Biological Classification: Toward a Synthesis of Opposing Methodologies

Ernst Mayr

Summary. Currently a controversy is raging as to which of three competing methodologies of biological classification is the best: phenetics, cladistics, or evolutionary classification. The merits and seeming deficiencies of the three approaches are analyzed. Since classifying is a multiple-step procedure, it is suggested that the best components of the three methods be used at each step. By such a synthetic approach, classifications can be constructed that are equally suited as the basis of generalizations and as an index to information storage and retrieval systems.

For nearly a century after the publication of Darwin's *Origin* (*1*) no well-defined schools of classifiers were recognizable. There were no competing methodologies. Taxonomists were unanimous in their endeavor to establish classifications that would reflect "degree of relationship." What differences there were among competing classifications concerned the number and kinds of characters that were used, whether or not an author accepted the principle of recapitulation, whether he attempted to "base his classification on phylogeny," and to what extent he used the fossil record (*2*). As a result of a lack of methodology, radically different classifications were sometimes proposed for the same group of organisms; also new classifications were introduced without any adequate justification except for the claim that they were "better." Dissatisfaction with such arbitrariness and seeming absence of any carefully thought out methodology, led in the 1950's and 1960's to the establishment of two new schools of taxonomy, numerical phenetics and cladistics, and to a more explicit articulation of Darwin's methodology, now referred to as evolutionary classification.

The Major Schools of Taxonomy

Numerical phenetics. From the earliest preliterary days, organisms were grouped into classes by their outward appearance, into grasses, birds, butterflies, snails, and others. Such grouping "by inspection" is the expressly stated or unspoken starting point of virtually all systems of classification. Any classification incorporating the method of grouping taxa by similarity is, to that extent, phenetic.

In the 1950's to 1960's several investigators went one step further and suggested that classifications be based exclusively on "overall similarity." They also proposed, in order to make the method more objective, that every character be given equal weight, even though this would require the use of large numbers of characters (preferably well over a hundred). In order to reduce the values of so many characters to a single measure of "overall similarity," each character is to be recorded in numerical form. Finally, the clustering of species and their taxonomic distance from each other is to be calculated by the use of algorithms that operationally manipulate characters in certain ways, usually with the help of computers. The resulting diagram of relationship is called a phenogram. The calculated phenetic distances can be converted directly into a classification.

The fullest statement of this methodology and its underlying conceptualization was provided by Sokal and Sneath (*3*). They called their approach "numerical taxonomy," a somewhat misleading designation, since numerical methods, including numerical weighting, can be and have been applied to entirely different approaches to classification. The term numerical phenetics is now usually applied to this school. This has introduced some ambiguity since some authors have used the term phenetic broadly, applying it to any approach making use of the "similarity" of species and other taxa, while to the strict numerical pheneticists the term phenetic means the "theory-free" use of unweighted characters.

Cladistics (or cladism). This method of classification (*4*), the first comprehensive statement of which was published in 1950 by Hennig (*5*), bases classifications exclusively on genealogy, that is, on the branching pattern of phylogeny. For the cladist phylogeny consists of a sequence of dichotomies (*6*), each representing the splitting of a parental species into two daughter species; the ancestral species ceases to exist at the time of the dichotomy; sister groups must be given the same categorical rank; and the ancestral species together with all of its descendants must be included in a single "holophyletic" taxon.

Evolutionary classification. Phenetics and cladistics were proposed in the endeavor to replace the methodology of classification that had prevailed ever since Darwin and that was variously designated as the "traditional" or the "evolutionary" school, which bases its classifications on observed similarities and differences among groups of organisms, evaluated in the light of their inferred evolutionary history (*7*). The evolutionary school includes in the analysis all available attributes of these organisms, their correlations, ecological stations, and patterns of distributions and attempts to reflect both of the major evolutionary processes, branching and the subsequent diverging of the branches (clades). This school follows Darwin (and agrees in this point with the cladists) that classification must be based on genealogy and also agrees with Darwin (in contrast to the cladists) "that genealogy by itself does not give classification" (*8*).

The results of the evolutionary analysis are incorporated in a diagram, called a phylogram, which records both the branching points and the degrees of subsequent divergence. The method of inferring genealogical relationship with the help of taxonomic characters, as it was first carried out by Darwin, is an application of the hypothetico-deductive approach. Presumed relationships have to be tested again and again with the help of new characters, and the new evidence frequently leads to a revision of the inferences on relationship. This method is not circular (*9*) as has sometimes been suggested.

The author is Alexander Agassiz Professor Emeritus, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138.

0036-8075/81/1030-0510$01.00/0 Copyright © 1981 AAAS

Is There a Best Way to Classify?

Each of the three approaches to classification—phenetics, cladistics, and evolutionary classification—has virtues and weaknesses. The ideal classification would be one that would meet best as many as possible of the generally acknowledged objectives of a classification.

A biological classification, like any other, must serve as the basis of a convenient information storage and retrieval system. Since all three theories produce hierarchical systems, containing nested sets of subordinated taxa, they permit the following of information up and down the phyletic tree. But this is where the agreement among the three methods ends. Purely phenetic systems, derived from a single set of arbitrarily chosen characters, sometimes provide only low retrieval capacity as soon as other sets of characters are used. The effectiveness of the phenetic method could be improved by careful choice of selected characters. However, the method would then no longer be "automatic," because any selection of characters amounts to weighting.

