6

The Study of Phylogeny

As soon as phylogeny becomes a consideration to be dealt with during the classification of a group of organisms, we must be concerned not only with the phenetic relationship among the end points of the branching sequence, but also with phenetic relationships among any points that have at one time or another been occupied by organisms belonging to the phyletic branch under consideration. We have furthermore to concern ourselves with the sequence of branching as well as the time dimension. The nature of these relationships and the problems inherent in their analysis have been discussed in detail in Chapter 2. In Section 6.1 we shall be primarily concerned with the time dimension. This leads to a discussion of rates of evolution in Section 6.2, followed by the theory and practice of cladistic analysis in Sections 6.3 and 6.4. Finally in Section 6.5 we cover the applications of numerical taxonomy to paleontology.

6.1 PHENETICS AND THE TIME DIMENSION

Although we have already defined our terms carefully in earlier sections it will be useful to review them briefly for a discussion of phenetics and time dimensions. Phenetic taxonomy of recent organisms estimates their resemblances on the present time plane. Note that phenetic taxonomy is not restricted to comparisons

at a single time plane; as we shall see in Section 6.5, it is possible and of some interest to estimate phenetic relationships between OTU's from different time planes. Cladistic taxonomy in the generally accepted sense reconstructs the branching pattern leading to the present phenetic positions but does not deal with the phenetic relationships among the ancestral forms, except to estimate the cladograms. Phylogenetic taxonomy would aim to reconstruct not only the branching pattern itself, but also the detailed phenetic resemblances between the stems in the past. In theory, therefore, phylogenetic taxonomy aims to reconstruct the characters of the ancestral organisms and thus to obtain such properties (e.g., evolutionary rates, convergence, parallelism) as would be obtained from a complete fossil record. Developments on these lines are now being attempted based on conventional morphological data and studies of protein sequences.

Can Ancestral Forms Be Included in Phenetic Classifications? What Would Be Their Rank? Plainly any division of continuous lineages is to some extent arbitrary. When two taxa ranked as classes are separated at a given line, the species on either side of this line, though both genetically and phenetically closely related, will be grouped in different classes. In discussing this problem, Remane (1956) points out that in lineages such as the one in Figure 6-1 (representing the phylogeny of a family) the species **H** and **I** have a dual but partly contradictory relationship. They are very similar both in properties and closeness of ancestry and should therefore be placed in one genus. However, they are also ancestors respectively of the genus (**A, B, C, D**) and the genus (**E, F, G**), and the closeness of their phenetic and phyletic relationship indicates they could legitimately be included in these two genera to give the genus (**A, B, C, D, H**) and the genus (**E, F, G, I**).

Establishing taxa (**A, B, C, D**), (**E, F, G**), and (**H, I**) is horizontal classification. The second method—establishing taxa (**A, B, C, D, H**) and (**E, F, G, I**)—is classifi-

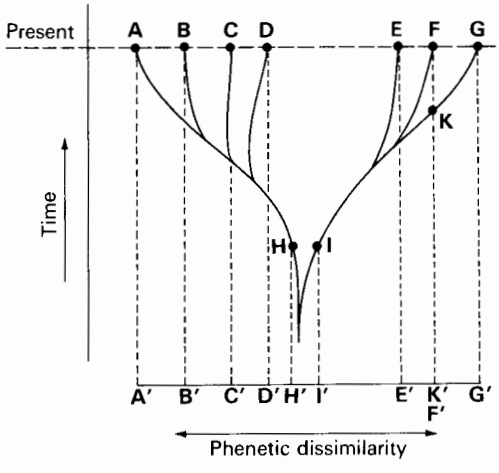


FIGURE 6-1
A phylogenetic tree as commonly represented, with a time dimension and one phenetic dimension. For explanation, see text.

cation by *clades* (Huxley, 1958). In such complex relationships it is a serious problem to know where to draw the dividing lines (made even more difficult if lineages fuse by hybridization, which is often true of plants). Those taxonomists who prefer monophyletic lineages would divide the group into clades wherever possible, though by phenetic criteria the OTU's would cluster to give the most cohesive taxa. Mainly such phenetic clusters would be monophyletic, but need not be so.

One misleading point about a figure such as Figure 6-1 is that it does not represent the phenetic relations at all accurately. In Figure 6-1, for example, the form (species?) **K** is directly below **F**; it is, however, exceedingly unlikely that it would be phenetically identical with **F**. The phenetic relations of the forms in this diagram are their relations in the horizontal plane, as would be shown by throwing a shadow of them onto the base line. This is equivalent to making projections of all the points on the dendrogram onto the abscissa, as is shown for a few points in Figure 6-1, for in general a single dimension is insufficient for representing phenetic relationships. These are represented somewhat more clearly in Figure 6-2, where a three-dimensional model of a substantially similar diagram is shown. Here the horizontal plane shows phenetic dissimilarity in two dimensions; the vertical dimension represents time, as before. Now the phenetic relations can be shown by the shadow of the phylogenetic "tree" projected onto the base plate. This shadow is shown again in Figure 6-3,*a*. It is a fronded figure in which the organisms of all time periods are shown without overlapping one another (on this scale of taxonomic discrimination). If we wished to divide it phenetically, we would divide it into two main groups, genera perhaps, roughly as shown. The exact place of the division line could be determined mathematically if we wished, though for most purposes a division at the point of branching would suffice. We would rarely have enough fossil data to have a complete shadow, however; we would be more likely to have an incomplete set of organisms giving a shadow such as in Figure 6-3,*b*, and most

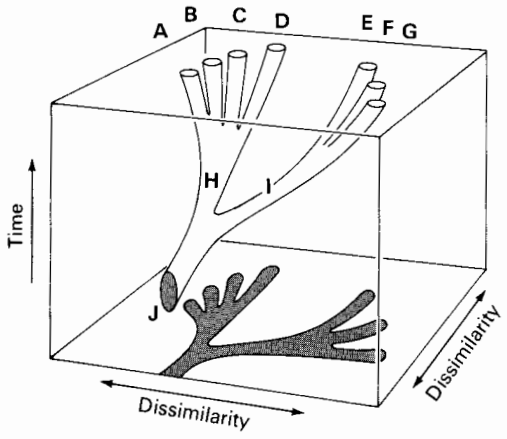


FIGURE 6-2
A phylogenetic tree in three dimensions, one of time and two of phenetic dissimilarity. The "shadow" of the tree on the base indicates the purely phenetic relationships. For explanation, see text.

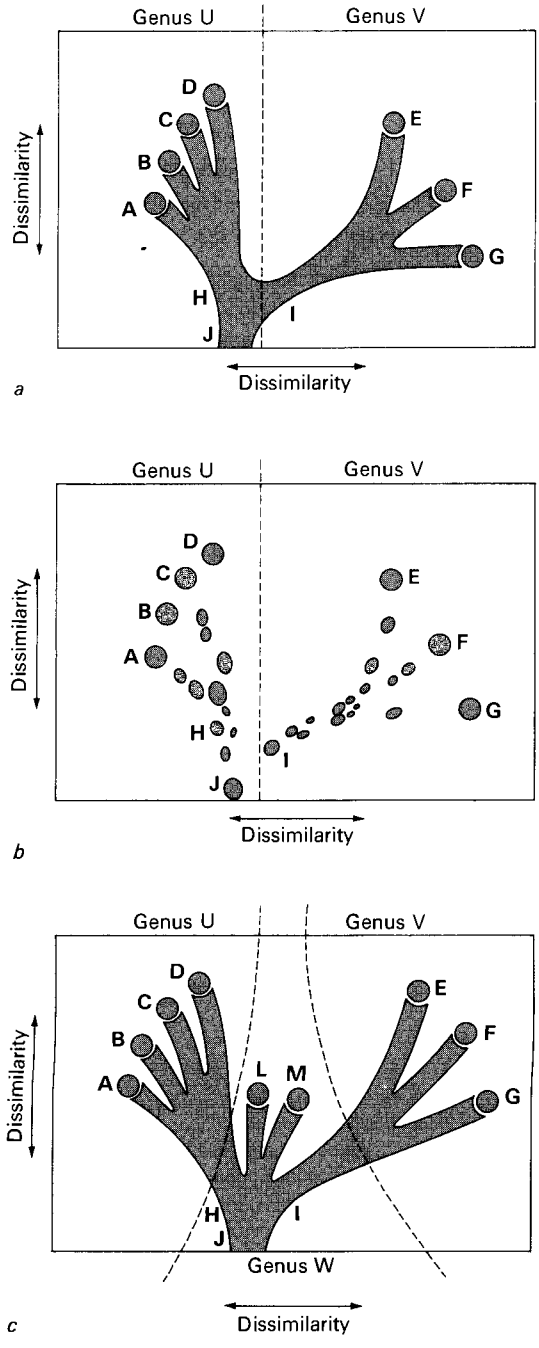


FIGURE 6-3
The "shadow" from Figure 6-2. The phenetic relationships are represented as shadows on the horizontal plane. *a*, The shadow from Figure 6-2 with a dashed line dividing it into two phenetic taxa, such as two genera. *b*, A patchy shadow. This is a more realistic representation because of the usual scarcity of fossils. *c*, Division into three phenetic taxa (such as genera) when branches **L** and **M** have been added.

often we would be lucky to get this amount of information. The exact position of the dividing line would then not be worth much argument.

We might have had some subsidiary branches (say, **L** and **M**) near the common ancestor, as shown in Figure 6-3,*c*. If so, one would, on phenetic grounds, divide it more or less as shown into the three genera **U**, **V**, and **W**. Note that in all cases the phenetic divisions—divisions made on the basis of the shadows—are, as we would expect, fairly close approximations of monophyletic groups or single lineages, though divergent clusters of branches may be excluded from the basal taxon. This seems to the authors the only honest thing to do; we believe that phylogenies are deduced necessarily from the phenetic relations.

6.2 RATES OF EVOLUTION

When it is possible to study fossil material, thus obtaining data from several known points of time, the resemblance coefficients will allow estimates of overall evolution rates. The dissimilarity between ancestral and descendant forms will be the measure of the overall evolution that has occurred in the intervening period. Simpson (1944) has discussed the great advantages of measuring the overall rate of evolution (what he calls the "organism rate"), as well as the rate of evolution in one or a few characters ("character" and "character complex" rates such as those studied by Haldane, 1949; Kurtén, 1958, 1959; and Buzzati-Traverso, 1959). Numerical taxonomy therefore offers a solution to many of the problems propounded by Simpson and by Huxley (1957). Simpson (1944) uses the terms tachytelic, horotelic, and bradytelic to describe rates that are respectively rapid, moderate, and slow. Little is known about tachytelic evolution, since the changes are so rapid that there is small chance of finding fossils of the relevant period. Bradytelic evolution is the kind shown by "living fossils" such as *Lingula*, *Ginkgo*, *Metasequoia*, and the coelacanth *Latimeria*.

Character Rates

Haldane (1949) has suggested a measure of the rate of evolution of a single character (for example, the length of an organ), so that the unit rate, the *darwin*, corresponds to a change by a factor of e in one million years—that is,

$$\frac{\ln X_T - \ln X_0}{T} = 1, \qquad \text{or} \qquad \frac{X_T}{X_0} = e^T$$

when the character has the value X_0 at time 0 and X_T at time T, with T measured in *crons* (the units of X are immaterial). (The word *cron* (Huxley, 1957) is a convenient term for one million years.) A darwin is approximately equivalent to a change by a factor of 1/1,000 in 1,000 years. If the allometry equation $\log y = \log$

$a + b \log x$ is used, the constant b should be used without transformation into logarithms since it is itself effectively a logarithm. In the examples studied by Haldane, the rate of change in horotelic evolution was around 0.04 darwins (40 millidarwins), but he noted that domestic animals have changed at rates of kilodarwins, so that increased selection can evidently greatly increase the usual rate of evolution.

In a study of fossil horses, Downs (1961) found rates of 12.3 to 124.3 millidarwins for tooth characters, their mean being about 57 millidarwins. These rates seem fairly typical of horotelic evolution, though Simpson (1953) notes that there is considerable variation between horotelic lines. The rate may, of course, vary within any one line as well. It may be that in the work to date there has been a bias to select for measurement the more rapidly changing characters.

The darwin cannot be thought of as an absolute measure of evolution, since its value depends on the manner of scaling of the character (Sokal and Sneath, 1963, p. 239). Haldane (1949) also suggested measuring evolution rates as the time for the mean of a given character to change by one standard deviation, but this seems unduly dependent on the variability of the character, which may perhaps change erratically with time or, as has recently been suggested, may in fact be determining the rate of evolution (see Section 6.3).

Recent work on protein sequences has thrown some light on the rates of evolution as inferred from the time of separation of phyletic lines from geological evidence. Margoliash and Smith (1965), Zuckerkandl and Pauling (1965), and Kimura (1969) showed that the estimated rates were fairly constant over long periods of geological time, and suggested corrections to account for repeated mutations at the same site (see Section 5.12). There are some uncertainties with regard to very long periods, such as time back to the common ancestors of animals and plants. The rates vary widely with different types of protein: the average number of mutations becoming established in a lineage for a sequence of 100 amino acids in a period of 100 crons varies from about 90 for fibrinopeptides to 0.06 for histones (McLaughlin and Dayhoff, 1969), with more typical values such as 12 for hemoglobin and 3 for cytochromes.

Little is yet known of the relation of these rates of protein evolution to those of DNA. Hoyer et al. (1965) have shown that there is a roughly logarithmic relation between the degree of nucleic acid pairing and the time to a common ancestor, with a drop in pairing by one power of 10 in every 300 crons. Studies on the relations between phenetic change and change in DNA include those of Sneath (1964d) and Laird, McConaughy, and McCarthy (1969). The existence of numerous single mutation differences between almost all proteins of closely related mammals poses questions for population biology, because there are several thousand different proteins in mammals. The replacement of such large numbers of alleles by selection implies very heavy evolutionary cost (see Kimura, 1968).

Character Complex Rates and Mosaic Evolution

Incongruence between numerical taxonomies based on different stages of the life cycle or different organs implies that there have been different rates of evolution in various character complexes. It is easy to imagine, for example, that the larva of an insect species might become adapted to a special habitat, giving rise to marked phenetic changes, while the adults of this and closely related species might remain in much the same habitat as before and retain high taxonomic similarity. When applied to different rates of evolution in different organ systems this phenomenon has been called mosaic evolution (de Beer, 1954). A good discussion of this problem in human evolution is given by Campbell (1964). If there is great difference in the evolutionary rates of different character complexes, this clearly raises problems in measuring overall evolution (organism rates, discussed next) analogous to the problems in making a taxonomy when incongruence is marked. Marcus (1969) has used Mahalanobis' D^2 to measure character complex evolution rates in teeth of the orangutan.

