–365.
- . 1981. Taxonomic congruence in the Leptopodomorpha re-examined. *Syst. Zool.* 30:309–325.
- Sokal, R. R., and K. T. Shao. 1985. Character stability in 39 data sets. *Syst. Zool.* 34:83–89.
- Sokal, R. R., and P. H. A. Sneath. 1963. *Principles of numerical taxonomy*. Freeman, San Francisco.
- Sokal, R. R., K. L. Fiala, and G. Hart. 1984. OTU stability and factors determining taxonomic stability: examples from the Caminalcules and the Leptopodomorpha. *Syst. Zool.* 33:387–407.
- Stinebrickner, R. 1984. An extension of intersection methods from trees to dendrograms. *Syst. Zool.* 33:381–386.
- Tateno, Y., M. Nei, and F. Tajima. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *J. Mol. Evol.* 18:387–404.
- Wiley, E. O. 1981. *Phylogenetics*. Wiley, New York.