Cladists use only as much information for the construction of the classification as is contained in the cladogram. They convert cladograms, quite unaltered, into classifications, only when the cladograms are strictly dichotomous. Even though cladists lose much information by this simplistic approach, the information on lines of descent can be read off their classifications directly. However, a neglect of all ancestral-descendant information reduces the heuristic value of their classifications. By contrast, since evolutionary taxonomists incorporate a great deal more information in their classifications than do the cladists, they cannot express all of it directly in the names and ranking of the taxa in their classifications. Therefore, they consider a classification simply to be an ordered index that refers them to the information that is stored elsewhere (in the detailed taxonomic treatments).

A far more important function of a classification, even though largely compatible with the informational one, is that it establishes groupings about which generalizations can be made. To the extent that classifications are explicitly based on the theory of common descent with modification, they postulate that members of a taxon share a common heritage and thus will have many characteristics in common. Such classifications, therefore, have great heuristic value in all comparative studies. The validity of specific observations can be generalized by testing them against other taxa in the system or against other kinds of characters (*10–12*).

Pheneticists, as well as cladists, have claimed that their methods of constructing classifications are nonarbitrary, automatic, and repeatable. The criticisms of these methods over the last 15 years (*13*) have shown, however, that these claims cannot be substantiated. It is becoming increasingly evident that a one-sided methodology cannot achieve all the above-listed objectives of a good classification.

The silent assumption in the methodologies of phenetics and cladistics is that classification is essentially a single-step procedure: clustering by similarity in phenetics, and establishment of branching patterns in cladistics. Actually a classification follows a sequence of steps, and different methods and concepts are pertinent at each of the consecutive steps. It seems to me that we might arrive at a less vulnerable methodology by developing the best method for each step consecutively. Perhaps the steps could eventually be combined in a single algorithm. In the meantime, their separate discussion contributes to the clarification of the various aspects of the classifying process.

Establishment of Similarity Classes

The first step is the grouping of species and genera by "inspection," that is, by a phenetic procedure. (I use phenetic in the broadest sense, not in the narrow one of numerical phenetics.) All of classifying consists of, or at least begins with, the establishment of similarity classes, such as a preliminary grouping of plants into trees, shrubs, herbs, and grasses. The reason why the method is so often successful is simply that—other things being equal—descendants of a common ancestor tend to be more similar to each other than they are to species that do not share immediate common descent. The method is thus excellent in principle. Numerical phenetics has nevertheless proved to be largely unsuccessful because (i) claims, such as "results objective and strictly repeatable," were not always justifiable since in practice different results are obtained when different characters are chosen or different programs of computation are used; (ii) the method was inconsistent in its claim of objectivity since subjective biological criteria were used in the assigning of variants (for example, sexes, age classes, and morphs) to "operational taxonomic units" (OTU's); and, most importantly, the method insisted on the equal weighting of all characters.

It is now evident that no computing method exists that can determine "true similarity" from a set of arbitrarily chosen characters. So-called similarity is a complex phenomenon that is not necessarily closely correlated with common descent, since similarity is often due to convergence. Most major improvements in plant and animal classifications have been due to the discovery of such convergence (*14*).

Different types of characters—morphological characters, chromosomal differences, enzyme genes, regulatory genes, and DNA matching—may lead to rather different grouping. Different stages in the life cycle may result in different groupings.

The ideal of phenetics has always been to discover a measure of total (overall) similarity. Since it is now evident that this cannot be achieved on the basis of a set of arbitrarily chosen characters, the question has been asked whether there is not a method to measure degrees of difference of the genotype as a whole. Improvements in the method of DNA hybridization offer hope that this method might give realistic classifications on a phenetic basis, at least up to the level of orders (*15*). The larger the fraction of the nonhybridizing DNA, the less reliable this method is, because it cannot be determined whether the nonmatching DNA is only slightly or drastically different.

Testing the Naturalness of Taxa

In the first step of the classifying procedure clusters of species were assembled that seemed to be more similar to each other than to species in other clusters. These clusters are the taxa we recognize tentatively (*16*). In order to make these clusters conform to evolutionary theory, two, operationally more or less inseparable tests, must be made: (i) determine for all species of a cluster (taxon) whether they are descendants of the nearest common ancestor and (ii) connect the taxa by a branching tree of common descent, that is, construct a cladogram. An indispensable preliminary of this testing is an analysis of the characters used to establish the similarity clusters.

Character analysis. A careful analysis shows almost invariably that some characters are better clues to relationship (have greater weight) than others. The fewer the number of available characters, the more carefully the weighting must be done. This weighting is one of the most controversial aspects of the classifying procedure. Investigators who come to systematics from the outside, say from mathematics, or who are beginners tend to demand objective or quantitative methods of weighting. There are such methods, principally ones based on the covariation of characters, but they are not nearly as informative as methods based on the biological evaluation of characters (*17*). But such an evaluation requires an understanding of many aspects of the to-be-classified group (that is, its life history, the inferred selection pressures to which it is exposed, and its evolutionary history) that may not be available to an outsider. This creates a genuine dilemma. If strictly taxometric methods were available that would produce satisfactory weighting, everyone would surely prefer them to weighting based on experience and biological knowledge. But so far such methods are still in their infancy.

The greatest difficulty for a purely phenetic method, indeed for any method of classification, is the discordance (noncongruence) of different sets of characters. Entirely different classifications may result from the use of characters of different stages of the life cycle as, for instance, larval versus adult characters. In a study of species of bees, Michener (*18*) obtained four different classifications when he sorted them into similarity classes on the basis of the characters of (i) larvae, (ii) pupae, (iii) the external morphology of the adults, and (iv) male genitalic structures. Phenetic delimitation of taxa unavoidably necessitates a great deal of decision-making on the use and weighting of characters. Often, when new sets of characters become available, their use may lead to a new delimitation of taxa or to a change in ranking.