Organism Rates

When we come to a discussion of evolutionary rates in organisms, we enter a subject fraught with many pitfalls. In a simple, superficial sense it seems quite clear that different groups have evolved at different rates. For example, the statement that coelacanths have evolved more slowly than horses since the Eocene appears self-evident and is not likely to be challenged. However, when we attempt to analyze in detail the meaning of statements such as these, we run into considerable difficulty because we do not have any absolute scale of phenetic similarity. We think that we can make two definite statements. First, the absolute evolutionary rates of no group should be investigated until its characters and phenetic resemblances have first been investigated and evaluated with relation to the higher ranking groups to which they belong, and also to neighboring taxa. Second, a paleontologist, in comparing the rates of evolution of coelacanths and horses, has in the back of his mind an idea of the range of characters within vertebrates that have to his knowledge changed, and the kinds and degrees of change that have happened in all the vertebrate classes. He is evaluating the changes in the coelacanths and in horses against this unexpressed standard. A possible mode of evaluation may be the following: once a large group, such as the mammals, has been sufficiently studied by means of numerical taxonomy, a series of marker taxa may be chosen, with which evolutionary change can be compared. Thus a single representative of each order of mammals might be appropriately included in the matrix of resemblances and serve to furnish the proper scale for a group, such as the horses, for example. A fairly comprehensive standard set of taxa of this sort would comprise a reasonably stable standard of comparison.

An advantage of organism rates is that they are likely to be more steady than character rates, since bursts of rapid change in individual characters will tend to be smoothed out. Simpson (1944) has estimated organism rates (the "taxonomic rates" of Kurtén, 1959) by measuring the time for a phyletic lineage to change morphologically (phenetically) from one genus to another. From the data of Simpson (1944, 1961) and Kurtén (1959) and a consideration of the time of appearance of different taxa in vertebrate evolution, we may estimate that the time corresponding to change in rank in horotelic evolution in vertebrates is approximately as follows: morphospecies, 0.5 crons; genus, 7 crons; family, 20 crons; order, 45 crons; class, 80 crons. Myers (1960) discusses the rate of evolution of fishes after their introduction into lakes. With the exception of one lake in the Philippines (where very rapid evolution of several genera may have occurred in as little as 10,000 years), the usual pattern is the evolution of a few new species and subspecies after about 1/10 cron, many new species and some new genera after 1/2 cron, and many new genera and some new families after one to two crons. These rates are somewhat faster than those given above.

The rates appear to be much slower in some other phyla, such as insects (Crowson, 1958) and many lines of mollusks. In flowering plants rates have also been slower (Stebbins, 1950, pp. 529, 547 ff.) though there has been rapid evolution of some groups during the Pleistocene, with new species arising within one cron, including some arising in historic times. It should be made clear that by the phenetic change corresponding to a genus (for example) we mean the minimum phenetic difference between two forms that would just necessitate the placing of the two forms in different but closely similar genera instead of placing both in one genus (according to the criteria established by the investigator).

Early work on organism rates includes that of Westoll (1949) on lungfishes. Kurtén (1958) suggested that the percentage of significantly differing allometric growth gradients between two populations could be used as a measure of the organism rate. He calls this the "differentiation index." The index increases as a geometric series, with the limit value of 100; for example, the steps from 0 to 50, from 50 to 75, and from 75 to 87.5 are all equivalent. It runs parallel with taxonomic (phenetic) change but has the disadvantage of not taking into account the magnitude of the differences in the gradients (except so far as the magnitudes make the differences statistically significant). In most instances in mammals the rate of change was about 0.2 percent per thousand years, but periods of more rapid evolution also occurred. He found that the morphological difference between two subspecies was equivalent to an index of about 50 percent and that between two species was about 75 percent (Kurtén, 1959). It is clear from the context that Kurtén here uses subspecies to indicate a major morphological subdivision of a species rather than a trivial variant, and that by species he means a category approximating a morphospecies.

We should emphasize that all the above considerations are based on conventional judgments of taxonomic rank and are only as precise as these evaluations. Lacking

better ones, we cite them to give a general indication of the nature of the problem. Numerical taxonomy will yield resemblance values between chronologically successive organisms, which can be used as measures of evolutionary rate. Over small ranges the change in the resemblance values compared with time will be satisfactory expressions of the rate. Over larger ranges this may be unsatisfactory inasmuch as the similarity values may be in a scale (such as the index of Kurtén, just described) in which dissimilarities are not additive, so that for three OTU's, **a**, **b**, and **c**, with **b** midway between **a** and **c** in the agreed phenetic space, $U_{ab} + U_{bc} \neq U_{ac}$. If dissimilarities are not additive (this includes cosines of angular coefficients) it will be necessary to make appropriate transformations. Alternatively one can divide the resemblance scale to define different ranks and express the evolutionary rate as the time taken for a phyletic line to pass through the degrees of similarity values applicable to these ranks, as Simpson suggested.

Several numerical taxonomic studies on organism rates have recently been published. Grewal (1962) examined the rate of phenetic divergence in inbred lines of mice based on numerous skeletal characters, using a resemblance measure that corrects for sampling error. Lerman (1965a,b) reports on several paleontological examples, using as a distance measure the square root of Mahalanobis' D^2. His evolution rate measure is D/T, where T is the time between taxa **J** and **K**.

Lerman used rather few characters, but there is some evidence that the evolution rates were different in different parts of the lineages that he studied (horses, oreodonts, and molluscs). There would seem no strong reason to prefer what is in effect discriminant analysis units over other phenetic measures, such as taxonomic distance, because the estimation of evolution rates is not primarily a problem of discrimination. It may be mentioned that the most appropriate measures require some further investigation. If evolution is predominantly by replacement of successive single mutations the total evolutionary pathway in A-space is represented in Manhattan metric rather than taxonomic distance, since the pathway consists of successive displacements on each character axis in turn and not on the diagonal joining the starting and ending positions. There is some support for this from protein evolution data, because the difference in two sequences is roughly proportional to time, and this difference is equivalent to the distance along successive sides of a hypercube, whereas the taxonomic distance is the square root of this (see Sections 4.4, 5.12).

Whether evolution rates are constant and divergent throughout all parts of a phylogeny has important consequences for cladistic analysis, which is taken up in the next two sections. The evidence on whether rates are constant and divergent may be briefly discussed here. It is clear that for morphological characters evolution rates have been far from uniform, but in the absence of numerical phenetic data little more can be said about this. There is however a good deal of numerical evidence from protein sequences that bears on the question. Various authors, as mentioned above, have concluded that for any given protein, evolution has usually

been divergent and fairly constant in rate, although except for Kimura (1969) little attention has been paid to sampling error, which must be considerable, or to uncertainty in the paleontological dating of divergence between lineages, probably equally considerable. Kirsch (1969) points out that strictly constant evolution would rule out the possibility of biochemical "living fossils." One would be reluctant to deny their existence, although at present we have no way of demonstrating it. Exceptions to constant evolution rates are in fact now becoming evident in the insulins (Dayhoff and McLaughlin, 1969), for example, and mutations are far from random (Clarke, 1970). Similar considerations apply to serological evolution, which presumably reflects fairly closely the evolution of the proteins concerned: Sarich (1969a,b) and others (e.g., Read and Lestrel, 1970) consider that serological rates have been fairly constant, but Farris (1972) has shown very large heterogeneities in rates of albumin evolution in carnivores, on the basis of immunological evidence.

However, Kirsch (1969), Jardine, van Rijsbergen, and Jardine (1969), and Moore (1971), have made an important observation: if rates are constant and divergent, then the resemblances between extant organisms will be an ultrametric; by the usual methods of clustering the cophenetic correlation will be 1, and single linkage and complete linkage (as well as most forms of average linkage) will yield identical phenograms. Kirsch points out this is demonstrably not so for most matrices of serological resemblance although the evidence from serology may be questioned on the basis of inappropriate resemblance measures or techniques, or lack of correspondence between serological reactivity and protein differences.

Nor has evidence from protein sequences about constancy and divergency of evolution been carefully assessed either. The matrices of differences between organisms in Dayhoff (1969a) show, for many proteins, the long rows and columns of almost constant values that would be expected from an ultrametric with superimposed sampling error. J. S. Farris (personal communication) has recently shown that such seemingly ultrametric structure is to be expected in protein sequence data, whether the rates of evolution in the phyletic lines are really homogeneous or not. This is so because sites differing between modern taxa **A** and **B** are likely also to become modified in the evolutionary path connecting **X**, the common ancestor of **A** and **B**, to a third taxon, **C**. If **C** is quite distant from **X**, then the number of amino acid sequence differences between taxa **A** and **C** and taxa **B** and **C** tend to approximate equality—even though these differences are substantial for taxa **A** and **X** and **B** and **X**. It is of critical importance to investigate this carefully, including the consideration of alternatives like local constancy of rates (Jardine et al., 1969). The squared discrepancies between the resemblance matrix and the UPGMA cophenetic matrix could be compared with the sampling variances, because UPGMA clustering might be expected on theoretical grounds to give the closest approximation to the correct cladogeny (see Moore, 1971). Kirsch (1969) has made some simulation studies along these lines. If a hypothesis

of constant divergence can be strongly supported it would open the way for attack on many intractable cladistic problems.

Colless (1970) has pointed out that less stringent conditions than constant and divergent evolution may still permit reconstruction of cladogenies; phenograms may be quite good estimators of cladistics, provided that (1) later stocks do not evolve very much faster than earlier ones for appreciable periods of time, (2) sister species do not divide again before they have diverged appreciably, and (3) later lines do not converge upon each other less than upon earlier lines. Colless gives some reasons for preferring average linkage cluster methods for constructing the phenograms, while Moore (1971) has constructed a set of assumptions under which UPGMA clustering yields optimal estimates of cladograms.

6.3 CLADISTIC ANALYSIS

The problems of making inferences about cladistic relationships from data on recent organisms (even when fossil evidence is available) have been detailed in Sections 2.5 and 2.6. In spite of the manifold difficulties discussed there, considerable progress has been made in recent years in developing techniques of numerical cladistics. In this section we shall discuss in some detail the types of evidence on which phylogenetic work has largely rested as well as the philosophical bases of cladistic inferences.

The sources of data for cladistic inferences are invariably phenetic. This is obvious when cladistic analysis on morphological data follows the methods of Hennig (1966) or Maslin (1952). Even phylogenetic inferences from fossil material are based on phenetics; Colless (1967b) and Rowell (1970) have pointed this out. In certain special instances the data are not immediately recognized as being phenetic. These include cytogenetic data, especially those on chromosome inversion sequences in some species of diptera, which are believed by some (Stalker, 1966) to yield incontrovertible evidence of branching patterns. Yet when the reasoning about inversion sequences is examined carefully it is found that a linear order of inversions must be postulated before any cladistic inferences can be made. This linear order is based on the (phenetic) homology of bands in the chromosomes, on postulates about the likelihood of inversions (simultaneous breaks in the chromosomes with concomitant inversion of a middle segment), and on the relative improbability of such an event happening in the identical chromosome segments two or more times in any one lineage. The phenetic evidence must be employed together with the postulations of linear order and the likelihood of inversions before cladistic inferences can be made. When such information is subjected to a numerical cladistic procedure (that of Camin and Sokal, 1965 in this instance) in which the cytogenetic characteristics have been recoded in a linear ordering following the assumptions about their order of origin, the resulting cladogenies match those established by cytogenetic methods (unpublished work by John Hendrickson, Jr. on data by

Stalker, 1966) or are more parsimonious than previously published evolutionary trees (analyses of blackflies; unpublished information by R. Hansell about studies by Rothfels, 1956, and Basrur, 1959, 1962). The trees postulated by Rothfels and Basrur differ from the numerically obtained cladograms by allowing hypothetical intermediate forms in which alternative banding patterns were permitted to co-exist. This is a model similar to one developed in a cladistic computer program by J. H. Felsenstein.

Genetic information is often thought to be a sounder way of establishing cladistic relationships than phenetic characteristics. However, we are rarely in a position to establish genetic relationships among representatives of fairly divergent taxa (for exceptions see the work on protein structure and on DNA pairing discussed in the next paragraph). In genetic relationships of low ranked taxa the problems become more those of population structure and species definition than of cladogeny in the strict sense, although the general line of argument would still hold. However, even evidence on crossability is a phenetic character as will be obvious on some reflection. Morishima (1969b) has employed interstrain cross-ability of rice in just such a manner.

Biochemical evidence is often thought to be a "true indicator" of cladistic relationship. Yet the two major criteria, namely mutation distances found in comparing two sets of proteins and the degree of DNA pairing are only different types of estimates of phenetic resemblance between the genomes or their products. We have already discussed DNA pairing in Section 5.12 but shall deal in some more detail with matching and analysis of protein sequences in this and the next section. A difference in an amino acid at a given site in a protein is clearly a phenetic difference whether expressed in number of mismatches of the amino acids or in the inferred number of mutational steps necessary to bring about a change in the nucleotides coding for the given amino acid. Fitch and Margoliash (1968) recognize that the trees they reconstruct are phenograms because differences between amino acid sequences in proteins are no more or no less phenetic than conventional morphological characters.

We have already seen in Section 2.5 that an ordering of character states into an evolutionary sequence is at best a tricky procedure and at worst can be grossly misleading. Any attempt to reach decisions about the order of the evolutionary sequences in character states leads inevitably to statistico-phenetic taxonomy (Colless, 1969a) and some of the protein chemists have more frankly relied upon what they have descriptively termed decision by majority vote (Dayhoff and Eck, 1969).

Direct phenetic evidence is not the only kind used for making decisions on cladistic relationships. Traditionally other evidence such as the recapitulation hypothesis is employed to provide ancillary support. Biogeographic evidence is frequently used at the species level and for higher ranked taxa. There is little doubt that such evidence can lend support to a given cladistic hypothesis, or weaken it, as the case may be. However, it is unlikely to serve as primary evidence. Some

function needs to be constructed for incorporating ancillary evidence of various types into the decision algorithm for arriving at cladistic sequences. No attempts at such an incorporation of ancillary evidence has yet been made. It would involve a procedure for weighting these ancillary characters in relation to orthodox characters and would yield modified cladograms.

Several principles are either explicit or implicit in the operations and reasoning of cladistic systematists. Those who would largely use phenetic similarity as evidence for recency of cladistic ancestry must assume at least some uniformity of evolutionary rates in the several clades, although Colless (1970) has shown that the requirements need not be as stringent as is commonly thought (see Section 6.2). We stress here uniform rather than constant rates of evolution. As long as rates of evolution change equally in parallel lines it is unimportant whether these rates are constant through an evolutionary epoch. We may use the analogy of multiple clusters of fireworks, smaller clusters bursting from inside large clusters, a familiar sight to most readers. While the small rays of the rocket have not "evolved" at all until the small rocket exploded, their rates of divergence from the center of their rocket are identical but not constant, since they were zero during the period of the early ascent of the rocket.

Clearly all the evidence at hand indicates evolutionary rates in different clades are not uniform. Different lines do evolve at different rates. Taking this fact into account Farris (1966) has argued that present intrataxon variation is inversely related to conservatism of the character in the taxon. Kluge and Farris (1969) have used this assumption for weighting characters in their method of constructing Wagner trees (see Section 6.4). A. G. Kluge (personal communication) has recently obtained evidence from a variety of animal groups that tends to bear out the correlation between past evolutionary rate and present character variation within a taxon. This hypothesis is a modern version of the doctrine of uniformitarianism that, if it can be verified, would be of great importance to evolutionary theory. Some evidence tends to contradict it. Thus Selander et al. (1970) find that *Limulus*, a "living fossil," is as variable as other organisms. In any case, for such a hypothesis to be operationally useful one must be able to define taxa of the desired rank in order to measure their intrataxon variability.

A principle that has almost uniformly guided evolutionists in devising the most likely evolutionary trees has been that of minimum evolution. This can be construed as the minimum number of evolutionary steps (Camin and Sokal, 1965), the minimum number of mutational steps (Fitch and Margoliash, 1967), or a minimum length tree (Edwards and Cavalli-Sforza, 1964; Kluge and Farris, 1969). It is not easy to justify minimum length evolutionary trees in other ways than through the generally accepted principle of parsimony. However, Cavalli-Sforza and Edwards (1967) feel that such a principle cannot be accepted as an article of faith but must be justified on independent evidence. It is this type of evidence that by the nature of the problem is so hard to come by.

One must also decide whether minimum length trees should be rooted and directed. If so, one must postulate the nature of the ancestor and also whether the characters should be ordered from evolutionarily primitive to advanced and whether reversal of states should be permitted. Another implicit assumption that is often made, resting in part on the principle of parsimony, is that the common ancestor most probably possessed character states close to the central values of the states in present day organisms (i.e., it was similar to the centroid of extant OTU's). Certain other related concepts were discussed in our earlier volume (Sokal and Sneath, 1963, p. 230 ff.). On none of these issues have we found solid evidence either at the morphological or the molecular level, and some of the difficulties in the operational procedures for numerical cladistics described in the next section stem from this fundamental difficulty.

One aspect of cladistic analysis that is well known to evolutionary taxonomists, but whose implications are often neglected, is that ancestors are unknown: almost all of the work deals with recent organisms. Even if some recent organisms are shown to be close to the common ancestors of a set of OTU's, it is unlikely that the organisms living at present actually are unmodified descendants of these common ancestors. Some uncritical statements are frequently made for this reason. If cladogenies are deduced from present day organisms only, so that these are the only OTU's, then all earlier forms must be to some extent hypothetical and none of the OTU's should be shown as an ancestor of any other OTU. The caution must extend to fossil material. It is rare that a known fossil can be assumed to be a direct ancestor of a later fossil or extant form, though it may be phenetically close to the ancestor. More likely it is a form more similar to the common ancestor of both forms than is the extant form. Yet the construction of cladogenies of necessity involves branching points that are assumed to be the last common ancestors of the diverging lines. Whenever such cladogenies are based on suites of characters there is a temptation to reconstruct a description of the common ancestor from the postulated states for the entire suite of characters. Farris (1970) has called such postulated ancestors HTU's, for *hypothetical taxonomic units*. We must realize that the observed OTU's are points in a hyperspace that we are trying to connect in some defined optimal manner. The lines connecting these points are hypothesized as are most of the branching points (HTU's). But once a cladogram is constructed, the lines are very apt to assume some sort of reality in the minds of the beholder and so do the HTU's. One should always be on guard against this attitude.

A second problem is what lines or points should be studied. Should these represent individuals? The ancestry of a furcating line comprising many individuals gives a very complex network of reticulations and furcations, as Hennig (1966) has correctly pointed out. Do we therefore wish to make statements about some average of the mass of individuals when this mass has undergone a splitting process? Although it may sometimes be naively assumed that changes in sets of charac-

ters take place simultaneously in the internodes between the OTU's in a cladogram, this is in fact unlikely. Any given mutation may arise several times, dying out each time but eventually becoming established, or different mutations may occur at different times in any one line and thus heterochronous divergence or heterochronous parallelism may result. We have reasonably good evidence on these phenomena from studies of amino acid sequences in proteins. It is also possible that the cladistic separation at the species level occurred prior to the phenetic separation, as may well be the case in two sibling species that could be isolated by a sterility barrier before substantial phenetic differentiation.

6.4 NUMERICAL APPROACHES TO CLADISTIC ANALYSIS

General Considerations

In the published work on various numerical approaches to cladistic analysis there seem to be four main schools. Edwards and Cavalli-Sforza (1964) measure differences among OTU's in terms of gene frequencies at one or more loci expressed as angles. Distances between OTU's thus are continuous variables. These authors construct what seem to them plausible models of the evolutionary process and then suggest a variety of statistical procedures for fitting evolutionary trees to the data, given the models. Frequently such solutions are approximated by the construction of shortest length trees from these dissimilarity matrices. Camin and Sokal (1965) worked with discrete characters and attempted to find cladograms requiring the minimum number of evolutionary steps to represent the character state vectors of each OTU. Farris and coworkers in an extensive series of papers discussed at length below have used Manhattan distances between pairs of OTU's based on both continuous and discrete characters to form minimum length graphs and directed trees that estimate evolutionary sequences. Finally, investigators estimating protein phylogenies, for example, Fitch and Margoliash (1967) or Dayhoff (1969a,b), have computed directed trees in which the dissimilarity between amino acid sequences of proteins is expressed in nucleotide minimal mutation distances. Although the approaches of these schools differ, their results are frequently similar. They yield or approximate minimum length nondirected or directed trees based on a variety of similarity coefficients between OTU's. Some methods use similarity coefficients based on all characters, others weight characters by lowering the weight of those presumed to be convergent to obtain a better estimate of patristic similarity, because the latter should be more likely to reflect cladistic relationships than would overall phenetic similarity. We shall discuss such weighting methods later in this section.

Most similarity coefficients and clustering algorithms employed in numerical cladistics are also employed in numerical phenetics. The important distinction between phenetic and cladistic analysis lies not in the similarity coefficients or

clustering algorithms, therefore, but in the assumptions underlying their use in numerical cladistics and in the conclusions drawn from the results of the study.

In cladistics we encounter *minimally connected graphs* or *trees*. Three examples are shown in Figure 6-4. A new feature is that not all of the nodes (vertices) of the graphs are OTU's. Some will be postulated common ancestors (hypothetical taxonomic units, HTU's). Thus in Figure 6-4,*a*, we would interpret the nodes **a**, **c**, **f**, and **g** as OTU's (presumably extant organisms), while the nodes **b**, **e**, and **d** are HTU's (postulated ancestors; see Section 6.3). This situation differs from the phenetic graphs considered in Section 5.7 (Figure 5-14) where all nodes are (Recent) OTU's. As pointed out by Farris (1970) these graphs can be defined by a *connection function* g(**j**), where g(**j**) defines the node (OTU or HTU) connected to node **j** in some consistent manner. Thus in each of the three graphs in Figure 6-4, g(**a**) = **b**, g(**b**) = **d**, g(**c**) = **b**, g(**d**) = **e**, g(**e**) = **g**, g(**f**) = **e**. The trees all contain a unique base element **g**, reached by repeatedly applying connection function g(**j**) from any node. Function g(**g**) is undefined. Connection functions do not imply directionality. This is shown by the nondirected graph in Figure 6-4,*c*, Although the three graphs in Figure 6-4 give the appearance of being different, they represent topologically equivalent relationships. The trees can therefore be defined as ordered pairs W = (N, g) where N is a collection of nodes (OTU's or HTU's) in the graph and g is a function as defined above.

In a cladogram, directionality is added to yield a *minimally connected directed graph* or *directed tree*, and one of its nodes must be postulated to be the root. In the first two graphs in Figure 6-4 nodes **d** and **e**, respectively, are assumed to be the roots. A rooted connection function f(**j**) is called an *ancestor function* by Farris (1970), where f(**j**) defines the immediate ancestor of node **j**. In Figures 6-4,*a* and 6-4,*b* the directionality has been indicated by arrows. Repeated application of the ancestor function starting from any point will terminate at the root of the tree, which in a cladogram is the common ancestor. Thus the cladogram of Figure 6-4,*b* is rooted in the common ancestor **e**. Such a directed tree is called an *evolutionary tree* by Farris (1970) and defined as an ordered pair T = (N,f) where N is a collection of nodes (OTU's or HTU's) on the tree, and f is the ancestor function defined above. Estabrook (1968) has called T = (N, f) an *evolutionary hypothesis*. Readers wishing to consult Farris (1970) will be helped by knowing that he uses "simply connected network" for tree, and "tree" for directed tree. We have retained below his terms "Wagner network" and "Wagner tree" with the meanings coined by him to avoid terminological confusion.

For any sizeable number of Recent OTU's a large number of combinations for any one topology and a large number of different topologies can be formed. Cavalli-Sforza and Edwards (1967) report that $(2t - 3)!/[2^{t-2}(t - 2)!]$ different rooted trees may be recognized for t OTU's. When $t = 10$, this equals 34,459,425 trees. They also state that $(2t - 5)!/[2^{t-3}(t - 3)!]$ different nondirected trees can be found; this equals 2,027,025 when $t = 10$. Another way of looking at these

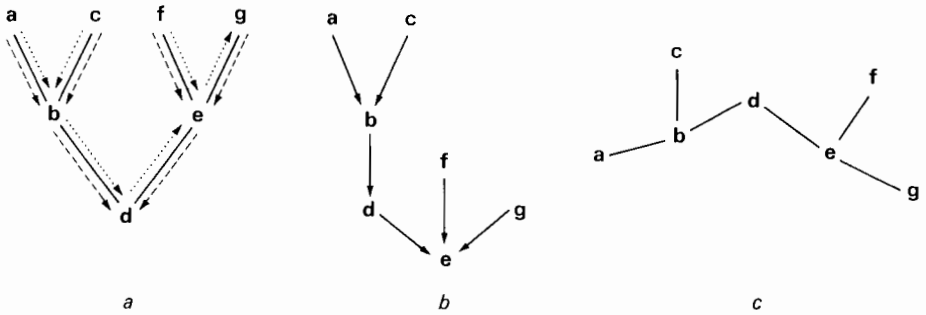


FIGURE 6-4

Three topologically equivalent trees (minimally connected graphs). Graphs in *a* and *b* are also directed trees (evolutionary trees) with roots at **d** and **e**, respectively. The direction of these trees is given by the ancestor function and is indicated by dashed arrows in *a*, and by the arrows in *b*. The connection function is shown in *a* by dotted arrows. For further explanation, see text. [Modified from Farris (1970).]

formulae is the observation by Fitch and Margoliash (1968) that the number of rooted trees for t OTU's equals the number of unrooted trees for $t + 1$ OTU's. The number of topologies (tree forms irrespective of which OTU's are placed on the terminal branches) is considerably less than the above. Cavalli-Sforza and Edwards computed 98 ways in which ten OTU's can be connected into a tree. The above computations limit the connections of one node to maximally three others. If more than three nodes can be connected to another one, the number of possible trees becomes even larger. Yet another complicating factor is the possibility of reticulate evolution, discussed at the end of this section. This greatly increases the number of possible configurations. Reversibility of evolution, discussed in the previous section, and the large number of possible tree forms are two of the fundamental problems of numerical cladistics. How can one give direction to evolutionary change as reflected in character state differences so that rooted trees rather than unrooted trees (simply connected graphs) are produced? And among the very large number of possible trees how is one to choose the most plausible one? The various techniques of numerical cladistics described below have made different approaches to a solution.

Trees

These nondirected graphs seem the most conservative representation of cladistic relationships. By setting up a minimum length graph among OTU's we are stating that the transition from one to the other most likely occurred along evolutionary pathways indicated by the edges or internodes of the tree. By leaving the relationship as a tree rather than as a cladogram (directed tree) the investigator is able to assign experimentally one node or the other as an ancestor and to make judgments about the reasonableness of the resulting cladogenies. The network

among OTU's will include additional nodes that help to minimize the overall length of the graph. The additional nodes will rarely if ever be extant organisms, and will normally be either fossils or HTU's. Such graphs are known in graph theory as *Steiner minimal trees* (Gilbert and Pollak, 1968). They are minimal length trees for t vertices (OTU's, terminal nodes). In order to achieve minimum length they may contain other vertices (HTU's, nodes, or in graph theoretical terminology —Steiner points) in addition to the t vertices.

The earliest numerical method for designing an evolutionary nondirected tree is that of Edwards and Cavalli-Sforza (1964), which was elaborated in several subsequent papers of which Cavalli-Sforza and Edwards (1967) is representative. The methods developed by these authors are applicable to continuous characters, but are primarily applied to gene frequencies at several loci. Edwards and Cavalli-Sforza prefer gene frequencies to other continuous characters such as morphometric ones, because they feel that they understand the underlying genetical mechanisms of these characters and therefore can make statements about the likelihood of selection, mutation, and random drift, which they could not do for morphometric characters. The gene frequencies are transformed by the cosine transformation (see Section 4.13).

Cavalli-Sforza and Edwards (1967) consider evolution as a branching process with recent OTU's in a "now" plane or hyperplane determined by the gene frequencies, with a normal to this hyperplane representing time. Branching points represent ancestral OTU's that are rarely known and are in fact HTU's. The projection of the tree on the hyperplane (the graph) is considered an estimate of the evolutionary topology. Although these authors considered various evolutionary models, they have mainly dealt with evolution as a branching Brownian-motion process at a constant rate for all characters at all times; populations do not become extinct but bifurcate at random time intervals.

Cavalli-Sforza and Edwards (1967) have worked on three basic approaches to a solution of the estimation problem. For uniform Brownian motion as applied to a Yule process they suggest the method of maximum likelihood. They treat this method in considerable mathematical detail but in fact are able to solve it only for the simplest cases, likely to be trivial in any actual evolutionary study. The reason for the difficulty is that the log-likelihood surfaces contain numerous singularities that lead to difficulties of estimation. In a recent paper Edwards (1970) showed that a maximum-likelihood solution can be found in principle but that considerable computational difficulties still stand in the way of an actual numerical solution. Because of these difficulties Cavalli-Sforza and Edwards have resorted to "intuitive approaches" that effectively are "minimum evolution" techniques. They wish to obtain a tree with minimal length on the phenetic hyperplane by introducing suitable branching points, thus giving a Steiner minimal tree in Euclidean space. To this end they suggest obtaining a shortest spanning tree based on the distances between the OTU's and HTU's, but this requires finding the optimal positions for

nodes (HTU's). Cavalli-Sforza and Edwards (1967) stress that the minimum-evolution approach is successful not because evolution actually proceeds parsimoniously but that the minimum-length tree solution is probably close to that of the projection of the maximum-likelihood tree on the "now" plane. Since the technique of computing a shortest spanning tree among a set of OTU's has already been discussed in Section 5.7, there is no need to deal further with it here. A third method by Cavalli-Sforza and Edwards is called the additive tree model, which is a least squares fit to a plausible tree structure. They believe that its justification is similar to that of the minimum evolution method. They also point out that hybridization, convergence, or parallelism cannot be handled by their model.