Determination of the genealogy. Each group (taxon) tentatively established by the phenetic method is, so to speak, a hypothesis as to common descent, the validity of which must be tested. Is the delimited taxon truly monophyletic (*19*)? Are the species included in this taxon nearest relatives (descendants of the nearest common ancestor)? Have all species been excluded that are only superficially or convergently similar?

Methods to answer these questions have been in use since the days of Darwin, particularly the testing of the homology of critical characteristics of the included species. However, Hennig (*5*) was the first to articulate such methods explicitly, and these have been modified by some of his followers. These methods can be designated as the cladistic analysis.

Such an analysis involves first the partitioning of the joint characters of a group into ancestral ("plesiomorph" in Hennig's terminology) characters and derived ("apomorph") characters, that is, characters restricted to the descendants of the putative nearest common ancestor (*20*). The joint possession of homologous derived characters proves the common ancestry of a given set of species. A character is derived in relation to the ancestral condition of the character. The end product of such a cladistic character analysis is a cladogram, that is, a diagram (dendrogram) of the branching points of the phylogeny.

Although this procedure sounds simple, numerous practical difficulties have been pointed out (*21, 22*). Very often the branching points are inferred by way of single or very few characters and are affected by all the weaknesses of single character classifications. More serious are two other difficulties.

1) *Polarity*. A derived character is often simpler or less specialized than the ancestral condition. For this reason it can be difficult to determine polarity in a transformation series of characters, that is, to determine which end of the series is ancestral. Tattersall and Eldredge (*23*) stressed that "in practice it is hard, even impossible, to marshall a strong, logical argument for a given polarity for many characters in a given group." Are they primitive (ancestral) or derived? Much of the controversy concerning the phylogeny of the invertebrates, for instance, is due to differences of opinion concerning polarity. Hennig tried to elaborate methods for determining polarity but, as others (*24, 25*) have shown, with rather indifferent success. Since characters come and go in phyletic lines and since there is much convergence, the problem of polarity can rarely be solved unequivocally. There are three best types of evidence for polarity reconstruction. First is the fossil record. Although primitiveness and apparent ancientness are not correlated in every case, nevertheless as Simpson (*26*) stressed, "for any group with even a fair fossil record there is seldom any doubt that characters usual or shared by older members are almost always more primitive than those of later members." Second is sequential constraints. Consecutive chromosomal inversions (as in *Drosophila*) or sets of amino acid replacements (and presumably certain other molecular events) form definite sequences. Which end of the sequence is the beginning can usually not be read off from the sequence itself, but additional information (polarity of other character chains, geographical distribution, and the like) often permits an unequivocal determination of the polarity. Third is the reconstruction of the presumed evolutionary pathway. This can sometimes be done by studying evidence for adaptive shifts, the invasion of new competitors or the extinction of old ones, the behavior of correlated characters, and other biological evidence (*11*, pp. 886–887; *24*). Particular difficulties are posed when the polarity is reversed in the course of evolution, as documented in the fossil record.

2) *Kinds of derived characters*. Two taxa may resemble each other in a given character for one of three reasons: because the character existed already in the ancestry of the two groups before the evolution of the nearest common ancestor (symplesiomorphy in Hennig's terminology), because it originated in the common ancestor and is shared by all of his descendants (homologous apomorphy or synapomorphy), or because it originated independently by convergence in several descendant groups (nonhomologous or convergent apomorphy) (*27*). Since, according to the cladistic method, sister groups are recognized by the possession of synapomorphies, convergence poses a major problem. How are we to distinguish between homologous and convergent apomorphies? Hennig was fully aware of the critical importance of this problem, but it has been quietly ignored by many of his followers. Both grebes and loons, two orders of diving birds, have a prominent spur on the knee and were therefore called sister groups by one cladist. However, other anatomical and biochemical differences between the two taxa indicate that the shared derived feature was acquired by convergence. The reliability of the determination of monophyly of a group depends to a large extent on the care that is taken in discriminating between these two classes of shared apomorphy (*11*, pp. 880–890).

There is a third class of derived characters, so-called autapomorphies, which are characters that were acquired by and are restricted to a phyletic line after it branched off from its sister group.

The pheneticists do not undertake a character analysis. Cladists and evolutionary taxonomists agree with each other in principle on the importance of a careful character analysis. They disagree, however, fundamentally in how to use the findings of the character analysis in the construction of classifications, particularly the ranking procedure.

The Construction of a Classification

Cladistic classification. Cladists convert the cladogram directly into a cladistic classification. In such a classification taxa are delimited exclusively by holophyly, that is, by the possession of a common ancestor, rather than by a combination of genealogy and degree of divergence (*19*). This results in such incongruous combinations as a taxon containing only crocodiles and birds, or one containing only lice and one family of Mallophaga.

Taxa based exclusively on genealogy are of limited use in most biological comparisons. Since, as Hull (*28*) pointed out, cladists really classify characters rather than organisms, they have to make the arbitrary assumption that new apomorph characters originate whenever a line branches from its sister line. This is unlikely in most cases. Surely the reptilian species that originated the avian lineage lacked any of the flight specializations characteristic of modern birds, except perhaps the feathers (*29*).

Two principles govern the conversion of a cladogram into a cladistic classification: (i) all branchings are bifurcations that give rise to two sister groups, and (ii) branchings are usually connected with a change in categorical rank. Cladistic classifications are only representations of branching patterns, with complete disregard of evolutionary divergence, ancestor-descendant relationships, and the information content of autapomorph characters. Because these aspects of evolutionary change are neglected, the cladistic method of classification "either results in lumping very similar forms (parasites and their relatives) or in recognizing a multitude of taxa (perhaps also of other categories) regardless of the extreme similarity of some of them. Such simplistic procedures do violence to most biological attributes other than the pattern of the cladistic branching system, as well as to the function of a classification for convenient information transmittal and storage," as Michener remarked (*18*).