Applications of these methods have been made to human blood groups in 15 populations (Edwards and Cavalli-Sforza, 1964) and in four populations (Cavalli-Sforza and Edwards, 1967) from different racial stocks. Similarly, Goodman et al. (1971) describe a method that minimizes the difference between patristic differences and mutational differences in protein sequences.

Although various authors have used minimum length trees to generate evolutionary hypotheses, we owe a formal treatment of the subject to Farris (1970). In addition to the t terminal nodes (OTU's) the Steiner minimal tree will have a variable number of $t^* - t$ additional nodes (HTU's or Steiner points), where $t^* \geq t$, yielding a total of t^* nodes for the entire tree. He defines the length of the internode between two nodes **j** and **k** as the Manhattan distance

$$d_1(\mathbf{j}, \mathbf{k}) = \sum_{i=1}^{n} |X_{i\mathbf{j}} - X_{i\mathbf{k}}|$$

between the two nodes (see Section 4.3). The character states may be any real numbers and thus include continuous characters as well as cases where character states can assume only integral numbers, the "evolutionary steps" of Camin and Sokal (1965). The length of a tree θ with t^* nodes is defined as

$$L(\theta) = \sum_{\mathbf{j} \neq \mathbf{q}} d_1(\mathbf{j}, \mathrm{g}(\mathbf{j})),$$

which is the summation of the $t^* - 1$ internode lengths $d_1(\mathbf{j}, \mathrm{g}(\mathbf{j}))$ within the tree. There is one less internode than the number of nodes t^*, since $d_1(\mathbf{q}, \mathrm{g}(\mathbf{q}))$ is undefined for the unique base element **q**. Shortest spanning trees can be computed, minimizing the length of the network defined as above. An efficient algorithm and a FORTRAN IV program for their computation is furnished by Farris (1970).

A *Wagner network* as defined by Farris (1970) is a Steiner minimal tree, measured in (multidimensional) Manhattan distance. Such trees in two dimensions were considered by Hanan (1966). Wagner networks also permit reversibility of character states. Thus if the character states for a given character i for two nodes are

$X_{ij} = 2$ and $X_{ik} = 3$, respectively, the distance $d_1(\mathbf{j}, \mathbf{k}) = |X_{ij} - X_{ik}| = 1$ may imply the change $2 \rightarrow 3$, as well as $3 \rightarrow 2$ within the same network. Wagner trees are based on concepts developed by W. H. Wagner (e.g., Wagner 1963, 1969) and illustrated by work such as that of Mickel (1962), Lellinger (1964), and Kesling and Sigler (1969). Farris shows that the character state $X_{i\mu}$ for any character i of HTU $\boldsymbol{\mu}$ connecting a triad of OTU's **a**, **b**, and **j** (see Figure 6-5) is the median value of the three character states X_{ia}, X_{ib}, and X_{ij}, if the sum of the Manhattan distance lengths of the internodes $d_1(\mathbf{a}, \boldsymbol{\mu})$, $d_1(\mathbf{b}, \boldsymbol{\mu})$, and $d_1(\mathbf{j}, \boldsymbol{\mu})$ is to be minimized. In this manner the HTU for these OTU's can be constructed as the vector of median states for all characters of the three connected OTU's.

An algorithm for constructing Wagner networks is given by Farris (1970) as follows. We find the pair of OTU's **a** and **b** whose distance $d_1(\mathbf{a}, \mathbf{b})$ is greater than that between any other pair of OTU's in the study. We compute the distance of all other OTU's **j** with the internode (**a**, **b**). This is done by computing the Manhattan distance $d_1(\mathbf{j}, \boldsymbol{\mu})$ between each OTU $\mathbf{j} \neq \mathbf{a}$ or **b** and the median character state vector $\boldsymbol{\mu}$ for each triad **a**, **b**, **j**. To simplify this process, Farris suggests computing the distance between **j** and a point representing the internode (**a**, **b**). This point is called an *interval* (symbolized by INT). This is given as

$$d_1(\mathbf{j}, \mathrm{INT}(\mathbf{a}, \mathbf{b})) = \tfrac{1}{2}[d_1(\mathbf{a}, \mathbf{j}) + d_1(\mathbf{b}, \mathbf{j}) - d_1(\mathbf{a}, \mathbf{b})]$$

Farris has shown that $d_1(\mathbf{j}, \mathrm{INT}(\mathbf{a}, \mathbf{b})) = d_1(\mathbf{j}, \boldsymbol{\mu})$. The OTU, say **c**, with the greatest distance $d_1(\mathbf{c}, \boldsymbol{\mu})$ is chosen and an HTU $\boldsymbol{\mu}_1$ is constructed as the median of OTU's **a**, **b**, and **c** (see Figure 6-6,*a*). The connection function for the network is defined as $g(\mathbf{a}) = \boldsymbol{\mu}_1$, $g(\mathbf{b}) = \boldsymbol{\mu}_1$, $g(\mathbf{c}) = \boldsymbol{\mu}_1$ and one proceeds to look for the unplaced OTU, say **e**, with the next greatest distance $d_1(\mathbf{e}, \boldsymbol{\mu})$ to internode (**a**, **b**). We now find the "interval" on the network constructed so far that is closest to the new OTU **e**. As before, we construct an HTU ($\boldsymbol{\mu}_2$) to represent this "interval" and connect the triad through $\boldsymbol{\mu}_2$. For example, if $d_1(\mathbf{e}, \mathrm{INT}(\mathbf{a}, \boldsymbol{\mu}_1))$ had been less than either $d_1(\mathbf{e}, \mathrm{INT}(\mathbf{b}, \boldsymbol{\mu}_1))$ or $d_1(\mathbf{e}, \mathrm{INT}(\mathbf{c}, \boldsymbol{\mu}_1))$, we would have constructed the internodes (**a**, $\boldsymbol{\mu}_2$),

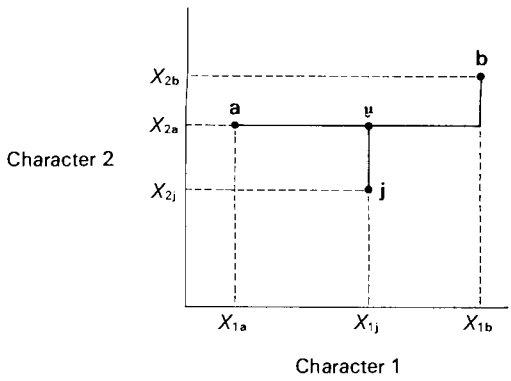


FIGURE 6-5
The calculation of character states of an HTU. The three OTU's are **a**, **b**, and **j**, with character state values on characters *1* and *2*. The HTU $\boldsymbol{\mu}$ is given the median values of the states of **a**, **b**, and **j** for each character, and the total length of the three internodes (solid lines), is then a minimum. For further details, see text.

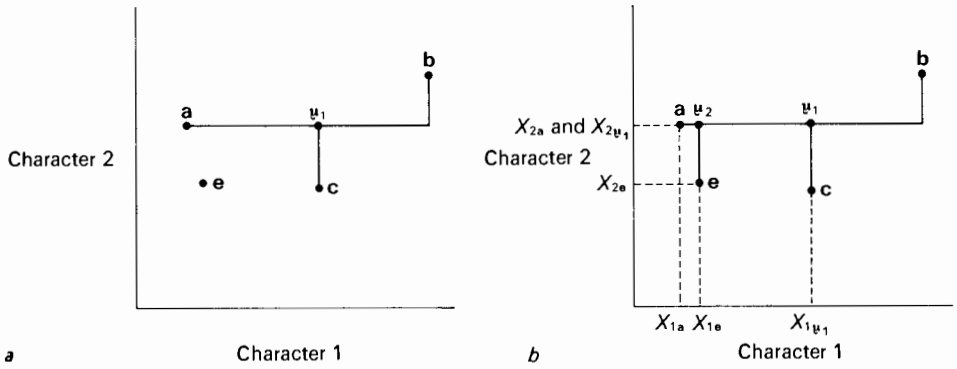


FIGURE 6-6

Construction of HTU's in making a Wagner network, as explained in the text. *a*, Construction of an HTU, μ_1 from OTU's **a**, **b**, and **c**. A fourth OTU, **e**, remains unconnected at this stage. *b*, Addition of OTU **e** to the network, with the construction of a new HTU, μ_2. This HTU is constructed as the median of the states for **e**, **a**, and μ_1 since μ_2 is closer to **e** (using the Manhattan metric) than the alternative new nodes for the triads **e**, **b**, μ_1, or **e**, **c**, μ_1. These alternative new nodes would be situated at the same point as μ_1 and just above **c** respectively (if the first had been required it would not have received a new symbol).

(μ_1, μ_2), and $(\mathbf{e}, \mu_2)$, with new connection functions defined as $g(\mathbf{a}) = \mu_2$, $g(\mu_1) = \mu_2$, and $g(\mathbf{e}) = \mu_2$ (see Figure 6-6,*b*). Functions $g(\mathbf{b}) = \mu_1$ and $g(\mathbf{c}) = \mu_1$, as before. The algorithm continues until all OTU's have been connected to the network. Another version of the algorithm searches for the OTU with the greatest distance to any of the available internodes. After the first cycle these will be internodes $(\mathbf{a}, \mu_1)$, $(\mathbf{b}, \mu_1)$, and $(\mathbf{c}, \mu_1)$. The new OTU is then linked to that internode with which it has the smallest Manhattan distance, again using the combinatorial formula given above. A new hypothetical taxonomic unit μ_2 is constructed intermediate between the two OTU's bounding the chosen internode and the most recent candidate for joining the network. The process continues until all candidate OTU's have been attached to the network by means of HTU's.

To illustrate the procedure for finding a Wagner network, we apply the first of the two algorithms proposed by Farris (1970) to the data matrix illustrated in Table 6-1. These are a subset of the Caminalcules (Camin and Sokal, 1965) mentioned in several places in this book. For reasons that will become apparent later we shall delete character (row) *4* from the data matrix for the computations to follow and base the Wagner network to be constructed on characters *1* through *3*, and *5* through *7*. Table 6-2 gives the successive steps necessary to obtain the Wagner network and in Figure 6-7 a graphic representation can be found. The length of the resulting network, $\Sigma d_1(\mathbf{j}, \mathbf{k}) = 15$, is minimal, since the sum of the absolute values of the ranges of the characters (the last column of Table 6-1) is indeed 15.

Necessarily, however, distances along the network between pairs of OTU's may be longer than in the initial distance matrix based on their character states. Thus, although $d_1(7,25) = 4$, the length of the path along the network between these OTU's adds up to 6.

TABLE 6-1

Data matrix for seven selected Caminalcules (Group A). [From Camin and Sokal, 1965.]

Characters	t OTU's							Number of character states, m_i	Range or minimum number of steps $m_i - 1$
	7	**8**	**13**	**14**	**15**	**25**	**28**		
1	1	0	1	0	1	1	1	2	1
2	1	1	0	0	0	2	0	3	2
3	1	1	1	2	0	2	2	3	2
4	3	2	0	0	3	1	0	4	3
5	1	1	2	1	0	1	3	4	3
6	1	0	0	0	0	−1	0	3	2
7	0	0	0	0	−1	0	1	3	2
									$\sum_{i=1}^{n} (m_i - 1) = 15$

TABLE 6-2

Finding a Wagner network

A. Manhattan distance matrix based on data matrix for Group A of the Caminalcules (Table 6-1 with character *4* omitted. The HTU's, numbered $\mu_1, \ldots, \mu_5$, are constructed as explained below.

OTU's and HTU's	OTU's and HTU's											
	7	**8**	**13**	**14**	**15**	**25**	**28**	μ_1	μ_2	μ_3	μ_4	μ_5
7	×											
8	2	×										
13	3	3	×									
14	4	2	3	×								
15	5	5	4	5	×							
25	4	4	5	4	7	×						
28	6	6	3	4	7	6	×					
μ_1	3	3	2	1	4	3	3	×				
μ_2	2	2	1	2	3	4	4	1	×			
μ_3	1	1	2	3	4	3	5	2	1	×		
μ_4	2	2	1	2	3	4	4	1	0	1	×	
μ_5	3	3	2	1	4	3	3	0	1	2	1	×

TABLE 6-2—(*continued*)

B. Matrix of character states for HTU's.

Characters	HTU's				
	μ_1	μ_2	μ_3	μ_4	μ_5
1	1	1	1	1	1
2	0	0	1	0	0
3	2	1	1	1	2
4	×	×	×	×	×
5	1	1	1	1	1
6	0	0	0	0	0
7	0	0	0	0	0

Step 1.
Take d_1(**15**,**25**) = 7. This is greater than other distances except for d_1(**15**,**28**), which equals it. In this instance and all others when decisions among criteria of equal magnitude must be made, the algorithm arbitrarily chooses the distance involving an OTU or HTU lowest in numerical order. The first internode is drawn. See Figure 6-7,*a*.

Step 2.
Compute the distance $d_1(\mathbf{j},\boldsymbol{\mu})$ for all OTU's **j** other than **15** and **25** to the HTU $\boldsymbol{\mu}_j$ representing the interval INT(**15**,**25**) for the given **j**. We employ the formula given in the text. For example, $d_1(7,\text{INT}(\mathbf{15},\mathbf{25})) = \frac{1}{2}[d_1(7,\mathbf{15}) + d_1(7,\mathbf{25}) - d_1(\mathbf{15},\mathbf{25})] = \frac{1}{2}[5 + 4 - 7] = 1$. Choose the highest $d_1(\mathbf{j},\boldsymbol{\mu})$, which is $d_1(\mathbf{28},\text{INT}(\mathbf{15},\mathbf{25})) = 3$.

Step 3.
HTU $\boldsymbol{\mu}_1$ is constructed between OTU's **15**, **25**, and **28**. The character states of this vector are the median states for the OTU's. Thus for character *1* the respective states are 1, 1, 1; the median is clearly 1. For character *2* the states are 0, 2, 0; the median state is 0. For character *5* the states are 0, 1, 3; the median state is 1. In this manner we obtain the character state vector of $\boldsymbol{\mu}_1$ as $\{1, 0, 2, 1, 0, 0\}$. The HTU $\boldsymbol{\mu}_1$ is drawn in to connect the three OTU's **15**, **25**, and **28** in Figure 6-7,*b*.

Step 4.
Choose the next highest $d_1(\mathbf{j},\boldsymbol{\mu})$ to the internode (**15**,**25**). Since all remaining OTU's, **7**, **8**, **13**, and **14**, have the same $d_1(\mathbf{j},\boldsymbol{\mu})$ we employ the arbitrary rule stated above and take the distance in order of OTU number. Therefore **7** is chosen and we compute $d_1(7,\text{INT}(\mathbf{j},\boldsymbol{\mu}_1))$ for all existing internodes (that is, **j** = **15**, **25**, or **28**). We choose the internode for which $d_1(7,\text{INT}(\mathbf{j},\boldsymbol{\mu}_1))$ is least. This is $d_1(7,\text{INT}(\mathbf{15},\boldsymbol{\mu}_1)) = 2$.

Step 5.
HTU $\boldsymbol{\mu}_2$ is constructed between OTU's **7** and **15** and HTU $\boldsymbol{\mu}_1$, in the same manner as in Step 3 above. The new character state vector for $\boldsymbol{\mu}_2$ is $\{1, 0, 1, 1, 0, 0\}$. The new HTU is drawn in to connect **7** to the network. See Figure 6-7*c*.

Step 6.
These procedures are continued, with OTU **8** joining internode (**7**,$\boldsymbol{\mu}_2$) via HTU $\boldsymbol{\mu}_3$, OTU **13** joining internode (**15**,$\boldsymbol{\mu}_2$) via HTU $\boldsymbol{\mu}_4$ and OTU **14** joining internode (**25**,$\boldsymbol{\mu}_1$) via HTU $\boldsymbol{\mu}_5$. When all OTU's are connected the network looks like Figure 6-7,*d*.

Step 7.
Since two pairs of HTU's, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_5$, and $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_4$, have zero distance between them they are deleted from the network that can then be represented in its simplest form as in Figure 6-7,*e*. The sum of the lengths of the internodes is 15.

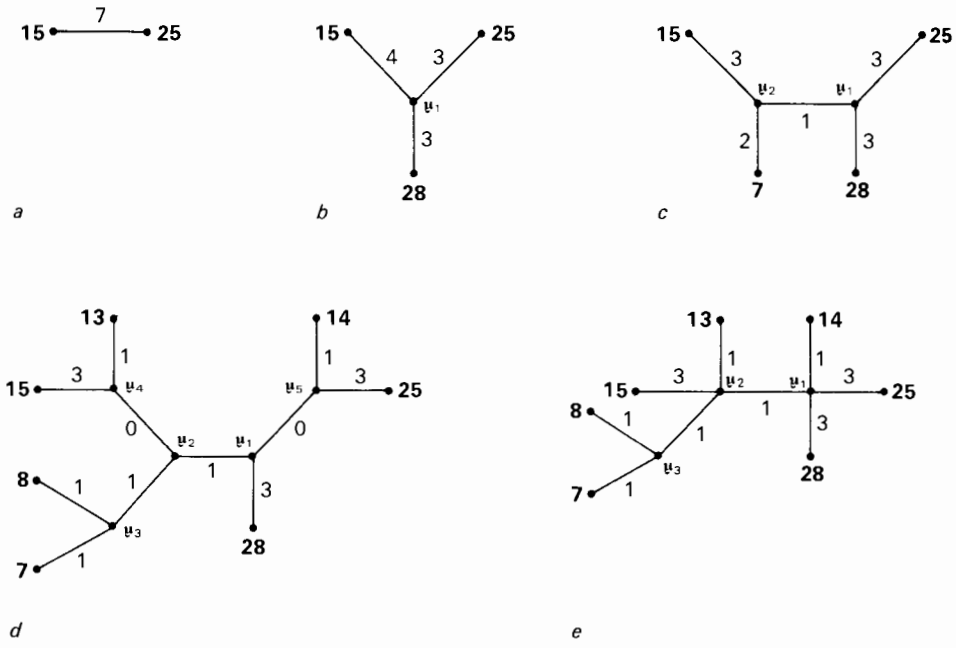


FIGURE 6-7
Construction of a Wagner network for Group A of the Caminalcules. OTU's are numbered, HTU's are subscripted **μ**. The various steps are explained in detail in Table 6-2.

In other data sets the HTU's produced during the successive clustering of OTU's to form the network are not likely to be optimal ones in the sense of minimizing the length of the entire network. There will be local minima but the tree may not be of minimum overall length. An iterative procedure for optimization can then be applied (Farris, 1970). When the internodes are arbitrarily assigned directions, the method is equivalent to the optimization procedure described below for Wagner trees. A worked application of Wagner networks in paleontology is given by Kesling and Sigler (1969), who also used character weighting similar to p_i of Farris discussed later in this section. It is important to repeat that this method permits the reversal of character states.

Cladograms

Cladograms are evolutionary reconstructions in the form of rooted directed trees. Although Camin and Sokal (1965), in their numerical cladistic method, dealt only with discrete characters, Farris (1970) generalized this to characters described by any real number. Camin and Sokal made the following four assumptions about their data sets. (1) Characters can be expressed in discrete states differing among at least some of the OTU's of the study. (2) The character states for any one character

can be arrayed in evolutionary order; thus any one character yields a character state tree (see Figure 6-8). (3) The most ancestral state in any study arose only once within the collection of OTU's under study. (These primitive character states, conventionally coded zero, are the unique character states of Wilson, 1965). (4) Evolution is irreversible. Thus a descendent character state cannot revert to an ancestral character state.

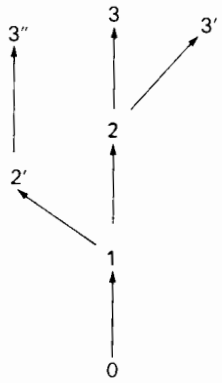


FIGURE 6-8
A character state tree. In this character the primitive state is 0. State 1 has given rise to two different states, labelled 2 and 2′. The latter has given rise to 3″, and the former has changed to both 3 and 3′.

Camin and Sokal (1965) phrased their methodology in terms of "evolutionary steps," the number of changes in discrete character state values needed to proceed from a prior (ancestral) OTU in a cladogram to a subsequent (descendent) OTU. This number of evolutionary steps was evaluated over all characters examined in a given study and was shown graphically by cross marks on the cladogram (see Figure 6-9). However, it is readily apparent that the number of evolutionary steps between an ancestral HTU **j** and a descendent OTU **k** on a cladogram is nothing but the Manhattan distance $d_1(\mathbf{j}, \mathbf{k})$ between these nodes over all characters. As an example let us examine the number of evolutionary steps or the Manhattan distance between OTU **25** and its immediate ancestor **a**, based on the data matrix in Table 6-1 and illustrated in Figure 6-9 (from which the states of **a** can be obtained). The character state vectors for these two OTU's are the following:

Character	Character state vectors for OTU's	
	25	**a**
1	1	1
2	2	1
3	2	1
4	1	1
5	1	1
6	−1	0
7	0	0

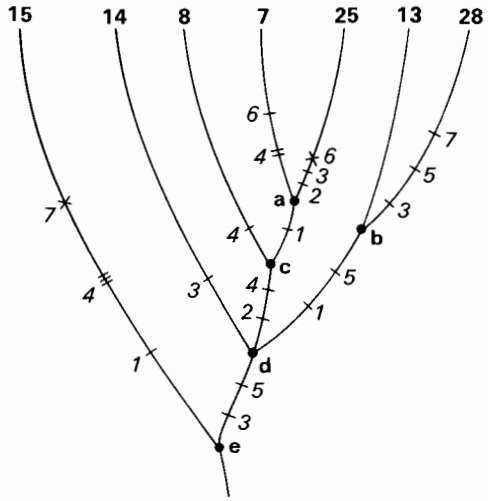


FIGURE 6-9

Reconstruction of a cladogram. The OTU's are at the tips, and are the Group A Caminalcules in Camin and Sokal (1965). They are identified by the numerals used in that paper.

The HTU's that are ancestors of the OTU's at the tips are indicated by black circles and identified by letters **a** to **e**. Their probable character states can be easily obtained from the cladogram by going from the base (whose states are all zero) to the HTU. A cross-bar indicates a change in character state for the character whose number is placed beside it, producing an increase by one unit, e.g., from state 0 to state 1. A cross-sign indicates a decrease by one unit, e.g., from state 0 to state −1. Thus for **c**, characters *1* through *7* will be 0, 1, 1, 1, 1, 0, 0 respectively. [Modified from Camin and Sokal (1965).]

By inspection, their Manhattan distance $d_1(\mathbf{25}, \mathbf{a})$ can be computed as 3, since the two OTU's differ by one evolutionary step (absolute differences of one) for characters *2*, *3*, and *6* only.

The character state vectors for any OTU can be thought of as points in a hyperdimensional lattice in which, because of the irreversibility assumption, the OTU's follow a monotonically increasing ($\geq$) or decreasing ($\leq$) trajectory for any one dimension. This concept will become clearer when looked at in a two-dimensional diagram (see Figure 6-10). OTU's **a** and **b** represent the following character state vectors.

Character	Character state vectors for OTU's	
	a	**b**
h	1	2
i	2	1

Panel *a* shows independent evolutionary trajectories for these two OTU's. Parallelism is present in characters *h* and *i*. Panels *b* and *c* show equivalent minimum length cladistic paths. Both OTU's share the evolutionary path part of the way. The cladistic representation of panel *a* is shown in cladogram *d* of Figure 6-10, that of panels *b* and *c* in cladogram *e*. Obviously, the maximum length of a cladogram will be the sum of the independent paths of each OTU, emerging as a separate line from a common ancestor—V-shaped for two OTU's in Figure 6-10,*d*, brushlike when there are many OTU's. Thus, using the convention that all character states in the common ancestor are equal to 0, the maximum distance from any OTU **j** to the common ancestor **o** must be $d_1(\mathbf{j}, \mathbf{o}) = \Sigma_{i=1}^{n} |X_{i\mathbf{j}}|$. Consequently the maximum length of the entire cladogram comprising *t* OTU's will be the sum of the separate evolutionary lengths for each OTU, $L_{\max} = \Sigma_{\mathbf{j}=1}^{t} \Sigma_{i=1}^{n} |X_{i\mathbf{j}}|$.

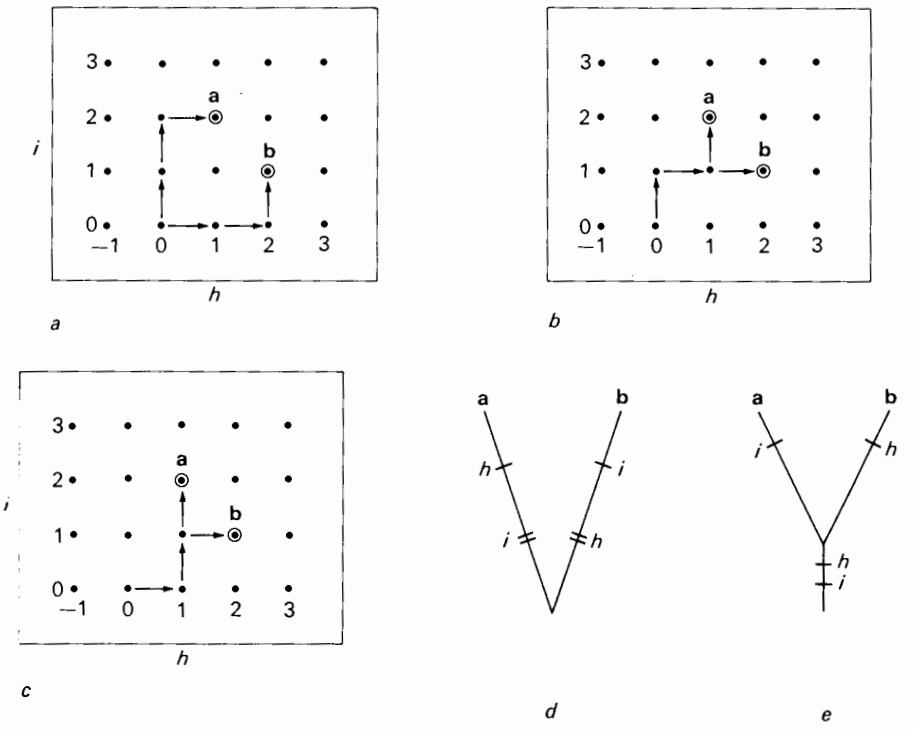


FIGURE 6-10

Three representations of an equivalent cladistic sequence. Two characters, *h* and *i*, are shown, and −1, 0, 1, 2 and 3 are character states. The two OTU's **a** and **b** have states (1,2) and (2,1) respectively. *a*, A maximum-length cladistic path (reversals not permitted). *b*, A minimum-length cladistic path, using the point (0,1) as an intermediate vertex to represent an HTU. *c*, The other minimum-length path, using (1,0) as an intermediate vertex. *d*, The cladogram representation equivalent to (*a*). *e*, The cladogram representation equivalent to *b*; that for panel *c* differs only in the order of characters *h* and *i* at the lowest internode, and since this order is arbitrary, *e* can represent both *b* and *c*. [Modified from Hendrickson (1968).]

Certain evolutionary steps can be *shared*, that is, changes in some OTU's occurred before their last common ancestor and can be placed on a single internode. Then the total length of the tree is diminished by the sum of the lengths of the shared internodes, each multiplied by $(t_{\mathbf{J}} - 1)$, where $t_{\mathbf{J}}$ is the number of terminal OTU's subtended by the shared internode. Thus, to compute the length of the cladogram in Figure 6-10,*e* we compute $(1 + 2) + (2 + 1) - (1 \times 2) = 4$. At the other extreme, the minimum length for a cladogram assuming the maximum possible number of shared steps for all OTU's, no parallelisms, and complete consistency among the characters, must be $L_{\min} = \Sigma_{i=1}^{n}(m_i - 1)$. In other words the absolute theoretical minimum length cannot be less than the number of character states less 1, summed over all the characters. In virtually all cladograms the actual number of evolutionary steps will be somewhere between $L_{\max}$ and $L_{\min}$, and it is the aim of numerical cladistic methods to minimize the overall length of the cladogram (or the number of evolutionary steps).