These objections show that the methodology of cladistic classification is not satisfactory. Anyone familiar with the history of taxonomy is strangely reminded of the principles of Aristotelian logical division when encountering cladistic classifications with their rigid dichotomies, the mandate that every taxon must have a sister group, and the principle of a straight-line hierarchy.

There has been much argument over the relationship between classification and phylogeny (*30*). Both cladists and evolutionary taxonomists agree that all members of a taxon must have a common ancestor. A phylogenetic analysis, and in particular a clear separation of homologous apomorphies from convergences, is a necessary component of the classifying procedure. Classificatory analysis often leads to new inferences on phylogeny, and new insights on phylogeny may necessitate changes in classification. These interactions are not in the least circular (*9*).

It is quite unnecessary in most cases to know the exact species that was the common ancestor of two diverging phyletic lines. An inability to specify such an ancestral species has rarely impeded paleontological research (*31, 32*). For instance, it is of little importance whether *Archaeopteryx* was the first real ancestor of modern birds or some other similar species or genus. What is important to know is whether birds evolved from lizard-like, crocodile-like, or dinosaur-like ancestors. If a reasonably good fossil record is available, it is usually possible, by the backward tracing of evolutionary trends and by the backward projection of divergent phyletic lines, to reconstruct a reasonably convincing facsimile of the representative of a phyletic line at an earlier time.

Simpson (*32*) has provided us with cogent arguments about why it is not permissible to reject information from the fossil record under the pretext that it fails to give the phylogenetic connections between fossil and recent taxa with absolute certainty. Hence, there is no merit in the suggestion to construct separate classifications for recent and for fossil organisms. After all, fossil species belong to the same tree of descent as living species. Indeed, enough evidence usually becomes available through a careful character analysis to permit relatively robust inferences on the most probable phylogeny. A number of recent endeavors have been made to develop a cladistic methodology that is quantitative and automatic. New methods in this area are published in rapid succession and it would seem too early to determine which is most successful and freest of possible flaws (*33*).

Evolutionary classification. The taxonomic task of the cladist is completed with the cladistic character analysis. The genealogy gives him the classification directly, since for him classification is nothing but genealogy. The evolutionary taxonomist carries the analysis one step further. He is interested not only in branching, but, like Darwin, also in the subsequent fate of each branch. In particular, he undertakes a comparative study of the phyletic divergence of all evolutionary lineages, since the evolutionary history of sister groups is often strikingly different. Among two related groups, derived from the same nearest common ancestor, one may hardly differ from the ancestral group, while the other may have entered a new adaptive zone and evolved into a novel type. Even though they are sister groups in the terminology of cladistics, they may deserve different categorical rank, because their biological characteristics differ to such an extent as to affect any comparative study. The importance of this consideration was stated by Darwin (*1*, p. 420): "I believe that the *arrangement* of the groups within each class, in due subordination and relation to the other groups, must be strictly genealogical in order to be natural, but that the *amount* of difference in the several branches or groups, though allied in the same degree in blood to their common progenitor, may differ greatly, being due to the different degrees of modification which they have undergone, and this is expressed by the forms being ranked under different genera, families, sections or orders." Darwin refers then to a diagram of three Silurian genera that have modern descendants; one has not even changed generically, but the other two have become distinct orders, one with three and the other with two families.

The question as to what extent an analysis of degrees of divergence is possible, is still debated. The cladist makes only "horizontal" comparisons, cataloging the synapomorphies of sister groups. The evolutionary taxonomist, however, also makes use of derived characters that are restricted to a single line of descent, so-called autapomorph characters (Fig. 1), which are apomorph characters restricted to a single sister group. The importance of autapomorphy is well illustrated by a comparison of birds with their sister group (*34*). Birds originated from that branch of the reptiles, the Archosauria, which also gave rise to the pterodactyls, dinosaurs, and crocodilians. The crocodilians are the sister group of the birds among living organisms; a stem group of archosaurians represents the common ancestry of birds and crocodilians. Although birds and crocodilians share a number of synapomorphies that originated after the archosaurian line had branched off from the other reptilian lines, nevertheless crocodilians are on the whole very similiar to other reptiles, that is, they have developed relatively few autapomorph characters. They represent the reptilian "grade," as many morphologists call it.

Birds, by contrast, have acquired a vast array of new autapomorph characters in connection with their shift to aerial living. Whenever a clade (phyletic lineage) enters a new adaptive zone that leads to a drastic reorganization of the clade, greater taxonomic weight may have to be assigned to the resulting transformation than to the proximity of joint ancestry. The cladist virtually ignores this ecological component of evolution.

The main difference between cladists and evolutionary taxonomists, thus, is in the treatment of autapomorph characters. Instead of automatically giving sister groups the same rank, the evolutionary taxonomist ranks them by considering the relative weight of their autapomorphies as compared to their synapomorphies (Fig. 1). For instance, one of the striking autapomorphies of man (in comparison to his sister group, the chimpanzee) is the possession of Broca's center in the brain, a character that is closely correlated with man's speaking ability. This single character is for most taxonomists of greater weight than various synapomorphous similarities or even identities in man and the apes in certain macromolecules such as hemoglobins and cytochrome *c*. The particular importance of autapomorphies is that they reflect the occupation of new niches and new adaptive zones that may have greater biological significance than synapomorphies in some of the standard macromolecules.