In a search for most parsimonious (shortest length) cladograms, Camin and Sokal (1965) developed a series of initial approaches that constructed cladograms close to a most parsimonious solution. A variety of clustering algorithms can be used to obtain first approximate cladograms (called procladograms by Camin and Sokal). One such procedure is a single linkage cluster analysis (see Section 5.5) of the pair function describing the number of common evolutionary steps (Camin and Sokal, 1965). This quantity has been called the total number of derived steps shared by OTU's **a** and **b**, and it has also been called the advancement index $h(\mathbf{j}_{\mathbf{a},\mathbf{b}})$ of the common joint ancestor $\mathbf{j}_{\mathbf{a},\mathbf{b}}$, which is the most recent common ancestor of OTU's **a** and **b** (Farris, Kluge, and Eckardt, 1970). We have adopted here the simpler symbolism consistent with the rest of our notation rather than the $h(\mathbf{J}(\{\mathbf{a}, \mathbf{b}\}))$ and $\mathbf{J}(\{\mathbf{a}, \mathbf{b}\})$ employed by Farris et al. (1970).

Camin and Sokal (1965) proposed the monothetic method for obtaining procladograms. Its algorithm runs as follows.

1. Count the number of zero character states for each OTU. (This provides some measures of "primitiveness," as OTU's with more characters of state zero should branch off the main trunk of the cladogram near its base). Go to 2.

2. Remove the OTU with the greatest number of zeros from the data matrix. Test the remaining data matrix for rows (characters) without zeros. If no such rows are found, replace the removed OTU and remove another one tied with it by number of zeros or by possession of the next largest number of zeros. If removal of any one OTU with a high number of zeros does not produce nonzero character rows, combinations of two or even three OTU's with large numbers of zero character states should be removed from the matrix. Once a nonzero row (or rows) appears, go to 3.

3. Draw a branch subtending the removed OTU or OTU's from the base of the cladogram. Go to 4.

4. Subtract unity from each character state code in each nonzero row (repeatedly

if necessary) until each row contains at least one zero. Rows that have become all zero are dropped from the matrix and from further consideration. Go to 5.

5. Are any OTU's unplaced on the branches? If yes, go to 1; if no, the procladogram is complete.

An example of such an algorithm is illustrated in Table 6-3. In this table characters having both negative and positive character states are, for convenience, recoded into all positive character states. Character *4* has been omitted as unreliable for reasons discussed later in this section. The algorithmic steps in Table 6-3 are self-explanatory. At the end of this procedure a procladogram is obtained, shown in Figure 6-11,*a*.

Camin and Sokal (1965) reduced the length of the procladogram by an iterative procedure. They first removed all internodes that do not bear any evolutionary steps (or whose lengths are zero on the Manhattan metric). This is done because there is no reason to assume separate branching points in the absence of intervening evolutionary steps. Next, all common evolutionary steps found on adjacent branches are moved to the base stem of these branches. This is followed by a trial and error moving of branches illustrated in Figure 6-11. In the procladogram resulting from the monothetic method (Figure 6-11,*a*), we move one of the terminal branches (that bearing OTU's **25** and **28**) down one branching level to level 3. This now becomes the terminal level (level ω) and OTU's **25**, **28**, **7** and **13** all emerge from this point (Figure 6-11,*b*). The length of the original procladogram was 17 (17 evolutionary steps), but the new length is 18 and is thus moving away from the intended direction. However, members of the cluster (**25**, **28**, **7**, and **13**) can now be examined for common steps that can be removed to the internode at the base. In Figure 6-11,*c* we note that OTU's **25** and **7** can be placed together with a step for character *2* in common and OTU's **28** and **13** can be joined with a step for character *5* in common. This necessitates parallel steps in OTU's **25** and **28** for character *3*, which previously was a common step, but we have now reduced the number of steps for the cladogram to 16, i.e., its length is 16. This is the same number required by the true cladogram, but Figure 6-11,*c* is not the correct solution. Moving the branch that bears OTU's **7** and **25** in Figure 6-11,*c* from branching level 3 to level 2 (Figure 6-11,*d*) and shifting OTU **8** to share a step of character *2* with (**7**, **25**) yields the correct cladogram of Figure 6-9. This illustrates that different but equally parsimonious solutions (trees of equal length) may occur and it is impossible to decide among them unless evidence from other characters can be relied upon.

The entire cladistic method of Camin and Sokal—compatibility matrix computation (see below), procladogram construction, and the iterative search for the most parsimonious tree—has been programmed (Bartcher, 1966) and has been found to yield generally satisfactory results. Its application to a cladogeny of the horses, the Caminalcules, and several other groups has been promising (Camin and Sokal, 1965).

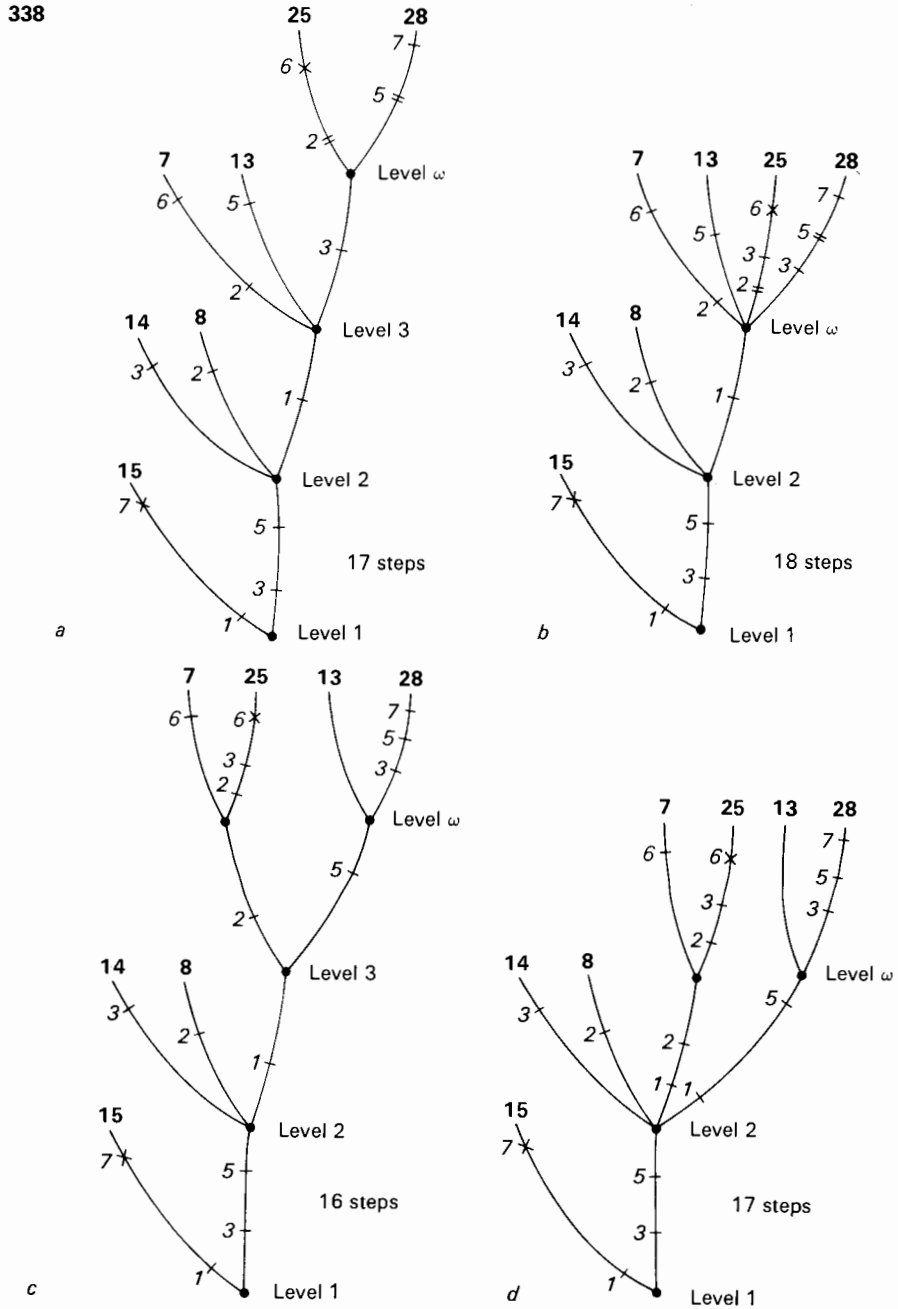


FIGURE 6-11

Steps in the reconstruction of the cladogram of Group A (Figure 6-9) by the monothetic method. Symbolism as in Figure 6-9. Level numbers and ω refer to levels of furcation from 1 to terminal level. *a*, Procladogram resulting from monothetic technique illustrated in Table 6-3. Total number of evolutionary steps is 17. *b*, OTU **25** moved down one branching point; 18 evolutionary steps result. *c*, OTU's **25** and **7** grouped, as are OTU's **13** and **28**. Achieved parsimony of 16 steps. This is equally parsimonious but not identical to the cladogram in Figure 6-9. Character *4* has been omitted from these cladograms. Further adjustments for parsimony result in a cladogram identical to Figure 6-9. *d*, Branch bearing OTU's **7** and **25** moves down one branching point; 17 evolutionary steps result. If OTU's **8**, **7**, and **25** are now rearranged so that they share their common step for character *2*, the cladogram of Figure 6-9 is obtained with 16 evolutionary steps. [Modified from Camin and Sokal (1965).]

TABLE 6-3

Camin and Sokal's monothetic method for reconstructing cladograms.

A. Data matrix for Caminalcule Group A (Table 6-1) with characters *6* and *7* recoded and character *4* omitted.

Characters	OTU's							Cycle 1, Step 2
	7	**8**	**13**	**14**	**15**	**25**	**28**	
1	1	0	1	0	1	1	1	OTU's **14** and **15** have 6 zeros each. Removal of OTU **14** leaves unchanged the number of "nonzero" rows. Removal of OTU **15** leaves rows *3* and *5* nonzero and *7*− all zero. Therefore, remove OTU **15**.
2	1	1	0	0	0	2	0	
3	1	1	1	2	0	2	2	
5	1	1	2	1	0	1	3	
6+	1	0	0	0	0	0	0	
6−	0	0	0	0	0	1	0	
7+	0	0	0	0	0	0	1	
7−	0	0	0	0	1	0	0	
Number of **Step** 1 zeros	3	5	5	6	6	3	4	

B. *Data matrix A with OTU* **15** *removed.*

Characters	OTU's						Cycle 1, Step 4
	7	**8**	**13**	**14**	**25**	**28**	
1	1	0	1	0	1	1	Subtract unity from rows *3* and *5*; delete row *7*−.
2	1	1	0	0	2	0	
3	1	1	1	2	2	2	
5	1	1	2	1	1	3	
6+	1	0	0	0	0	0	
6−	0	0	0	0	1	0	
7+	0	0	0	0	0	1	
7−	0	0	0	0	0	0	

C. Data matrix B with unity subtracted from rows *3* and *5* and row *7*− deleted.

Characters	OTU's						Cycle 2, Step 2
	7	**8**	**13**	**14**	**25**	**28**	
1	1	0	1	0	1	1	Recompute number of zeros for remaining OTU's. OTU's **8** and **14** have 6 zeros each. Removal of either **8** or **14** leaves unchanged the number of nonzero rows. Removal of both **8** and **14** leaves row *1* nonzero. Therefore, remove OTU's **8** and **14** together.
2	1	1	0	0	2	0	
3	0	0	0	1	1	1	
5	0	0	1	0	0	2	
6+	1	0	0	0	0	0	
6−	0	0	0	0	1	0	
7+	0	0	0	0	0	1	
Number of **Step** 1 zeros	4	6	5	6	3	3	

TABLE 6-3 (*continued*)

D. Data matrix C with OTU's **8** and **14** removed.

Characters	OTU's				Cycle 2, Step 4
	7	**13**	**25**	**28**	
1	1	1	1	1	Subtract unity from row *1*. Row *1* becomes all zero, so
2	1	0	2	0	delete.
3	0	0	1	1	
5	0	1	0	2	
6+	1	0	0	0	
6−	0	0	1	0	
7+	0	0	0	1	

E. Data matrix D with row *1* deleted.

Characters	OTU's				Cycle 3, Step 2
	7	**13**	**25**	**28**	
2	1	0	2	0	Recompute number of zeros for remaining OTU's
3	0	0	1	1	OTU **13** has 5 zeros. Removal of OTU **13** makes no
5	0	1	0	2	nonzero rows. Similarly for OTU **7** with 4 zeros
6+	1	0	0	0	Removal of OTU's **7** and **13** leaves row *3* nonzero
6−	0	0	1	0	Therefore, remove OTU's **7** and **13**.
7+	0	0	0	1	
Number of Step 1 zeros	4	5	3	3	

F. Data matrix E with OTU's **7** and **13** removed.

Characters	OTU's		Cycle 3, Step 4
	25	**28**	
2	2	0	Subtract unity from row *3*. Row *3* becomes all zero.
3	1	1	so delete. OTU's **25** and **28** are a terminal bifurcation
5	0	2	
6+	0	0	
6−	1	0	
7+	0	1	

[Modified from Camin and Sokal, 1965.]

However, the method of Camin and Sokal, while providing the user with numerous equally parsimonious solutions, had no criterion that indicated to the user when a minimum length tree was found or how many such minimum length trees there might be. This problem was solved by Estabrook (1968); using the topological technique of partial orders, he was able to find a general solution to obtaining the entire set of minimum length trees for a given collection of OTU's and characters. Since this set can sometimes be very large, one may need a substantial number of characters in order to reduce its size. We do not as yet have enough experience with the Estabrook procedure (programmed for computer processing by its author) to enable us to judge its usefulness, and to know whether increasing the number of characters substantially would reduce the number of equally parsimonious solutions for any given collection of OTU's.

The *weighted invariant step strategy* (WISS) of Farris, Kluge, and Eckardt (1970) consists of finding pairs of mutually highest advancement indexes (those pairs that have the highest number of shared derived steps) and replacing each pair by their most recent joint common ancestor. This approach has also been suggested by J. Hendrickson, Jr. (personal communication). The clustering cycle is repeated until all OTU's have been clustered. This method is thus a modification of a weighted pair group analysis using the total number of derived steps shared by OTU's **j** and **k** as the similarity coefficient. Since the similarity coefficient based on highest number of shared derived steps is a Manhattan metric if the primitive character is coded zero, we cannot recompute the dissimilarity coefficient between a recent joint common ancestor and a new branch by the combinatorial formula of Lance and Williams (1967b; see Section 5.5). However, Farris, Kluge, and Eckardt (1970) provide a computational formula for this operation,

$$h(\mathbf{j}_{\mathbf{a},\mathbf{b}}) = \tfrac{1}{2}[h(\mathbf{a}) + h(\mathbf{b}) - d_1(\mathbf{a}, \mathbf{b})]$$

similar to Farris' earlier mentioned formula for $d_1(\mathbf{j}, \boldsymbol{\mu})$. The clustering technique of Farris, Kluge, and Eckardt (1970) is very similar to that of Camin and Sokal (1965), except that the candidates for joining clusters are defined by their advancement index, $h(\mathbf{j}) = \Sigma_{i=1}^{n} |X_{ij}|$, which is the sum of the column vector of their absolute character states, whereas Camin and Sokal choose their candidates by the number of zero character states in the column vector. Thus it is possible that an OTU that was coded 1 for each character state might be preferred by the WISS method, although one that has many zeros but a few large character states might be preferred by Camin and Sokal's monothetic method.

Methods for finding *Wagner trees* are also given by Farris (1970). Since a Wagner tree is simply a directed Steiner minimal tree in Manhattan distance for a given set of OTU's (a Wagner network being a nondirected minimal Steiner tree for the same set of OTU's), any tree generated from a Wagner network by assuming one of the OTU's or HTU's to be the ancestor is a legitimate candidate for a Wagner

tree. Another algorithm suggested by Farris (1970) would assume an ancestor **e** for a given study. The ancestor could be an OTU (considered very primitive) or it could be a vector representing a hypothetical taxonomic unit. We compute all distances $d_1(\mathbf{j}, \mathbf{e})$ from the t OTU's to the ancestor. The OTU, say **b**, with the smallest $d_1(\mathbf{j}, \mathbf{e})$ is chosen and linked to **e**, forming internode (**e**, **b**). At this point the algorithm becomes very similar to that outlined in detail in Table 6-2, except that new candidate OTU's are chosen in the order of increasing $d_1(\mathbf{j}, \mathbf{e})$. As before, they are linked to that interval to which they are closest.

Using this algorithm and setting the assumed ancestor **e** as the null vector {0, 0, 0, 0, 0, 0, 0} and following Camin and Sokal's coding scheme, we obtain a tree that is similar, but not quite topologically equivalent, to that shown in Figure 6-9. It is equally parsimonious in terms of number of evolutionary steps. The HTU's generated during such a stepwise algorithm may not be globally optimal. Farris (personal communication) has described the following minor modification of his method for optimizing the HTU's for a tree with a fixed branching form (Farris, 1970). For each HTU that subtends two OTU's write down a vector whose elements are "character state sets" where each bracketed set is the range of character states for one character over the two subtended OTU's. Thus, referring to Table 6-1, notice that the HTU ancestral to OTU's **13** and **28** would possess, for the six characters, the following vector of character state sets: {[1,1], [0,0], [1,2], [2,3], [0,0], [0,1]}. HTU's that subtend either one or two other HTU's are characterized by a state set vector where each element or set is the intersection or overlap of the character state sets of the two subtended taxonomic units (thus, [0,1] and [1,3] yield [1,1]; or [0,2] and [1,3] yield [2,2]), or if the intersection is empty, i.e., if no state is held in common, the HTU's are characterized by the least range enclosing elements from both sets (thus [0,1] and [3,4] yield [1,3]). These rules are applied to successive HTU's descending the tree toward the root. A second procedure is to replace the character state sets by their intersections with the corresponding sets in the immediately ancestral HTU. Thus, for example, a state set of [0,1] with an ancestral state set of [1,2] is replaced by [1,1]. This second procedure is applied at two stages in the optimizing process. During the first, descending pass down the tree, after the state sets of each HTU have been computed, all immediately descendent HTU's are examined for state sets whose range is not zero (i.e., whose pair of states is not identical). Procedure 2 is then applied to these state sets. The same procedure is also applied in a second, ascending pass through the tree, from the root to the distal HTU's. Thus the HTU ancestral to OTU's **13** and **28** in group A of the Caminalcules is optimized to the following vector of character state sets: {[1,1], [0,0], [1,1], [2,2], [0,0], [0,0]}. Since the ranges of the state sets are 0, the character states of the HTU can be recorded as {1,0, 1,2, 0,0}.

When reversal of character states is permitted, as by Kluge and Farris (1969), the techniques of finding minimum length trees become somewhat more complicated. Ideally, character state trees should be worked out for the entire cladogram, which

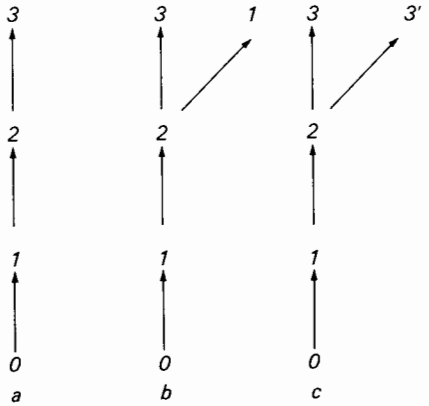


FIGURE 6-12
Recoding of character states. *a*, Original operational coding. *b*, Coded to represent presumed evolutionary sequence including reversed steps. *c*, Recoded to permit retention of the hypothesis of irreversibility. [Based on character *3* in the horse data of Camin and Sokal (1965).]

in effect means recoding characters so that reversed steps at a subsequent time are given a new character state (as suggested by Camin and Sokal, 1965). This is illustrated in Figure 6-12. If such character state trees cannot be constructed, then characters can be designated as primitive only in small local regions of the cladogram, as pointed out by Farris, Kluge, and Eckardt (1970). To carry out clustering of OTU's into cladograms under these circumstances they suggest a dissimilarity coefficient $U_{(\mathbf{a},\mathbf{b}),\mathbf{e}} = \frac{1}{2}(U_{\mathbf{a},\mathbf{e}} + U_{\mathbf{b},\mathbf{e}} - U_{\mathbf{a},\mathbf{b}})$, where values of U are in Manhattan distance, d_1. Note that the form of this coefficient is quite similar to the computation of the advancement index of the most recent joint ancestor of OTU's **a** and **b** except that here a local immediate ancestor **e** is assumed. The Manhattan distance should not be found for OTU's that are quite far apart in the network, for this would hide difference due to evolutionary reversals. Jardine, van Rijsbergen, and Jardine (1969) and van Rijsbergen (1970) have described clustering methods that will yield cladogenies if evolution rates are locally constant, as is assumed in much cladistic work.

In recent years there has been much progress in the construction of cladogenies from amino acid sequences in proteins. Early analyses of this type were carried out by Doolittle and Blombäck (1964) on fibrinopeptides and by Horne (1967) on erythrocyte catalase tryptic peptides. The classic paper in the field is that of Fitch and Margoliash (1967), about their work on amino acid sequences of cytochrome *c* for 20 OTU's ranging from yeast to man. Their phylogenetic (cladistic) reconstruction is based on the *minimal mutation distance* between two cytochromes. This distance is the sum of the *mutation values* for each pair of amino acids sequenced in the protein. A mutation value at a given site in the amino acid sequence is the *minimum number* of mutational changes in the nucleotides of a given codon necessary to produce the change from one amino acid to the other (the actual number may well be greater). Fitch and Margoliash furnish a table of mutation value for pairs of amino acids based on the known properties of the genetic code. Thus, for

example, a mutation from Cysteine to Lysine would require three mutational steps (mutation value = 3) since a UGU or UGC triplet would have to be replaced by an AAA or AAG triplet. By contrast, the mutation value from Aspartic acid to Glycine is 1, since a single nucleotide base substitution (for example, GAU to GGU or GAC to GGC) would suffice to bring about the change in the codon. Mutation distances by their nature are absolute values and can be quantified as $d_m(\mathbf{j}, \mathbf{k}) = n/(n - g)\,\Sigma_{i=1}^{n} m_{ijk}$ where m_{ijk} are the mutation distances at site i in a given protein of OTU's **j** and **k**. The mutation distances are adjusted proportionately to compensate for variable number of comparisons that involve gaps or deletions in the amino acid sequences. There are n sites, and comparisons are impossible (equivalent to NC in conventional numerical taxonomy) at g sites. Thus while the comparison in the study by Fitch and Margoliash was made over 110 amino acids, the mutation distances were multiplied by a proportionality factor of $110/x$ where $x = n - g$ is a number of positions ≤ 110 over which the comparison between any pair of OTU's was made.

Although Fitch and Margoliash (1967) provided a clustering algorithm for finding the evolutionary tree based on the mutation distances (basically single linkage clustering with the average of the as yet unclustered OTU's), these same authors in a more recent treatment (Fitch and Margoliash, 1970) deemphasize that algorithm and recommend any of several taxometric procedures for obtaining suitable dendrograms. Given several alternative trees, Fitch and Margoliash (1967) chose among them by an optimality criterion based on absolute differences of the mutation distances obtained in the original dissimilarity matrix from those implied by the cladogram. Each difference is divided by the observed mutation distance and multiplied by 100, i.e., it is expressed as a percentage. These percentages are treated as variates and the sample standard deviation of the $t(t - 1)/2$ variates is computed assuming a mean of zero. This standard deviation is employed as a measure of goodness of fit. There is no analytical or combinatorial solution to yield a minimum length tree, but the methods of Camin and Sokal (1965) as modified by Estabrook (1968) would clearly be applicable, the mutation values for any one site being equivalent to the number of evolutionary steps. There is close analytic similarity between the Fitch and Margoliash optimality criterion and the least squares fit measure of the additive tree model of Cavalli-Sforza and Edwards (1967). Recent phylogenetic work by Goodman and Moore (1971) and Goodman et al. (1971) has employed this criterion.

The purported cladogeny for cytochrome c of the 20 OTU's is shown in Figure 6-13. Note that it contains negative distances. This is because the optimizing procedure between the phenetic (mutation) distance matrix and the distances implied by the dendrogram (patristic differences, Farris, 1967a) forces the distances implied by the dendrogram into small negative distances that do not make biological sense but result from optimal fits. This is the phenomenon analogous to the reversals noted by Sokal and Michener (1958; and see Section 5.5).

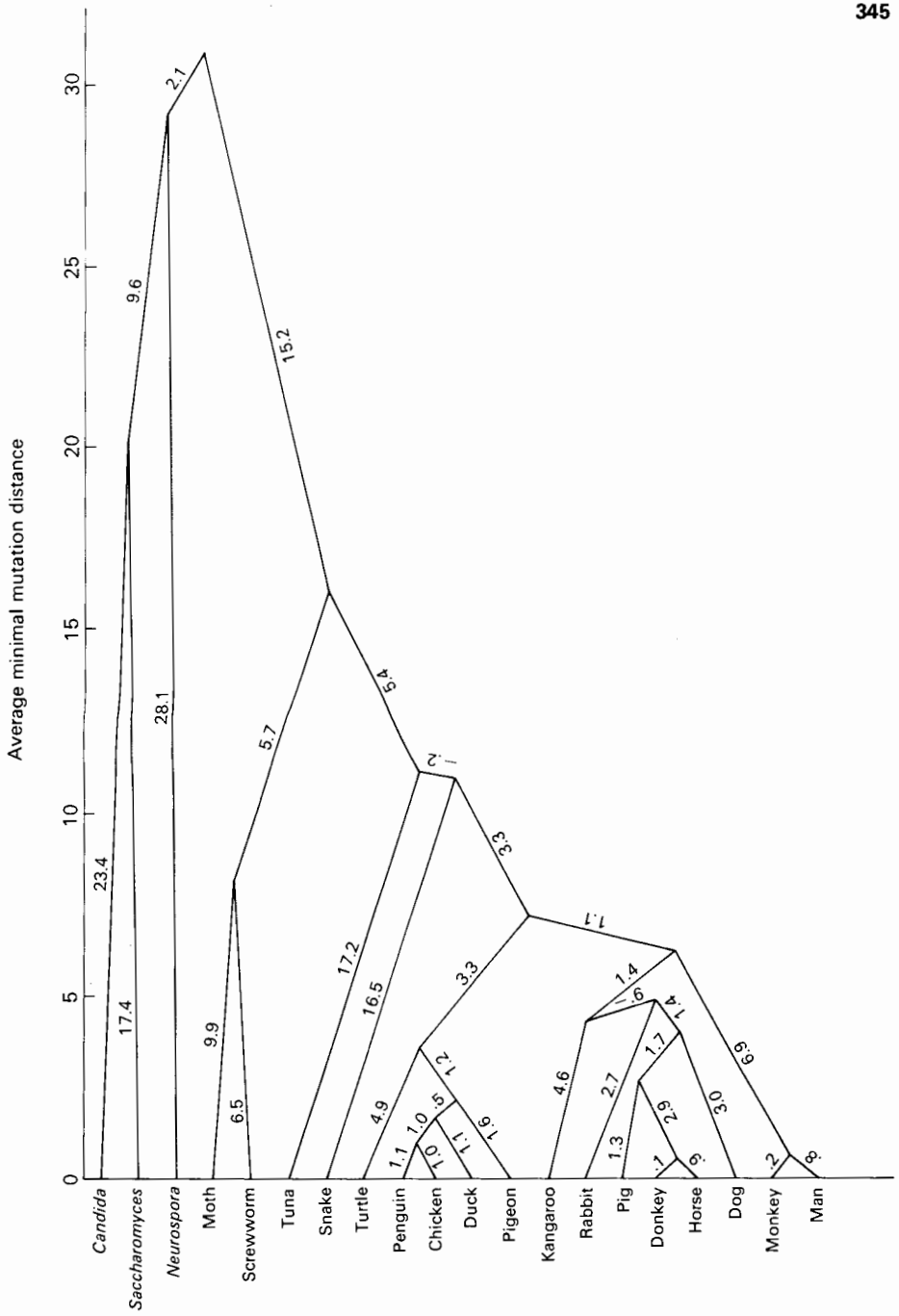


FIGURE 6-13

Phylogeny as reconstructed from observable mutations in the cytochrome *c* gene. Each number on the figure is the corrected mutation distance along the line of descent as determined from the best computer fit so far found. Each apex is placed at an ordinate value representing the average of the sums of all mutations in the lines of descent from the apex. [Modified from Fitch and Margoliash (1967). Copyright © 1967 by the American Association for the Advancement of Science.]

The method can be used not only for showing presumed cladogenies of proteins of different species, but also of genes. For example, Fitch and Margoliash (1967) publish a cladogeny of the ancestral gene for hemoglobin and its evolution into myoglobin and the α, β, γ, and δ chains of hemoglobin. In a later paper Fitch and Margoliash (1968) point out that the information they possess is basically phenetic information and that from that point of view their dendrograms are really phenograms. Only if certain assumptions are given about the evolutionary probability of given changes (or of the order and likelihood of certain nucleotide substitutions) can these dendrograms be turned into cladograms. The validity of their model was demonstrated on a simulated model in which the probability of mutation from one nucleotide to another, as observed in the first two positions of the codon, was employed. The authors caution about the difficulties inherent in double mutations (fixation of a mutation in two of its three nucleotides in the interval between two successive nodes).