I agree with Szalay (*35*) when he says: "The loss of biological knowledge when not using a scheme of ancestor-descendant relationship, I believe, is great. In fact, whereas a sister group relationship may . . . tell us little, a postulated and investigated ancestor-descendant relationship may help explain a previously inexplicable character in terms of its origin and transformation, and subsequently its functional (mechanical) significance." In other words, the analysis of the ancestor-descendant relationships adds a great deal of information that cannot be supplied by the analysis of sister group relationships.

It is sometimes claimed that the analysis of ancestor-descendant relationships lacks the precision of cladistic sister group comparisons. However, as was shown above and as is also emphasized by Hull (*36*), the cladistic analysis is actually full of uncertainties. The slight possible loss of precision, caused by the use of autapomorphies, is a minor disadvantage in comparison with the advantage of the large amount of additional information thus made available.

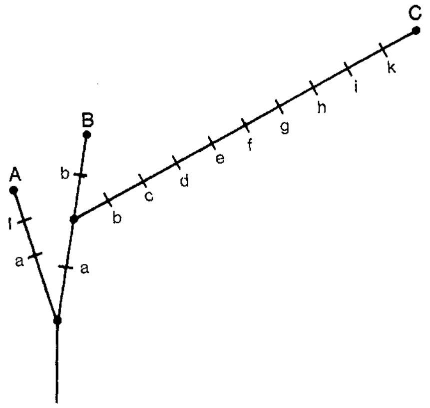

Fig. 1. Cladogram of taxa A, B, and C. Cladists combine B and C into a single taxon because B and C share the synapomorph character b. Evolutionary taxonomists separate C from A and B, which they combine, because C differs by many (c through k) autapomorph characters from A and B and shares only one (b) synapomorph character with B.

The information on autapomorphies permits the conversion of the cladogram into a phylogram. The phylogram differs from the cladogram by the placement of sister groups at different distances from the joint common ancestry (branching point) and by the expression of degree of divergence by different angles. Both of these topological devices can be translated into the respective categorical ranking of sister groups. These methods (*37*) generally attempt to discover the shortest possible "tree" that is compatible with the data. Yet, anyone familiar with the frequency of evolutionary reversals and of evolutionary opportunism, realizes the improbability of the assumption that the tree constructed by this so-called "parsimony method" corresponds to the actual phylogenetic tree. "To regard [the shortest tree method] as parsimonious completely misconceives the intent and use of parsimony in science" (*38*).

It is not always immediately evident whether a tree construction algorithm is based on cladists principles or on the methods of evolutionary classification. If the "special similarity" on which the trees are based are strictly synapomorphies, then the method is cladistic. If autapomorphies are also given strong weight, then the method falls under evolutionary classification.

The particular aspect of the method of evolutionary taxonomy found most unacceptable to cladists is the recognition of "paraphyletic" taxa. A paraphyletic taxon is a holophyletic group from which certain strikingly divergent members have been removed. For instance, the class Reptilia of the standard zoological literature is paraphyletic, because birds and mammals, two strikingly divergent descendants of the same common ancestor of all the Reptilia, are not included. Nevertheless, the traditional class Reptilia is monophyletic, because it consists exclusively of descendants from the common ancestor, even though it excludes birds and mammals owing to the high number of autapomorphies of these classes. The recognition of paraphyletic taxa is particularly useful whenever the recognition of definite grades of evolutionary change is important.

The Ranking of Taxa

Once species have been grouped into taxa the next step in the process of biological classification is the construction of a hierarchy of these taxa, the so-called Linnaean hierarchy. The hierarchy is constructed by assigning a definite rank such as family or order to each taxon, subordinating the lower categories to the higher ones. It is a basic weakness of cladistics that it lacks a sensitive method of ranking and simply gives a new rank after each branching point. The evolutionary taxonomist, following Darwin, ranks taxa by the degree of divergence from the common ancestor, often assigning a different rank to sister groups. Rank determination is one of the most difficult and subjective decision processes in classification. One aspect of evolution that causes difficulties is mosaic evolution (*39*). Rates of divergence of different characters are often drastically different. Conventionally taxa, such as those of vertebrates, are described and delimited on the basis of external morphology and of the skeleton, particularly the locomotory system. When other sets of morphological characters are used (for example, sense organs, reproductive system, central nervous system, or chromosomes), the evidence they provide is sometimes conflicting. The situation can become worse, if molecular characters are also used. The anthropoid genus *Pan* (chimpanzee), for instance, is very similar to *Homo* in molecular characters, but man differs so much from the anthropoid apes in traditional characters (central nervous system and its capacities) and occupation of a highly distinct adaptive zone that Julian Huxley even proposed to raise him to the rank of a separate kingdom—Psychozoa.

It has been suggested that different classifications should be constructed for each kind of character, or at least for morphological and molecular characters. Yet there is already much evidence that the acceptance of several classifications based on different characters would lead to insurmountable complications. By taking all available data into consideration simultaneously, a classification can usually be constructed that can serve conveniently as an all-purpose classification or, as Hennig (*5*) called it, "a general reference system."

It is usually possible to derive more than one classification from a phylogram, because higher taxa are usually composed of several end points of the phylogram, and different investigators differ by the degree to which they lump such terminal branches into a single higher taxon (*40*). An example is the phylogram of the higher ferns on which, as Wagner (*41*) has shown, six different classifications have been founded (Fig. 2) and many more are possible. The extent to which investigators "split" or "lump" higher taxa, thus, is of considerable influence on the classifications they produce.

Comparison of the Three Major Schools

Each school believes that its classification is the "best." Pheneticists as well as cladists claim that their respective methods have also the great merit of giving automatically nonarbitrary results. These claims cannot be substantiated. To be sure grouping by phenetic characters and determination of holophyly by cladistic analysis are valuable components of the procedure of biological classification. The great deficiency of both phenetics and cladistics is the failure to reflect adequately the past evolutionary history of taxa.

What needs to be emphasized once more is the fact that groups of organisms are the product of evolution and that no classification can hope to be satisfactory that does not take this fact fully into consideration. Both pheneticists and cladists are ambiguous in their attitude toward the evolutionary theory. The pheneticists claim that their approach is completely theory-free, but they nevertheless assume that their method will produce a hierarchy of taxa that corresponds to descent with modification. On the basis of this assumption, they also claim to be "evolutionary taxonomists" (*42*), but the fact that different phenetic procedures may produce very different classifications and that their procedure is not influenced by evolutionary considerations refutes this assertion. The cladists exclude most of evolutionary theory (for example, inferences on selection pressures, shifts of adaptive zones, evolutionary rates, and rates of evolutionary divergence) from their consideration (*43*) and tend increasingly not to classify species and taxa, but only taxonomic characters (*28*) and their origin. The connection with evolutionary principles is exceedingly tenuous in many recent cladistic writings.

By contrast, the evolutionary taxonomists, as indicated by the name of their school and by well-articulated statements of some of its major representatives (*7*), expressly base their classifications on evolutionary theory. They aim to construct classifications that reflect both of the two major evolutionary processes, branching and divergence (cladogenesis and anagenesis). They make full use of information on shifts into new adaptive zones and rates of evolutionary change and believe that the resulting classifications are a key to a far richer information content.

Although the three schools still seem rather fundamentally in disagreement, as far as the basic principles of classification are concerned, the more moderate representatives have quietly incorporated some of the criteria of the opposing schools, so that the differences among them have been partially obliterated. For instance, Farris' (*44*) clustering of special similarities is a phenetic method based on the weighting of characters. The evolutionary school uses phenetic criteria to establish similarity classes and to construct a classification, and cladistic criteria to test the naturalness of taxa. Comparing what McNeill (*45*) says in favor of phenetics (appropriately modified) and Farris (*44*) against it, we find that the gap has narrowed. I have no doubt that moderates will be able to develop an eclectic methodology, one that contains a proper balance of phenetics and cladistics that will produce far more "natural classifications" (*16*) than any one-sided approach that relies exclusively on a single criterion, whether it be overall similarity, parsimony of branching pattern, or what not. Evolutionary taxonomy, from Darwin on, has been characterized by the adoption of an eclectic approach that makes use of similarity, branching pattern, and degree of evolutionary divergence.

Classification and Information Retrieval

Biological classifications have two major objectives: to serve as the basis of biological generalizations in all sorts of comparative studies and to serve as the key to an information storage system. Up to this point, I have concentrated on

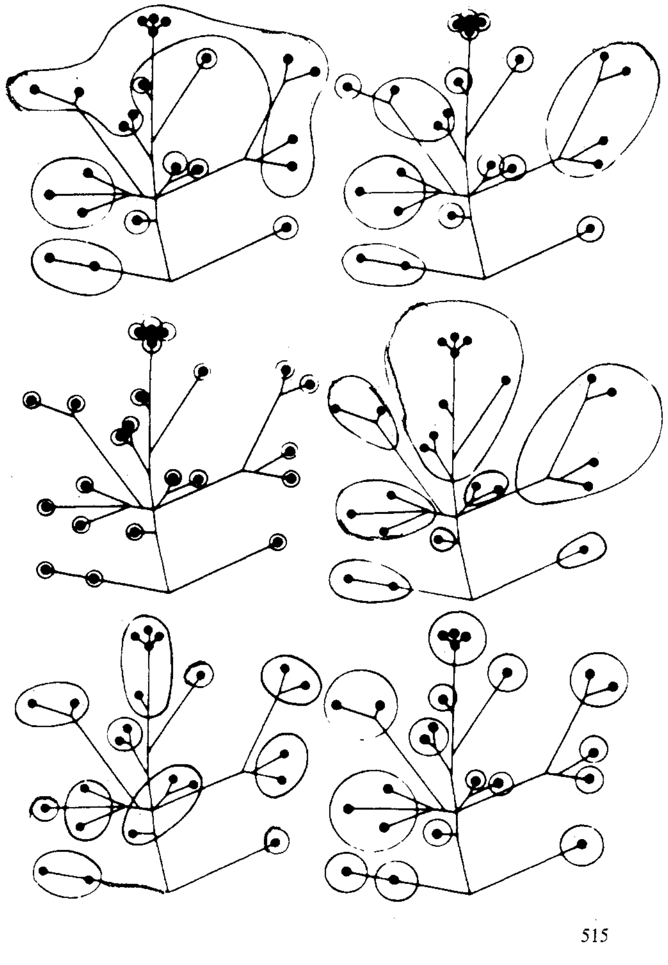

Fig. 2. Six different possible classifications of ferns, based on the same dendrogram. Each filled circle is a genus, and each open circle is a family. The differences are due to which and how many genera are combined to make up the families. [From W. H. Wagner (*41*, figure 7)].

those aspects of classifying that help to secure a sound basis for generalizations. This leaves unanswered the question of whether achievement of this first objective is, or is not, reconcilable with achievement of the second objective. Is the classification that is soundest as a basis of generalizations also most convenient for information retrieval? This, indeed, seems to have been true in most cases I have encountered. However, we can also look at this problem from another side.

It is possible at nearly each of the three major steps in the making of a classification to make a choice between several alternatives. These choices may be scientifically equivalent, but some may be more convenient in aiding information retrieval than others. If we choose one of them, it is not necessarily because the alternatives were "falsified," but rather because the chosen method is "more practical." In this respect, biological classifications are not unique. Scientific theories are nearly always judged by criteria additional to truth or falsity, for instance, by their simplicity or, in mathematics, by their "elegance." Therefore, it can be asserted that convenience in the use of a classification, including its function as key to information retrieval, is not necessarily in conflict with its more purely scientific objectives (*46–48*).

References and Notes

1. C. Darwin, *On the Origin of Species* (Murray, London, 1859).
2. For an illuminating survey of the thinking of that period see F. A. Bather [*Proc. Geol. Soc. London* **83**, LXII (1927)].
3. R. R. Sokal and P. H. A. Sneath, *Principles of Numerical Taxonomy* (Freeman, San Francisco, 1963). A drastically revised second edition was published in 1973.
4. The method was first published under the misleading name phylogenetic systematics, but since it is based on only a single one (branching) of the various processes of phylogeny, the terms cladism or cladistics have been substituted and are now widely accepted.
5. W. Hennig's original statement is *Grundzüge einer Theorie der Phylogenetischen Systematik* (Deutscher Zentralverlag, Berlin, 1950). A greatly revised second edition (reprinted in 1979) is *Phylogenetic Systematics*, D. D. Davis and R. Zangerl, Eds. (Univ. of Illinois Press, Urbana, 1966); see also W. Hennig (*47*). An independent phylogenetic analysis of characters was made by T. P. Maslin [*Syst. Zool.* **1**, 49 (1952)]. For an overview of the more significant recent literature see D. Hull (*36*) and J. S. Farris [*Syst. Zool.* **28**, 483 (1979)].
6. Some cladists in recent years have relaxed the requirements of strict dichotomy and have permitted tri- and polyfurcations or have quietly abandoned dichotomy by admitting empty internodes in the cladogram. Polyfurcations can be translated into several alternate bifurcations [see J. Felsenstein, *Syst. Zool.* **27**, 27 (1978)], and this makes the automatic conversion of the cladogram into a classification of sister groups impossible.
7. The classical statement of this theory is to be found in C. Darwin (*1*, pp. 411–434). G. G. Simpson [*Principles of Animal Taxonomy* (Columbia Univ. Press, New York, 1961)] and E. Mayr (*48*) provide comprehensive modern presentations of this theory. Several critical recent analyses are: W. Bock (*11*); C. D. Michener (*18*); *Syst. Zool.* **27**, 112 (1978); P. D. Ashlock (*12*).
8. F. Darwin, *Life and Letters of Charles Darwin* (Murray, London, 1887), vol. 2, p. 247.
9. D. Hull, *Evolution* **21**, 174 (1967); see also W. Bock (*11*).
10. F. E. Warburton, *Syst. Zool.* **16**, 241 (1967); W. Bock, *ibid.* **22**, 375 (1973).
11. W. Bock, in *Major Patterns in Vertebrate Evolution*, M. K. Hecht, P. C. Goody, B. M. Hecht, Eds. (NATO Advanced Study Institute Series, Plenum, New York, 1977), vol. 14, pp. 851–895.
12. P. D. Ashlock, *Syst. Zool.* **28**, 441 (1979).
13. I shall not, at this time, recount the almost interminable controversies among the three schools. For critiques of phenetics see E. Mayr (*48*, pp. 203–211), L. A. S. Johnson [*Syst. Zool.* **19**, 203 (1970)], and D. Hull [*Annu. Rev. Ecol. Syst.* **1**, 19 (1970)]. Some of the weaknesses pointed out by these early critics have been corrected in the 1973 edition of Sokal and Sneath (*3*) and by J. S. Farris (*44*). For critiques of cladistics see E. Mayr (*21*), R. R. Sokal (*22*), G. G. Simpson (*32*), D. Hull (*36*), P. D. Ashlock [*Annu. Rev. Ecol. Syst.* **5**, 81 (1974)]; and L. van Valen (*49*).
14. A particularly illuminating example is the breaking up of the plant group Amentiferae, which has been shown to consist of taxa secondarily adapted for wind pollination [R. F. Thorne, *Brittonia* **25**, 395 (1973)]. Examples among animals of radical reclassifications are the Rodentia, parasitic bees, certain beetle families, and the turbellarians.
15. C. G. Sibley, in preparation.
16. There have been arguments since before the days of Linnaeus about how to determine whether or not a system, a classification, is "natural." William Whewell, at a time before Darwin, had proclaimed his theory of common descent, expressed the then prevailing pragmatic consensus, "The maxim by which all systems professing to be natural must be tested is this: that the arrangement obtained from one set of characters coincides with the arrangement obtained from another set" [W. Whewell, *Philos. Inductive Sci.* **1**, 521 (1840)]. Interestingly, the covariance of characters is still perhaps the best practical test of the goodness of a classification. Since Darwin, of course, that classification is considered most natural that best reflects the inferred evolutionary history of the organisms involved.
17. For a tabulation and analysis of such qualitative methods of weighting, see E. Mayr (*48*, pp. 220–228).
18. C. D. Michener, *Syst. Zool.* **26**, 32 (1977).
19. I use the word monophyletic in its traditional sense, as a qualifying adjective of a taxon. Various definitions of monophyletic have been proposed but all of them for the same concept, a qualifying statement concerning a taxon. A taxon is monophyletic if all of its members are derived from the nearest common ancestor [E. Haeckel, *Natürliche Schöpfungsgeschichte* (Reimer, Berlin, 1868)]. Cladists have attempted to turn the situation upside down by placing all descendants of an ancestor into a taxon. Monophyletic thus becomes a qualifying adjective for descent, and a taxon is not recognized by its characteristics but only by its descent. The transfer of such a well-established term as monophyletic to an entirely different concept is as unscientific and unacceptable as if someone were to "redefine" mass, energy, or gravity by attaching these terms to entirely new concepts. P. D. Ashlock [*Syst. Zool.* **20**, 63 (1971)] has proposed the term holophyletic for the assemblage of descendants of a common ancestor. See also P. D. Ashlock (*12*, p. 443).
20. Terms like apomorph, synapomorph, derived, ancestral, and so forth always refer to characters of taxa at all levels. A genus may have synapomorphies with another genus, and so may an order with another order. It is this applicability of the same criteria for taxa of all ranks that permits the construction of the Linnaean hierarchy.
21. E. Mayr, *Z. Zool. Syst. Evolutionsforsch.* **12**, 94 (1974); reprinted in, E. Mayr, *Evolution and the Diversity of Life* (Harvard Univ. Press, Cambridge, Mass., 1976), pp. 433–478.
22. R. R. Sokal, *Syst. Zool.* **24**, 257 (1975).
23. I. Tattersall and N. Eldredge, *Am. Sci.* **65**, 204 (1977).
24. D. S. Peters and W. Gutmann, *Z. Zool. Syst. Evolutionsforsch.* **9**, 237 (1971).
25. O. Schindewolf, *Acta Biotheor.* **18**, 273 (1968); H. K. Erben, *Verh. Dtsch. Zool. Ges.* **79**, 116 (1979).
26. G. G. Simpson (*32*); see also L. van Valen (*49*).
27. For a diagram of these three categories of morphological resemblance see figure 1 in W. Hennig (*47*).
28. "Cladistic classifications do not represent the order of branching of sister groups, but the order of emergence of unique derived characters" [see D. Hull (*36*)].
29. G. G. Simpson [*The Major Features of Evolution* (Columbia Univ. Press, New York, 1953), p. 348] discusses the fallacy of the cladist assumption.
30. Phylogeny is equated by cladists with cladogenesis (branching), while the evolutionary taxonomist subsumes both branching and evolutionary divergence (anagenesis) under phylogeny.
31. C. W. Harper, *J. Paleontol.* **50**, 180 (1976).
32. G. G. Simpson, in *Phylogeny of the Primates*, W. Pluckett and F. S. Szalay, Eds. (Plenum, New York, 1975), pp. 3–19.
33. J. H. Camin and R. R. Sokal, *Evolution* **19**, 311 (1965); W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967); W. M. Fitch, in *Major Patterns of Vertebrate Evolution*, M. K. Hecht, P. C. Goody, B. M. Hecht, Eds. (NATO Advanced Study Institute Series, Plenum, New York, 1977), vol. 14, pp. 169–204.
34. There are literally hundreds of cases to illustrate this situation. I use again the classical case of birds and crocodilians because even a nonbiologist will understand the situation if such familiar animals are used. The holophyletic classification of the lice (Anoplura) derived from one of the suborders of the Mallophaga is another particularly instructive example [K. C. Kim and H. W. Ludwig, *Ann. Entomol. Soc. Am.* **71**, 910 (1978)].
35. F. S. Szalay, *Syst. Zool.* **26**, 12 (1977).
36. D. Hull, *ibid.* **28**, 416 (1979).
37. J. W. Hardin, *Brittonia* **9**, 145 (1957); W. H. Wagner, in *Plant Taxonomy: Methods and Principles*, L. Benson, Ed. (Ronald, New York, 1962), pp. 415–417; A. G. Kluge and J. S. Farris, *Syst. Zool.* **18**, 1 (1969); J. S. Farris, *Am. Nat.* **106**, 645 (1972).
38. L. H. Throckmorton, in *Biosystematics in Agriculture*, J. A. Romberger, Ed. (Wiley, New York, 1978), p. 237. Others who have questioned the validity of the so-called parsimony principle are M. Ghiselin [*Syst. Zool.* **15**, 214 (1966)] and W. Bock (*11*).
39. Unequal rates of evolution for different structures or for any other components of phenotypes or genotypes are designated mosaic evolution.
40. See E. Mayr (*48*, pp. 238–241) on the differences between splitters and lumpers.
41. W. H. Wagner, "The construction of a classification," in *Systematic Biology* (Publication 1692, National Academy of Sciences, Washington, D.C., 1969), pp. 67–90.
42. R. R. Sokal in (*22*), "I have yet to meet a nonevolutionary taxonomist."
43. Several leading cladists have recently published antiselectionist statements.
44. J. S. Farris, in *Major Patterns in Vertebrate Evolution*, M. K. Hecht, P. C. Goody, B. M. Hecht, Eds. (NATO Advanced Study Institute Series, Plenum, New York, 1977), vol. 14, pp. 823–850.
45. J. McNeill, *Syst. Zool.* **28**, 468 (1979).
46. For criteria by which to judge the practical usefulness of biological classifications, see E. Mayr (*48*, pp. 229–242).
47. W. Hennig, *Annu. Rev. Entomol.* **10**, 97 (1965).
48. E. Mayr, *Principles of Systematic Zoology* (McGraw-Hill, New York, 1969).
49. L. van Valen, *Evol. Theory* **3**, 285 (1978).
50. Drafts were read by P. Ashlock, J. Beatty, W. Bock, W. Fink, C. G. Hempel, and D. Hull, to all of whom I am indebted for valuable suggestions and critical comments, not all of which was I able to accept.