The method of Dayhoff and Eck (1968) is similar in general approach, searching for a minimum length unrooted tree in terms of mutation distances between two proteins. A distinctive feature of their method is that when reconstructing ancestral sequences they mark as unknown any portion for which the amino acid is very doubtful, rather than forcing a decision on slender grounds. However, the validity of the reconstructed sequences is undermined by the dependence of the method on the exact set of OTU's, because of the majority voting principle used to decide on ancestral sites. They too analyze cytochrome *c*. Evolutionary sequences of other proteins are found in papers by Dayhoff and Eck (1969) for the globins; Dayhoff, Sochard, and McLaughlin (1969) for immunoglobulins; and Dayhoff and McLaughlin (1969) for other proteins, all in the volume edited by Dayhoff (1969a). The techniques for analyzing amino acid sequences and their nucleotide encodings are still in their infancy. Among the important aspects to be considered are the *frame shift mutations*, which involve deletion of a single nucleotide and insertion of a single nucleotide elsewhere in a sequence, making the entire message code for a new sequence of amino acids. Such a change would greatly alter the gene products over the region considered. The cladistic problem here is that the change in amino acids, when considered as normal mutational changes, would require a very large mutation distance, though when considered as a frame shift mutation it could be accomplished in only two mutational steps. Fitch (1970c) has estimated the probability that a specific frame shift mutation might be selected in the course of evolution. A very informative review of the entire subject of reconstructing evolutionary sequences on the basis of amino acid and nucleotide sequences is furnished by Fitch and Margoliash (1970). It is evident, however, that some aspects require critical examination; for example, the homologizing of bacterial and mammalian cytochrome *c* sequences cited in that article would lead to the conclusion that the horse is closer to one bacterium than that bacterium is to a second bacterium.

Character Weighting

In cladistic analysis the problem of character weighting assumes new importance. Since there is a "true" dendrogram of relationships—i.e., the true cladogram—any weighting scheme is desirable that improves our chances of obtaining a closer estimate to the true cladistic relationships. This has been recognized by orthodox phylogenetic taxonomists who have traditionally weighted some characters more heavily than others in arriving at putative phylogenetic classifications (and this practice has been widely abused by taxonomists who weighted some characters with very little justification by claiming that their classifications were phylogenetic). Weighting has also been employed by the more critical phylogenetic systematists such as Hennig (1966), by considering corroborating evidence from large numbers of compatible apomorphous character states and neglecting evidence from few discordant ones. We have already discussed this issue in Section 2.5. This point is also discussed in some detail as axiom A IV by Farris, Kluge, and Eckardt (1970).

The problem of weighting characters also brings up that of the numbers of characters to be used in a cladistic study. In this connection an important distinction between phenetic and cladistic work must be stressed. It is almost intuitively obvious that an adequate measure of overall phenetic similarity cannot be determined with the use of only a few characters. This is so, not only because by definition overall similarity cannot be represented by few characters, but also because there is no parametric measure of overall similarity and, as we have seen in Section 3.8, relative stability in character hyperspace is obtained only when the space is of high dimensions. By contrast it may be believed that if there is a single true cladogeny for any given set of OTU's, then in principle it would be possible to find just a single character that uniquely defines this cladogeny. This would be true only if all branches of the tree were defined by unique character states. It is unlikely that for any large number of OTU's there would be an equal number of unique character states that could be logically ordered. If there are fewer character states than OTU's because of parallelism in several branches, then it would be impossible to arrive at a cladogram on the basis of this character alone. However, it might be feasible to define a cladogram uniquely on the basis of a few divergent characters. It is for this reason that weighting of characters (in favor of those showing derived similarity) is suggested for cladistic analysis by many authors.

Several approaches to character weighting have been suggested in numerical cladistics. Camin and Sokal (1965) developed the method of *compatibility matrices*. For any given set of OTU's and character states they draw for each character the most parsimonious character state tree covering all OTU's. They then fit all other characters in turn to the *pattern* defined by this character state tree and compute the number of extra steps needed to fit each character to the pattern. By extra steps

is meant any number greater than $m_i - 1$, where m_i is the number of states for character i. This is the minimum number of evolutionary steps necessary to attain the terminal state for any given character. For each of the patterns we find out whether the character is *compatible*, i.e., needs no more than its minimum number of evolutionary steps, or requires extra steps (parallelisms) to fit to the pattern. If $C(h,i)$ is the minimum number of extra steps for character h on a pattern (character state tree) that fits character i exactly, a nonzero $C(h,i)$ indicates some inconsistency between h and i. From these computations emerges a compatibility matrix as shown in Table 6-4. The numbers of *compatibilities* (no extra steps required) for each pattern and for each character are given in the margins of the table, as are the numbers of extra steps. Patterns that have many compatibilities and few extra steps seem promising starting points for the reconstruction of a cladistic sequence, although as we have seen there are better methods for forming an initial topology.

TABLE 6-4
Compatibility matrix for seven selected Caminalcules (Group A)

Characters (h)	Patterns (i)							Compatibilities	Extra steps
	1	*2*	*3*	*4*	*5*	*6*	*7*		
1	×	2	2	2	2	1	1	0	10
2	1	×	1	2	0	1	0	2	5
3	2	2	×	4	1	2	1	0	12
4	2	3	4	×	3	3	3	0	18
5	1	1	1	3	×	1	1	0	8
6	0	0	0	0	0	×	0	6	0
7	0	0	0	0	0	0	×	6	0
Compatibilities:	2	2	2	2	3	1	2	14	—
Extra steps:	6	8	8	11	6	8	6	—	53

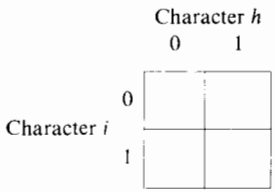
Values of $C(h, i)$ are given in the body of the matrix. Compatibilities in the margins are the frequencies of $C(h, i) = 0$ in a row or column.
Extra steps are $\sum_a C(h, i)$ for either rows or columns.

[From Camin and Sokal, 1965.]

However, those characters whose patterns require a large number of extra steps when the other characters are fitted to them, and which have few compatibilities with patterns made by other characters, are poor choices for the construction of minimum length trees. Characters may be incompatible because they have been incorrectly coded, and in fact the computation of a compatibility matrix may point out incorrect coding to the investigator. Such a case actually occurred in the coding of character 3 (average anteroposterior crown lengths of the fourth upper premolar and the second upper molar) in the fossil horses analyzed by Camin and Sokal (1965). Characters may also emerge as undesirable on the basis of compatibility

matrices if they have an unusually large number of parallelisms. The practice in the numerical cladistic method of Camin and Sokal has been to examine the compatibility matrix and to remove from it those characters that seem to be discordant with the majority of the others. This is why character *4*—clearly the most discordant in Table 6-4—was removed from consideration in the discussion of cladistic techniques (Tables 6-2 and 6-3). There is, of course, a danger in such a procedure. If most of the characters that are analyzed, which seem to be mutually compatible, are misleading, and only one or two characters incompatible with the others are the true indicators of the cladistic sequence, erroneous results are likely to follow. However, we know of no method that would reliably guard against this type of error. The philosophy of preponderance of evidence guides us here as in the more traditional approaches. Farris (1969b) has also pointed out that the relationship between the consistency of character h with the tree based on all characters and the number of extra steps for any given character h over all patterns as computed by $\sum_i C(h,i)$ is not a simple one, and one may not safely assume the cladistic reliability of character h from this value.

Le Quesne (1969) has shown that if the characters of a study are two-state, useful information on compatibility can be obtained by tallying the frequency of OTU's possessing character state combinations (0,0), (0,1), (1,0), and (1,1) for characters h and i in a customary 2×2 contingency table.



If all four cells of the contingency table are filled, the two characters cannot be compatible. If only three or fewer cells are filled, the two characters are compatible for this set of OTU's, and *might* be "uniquely derived characters," that is, both might have arisen only once in a phylogeny. He suggests removing all characters that are incompatible with others and using the remainder to construct a cladogram by assuming that the rarer state is derived. Le Quesne (1969) developed a *coefficient of character state randomness*, which can be defined as $\sum_i H(h,i)$, divided by the sum of the probabilities that all four states will occur, where $H(h,i) = 0$ when $C(h,i) = 0$, and $H(h,i) = 1$ when $C(h,i) > 0$. The lower this ratio the greater the proportion of presumed uniquely derived characters, and by implication the more reliable the cladogeny. He also extends this to characters of three ordered states. J. S. Farris (see Crovello, 1971) has shown that for binary characters Le Quesne's character-pair matrix is a compatibility matrix. M. J. Sackin (personal communication) points out to us that if there are n two-state characters with all pairs presumptively

uniquely derived, then there can be up to only $n + 1$ different OTU's. Furthermore, a rootless tree can be constructed in which every character is uniquely derived (i.e., each character mutates only once throughout the tree). If there are $n + 1$ OTU's then the tree is unique; if there are less, $n - x$ OTU's, then the tree is still unique, but $x + 1$ characters are nonindependent in the OTU's. A similar observation was made by J. S. Farris (see Crovello, 1971).

Farris (1969b) has shown that Le Quesne's method can be applied to multistate characters by recoding them by means of additive binary coding (see Section 4.8), except that there are m binary states for an m-valued character, not $m - 1$, and that the coding follows the steps in the character state tree, not necessarily the order of magnitude of the states (as when the tree branches, see Farris, Kluge, and Eckardt, 1970). One method of weighting proposed by Farris using Le Quesne's methods (and applied to two-state characters) is to use for each character h the weight $w_h = (n/N_h)^3 - 1$, where n is the total number of characters in the study, and N_h is the sum of $H(h,i)$ over the other $n - 1$ characters.

Another weighting criterion is the ratio s^2_W/s^2_B, where the variances describe variation of characters within and between OTU's. Other weighting coefficients proposed by Farris (1969b) relate to a statistic called by him *unit character consistency*, which is defined as $C_i = r_i/l_i$ for character i where r_i is the range of the character and l_i is the patristic unit character length. This is simply the sum of the lengths of the internodes for that particular character state tree. If the length of the internodes over all the OTU's of the study is no longer than the range of the character, then the character will be fully compatible and the unit character consistency $C_i = 1$. To obtain a weighting criterion Farris suggests $p_i = l_i/(t^* - 1)$ where t^* is the number of nodes (OTU's and HTU's) on the tree. Thus p_i is the average length of the character state tree for a given character i. Since in a two-state character $r_i = 1$ we can rewrite $p_i = 1/(t^* - 1)C_i$.

Farris (1969b) experimented with various weighting procedures based on the concepts defined above, using an artificial tree with 31 nodes and 30 consistent characters. From 5 to 150 poorly consistent characters had been assigned at random to the nodes. The data matrix (including the inconsistent characters) was at first weighted by a weight of $w_i = 1$. A shortest spanning tree was then constructed based on Manhattan distances and the number of evolutionary steps (patristic unit character length, network length) for each character was recomputed. This led to a computation of a weighting function $w_i = f(p_i)$, which was applied to the data matrix. A new shortest spanning tree was produced from the weighted data matrix, and this in turn yielded new parameters for the weighting function. The iterative procedure was terminated when the ancestor function of the tree remained unchanged on successive runs. Of various weighting functions employed the concave and unbounded function $w_i = p_i^{-3} - 1$ appeared most successful. It reconstructed the correct cladogram even when as many as 150 inconsistent characters had been added to the data matrix.

Finally, we need to consider a type of weighting introduced into the computations of cladogenies from protein sequences. Relying on knowledge of chemical structure and function, but mainly on the frequency of presumed mutations in the large number of proteins that have been sequenced, one can calculate the probability of a given point mutation from one amino acid to another. When such a study is made (see Dayhoff, Eck, and Park, 1969) great discrepancies in the likelihood of transition from one amino acid to another can be shown. In the reconstruction of ancestral sequences of proteins it seems reasonable therefore to take these probabilities into consideration as weights, and this has been done in some recent work as detailed by these authors. However, there is the slight danger of a logical trap. If we are to investigate the evolution of an as yet unstudied protein or taxon, the use of a matrix of probabilities or weights that effectively summarizes past history based on other proteins and organisms is likely to prejudice our conclusions in the direction of prior knowledge.

Numbers of Characters

We should briefly return to the question of the numbers of characters required for cladistic analyses, which were mentioned earlier in this section in connection with character weighting. This is an area where our knowledge is as yet rather scanty. It is clear that even if characters can be appropriately weighted there is a limit to the number of OTU's whose cladogeny can be established from n characters. As noted earlier, with two-state characters, at the most $n + 1$ OTU's can be arranged in a unique (but unrooted) tree, even if all of them are presumed to be uniquely derived using the methods of Le Quesne (1969), and Le Quesne observed that substantial fractions of characters had to be discarded because they were not uniquely derived. Hendrickson (1967) found empirically that it was necessary to use $2t$ to $3t$ characters to give most parsimonious cladograms of t OTU's if these cladograms were to be regarded with much confidence. It is not very satisfactory to have to choose between cladograms whose parsimony differs in only one or two evolutionary steps, and work is needed on the statistical significance of such differences if appropriate statistical models are to be developed.

Kidd and Cavalli-Sforza (1971) have made a first attempt in this direction by determining the probability of recovering the cladogram of correct topology for a set of four OTU's, using cladograms that were generated by a Monte Carlo technique. The OTU's branched from a common ancestor under conditions of random mutation at a constant rate. The authors showed the difficulty of obtaining the correct cladogram if the branch points were close together, and also demonstrated some differences in the efficiency of their different methods described earlier in this section. What is not so obvious from their presentation, but can be deduced with reasonable certainty by graphical integration of their findings, is that if all configurations are equally probable, then all the methods have much the same

performance. Furthermore, even when the number of characters is many times larger than t the chance of recovering the correct cladogram is not very high; e.g., with $t = 4$ and even 100 characters the chance of recovering the correct tree would be only about 70 percent for the rooted tree and a little over 90 percent for the unrooted tree. There are evidently severe limitations on the certainty of reconstructing cladogenies. Haigh (1970, 1971) has made a statistical study of the probability of determining the true root of a branching tree, which is a step in the desired direction (though perhaps his model is not entirely appropriate).

The methods described so far have dealt mostly with the problem of constructing estimated cladograms on the basis of character state information about recent OTU's. Estimating phylogenetic trees when character state information is not available, and when information about the OTU's is represented only in the form of a distance matrix between the OTU's, as for example, in immunological and DNA pairing studies, presents rather different technical problems, and has been little studied. The maximum likelihood inference model of Cavalli-Sforza and Edwards, mentioned above, formally depends only on a distance matrix between recent OTU's. This method, however, is not appropriate in general for phylogenetic studies, since it can be justified only by recourse to Cavalli-Sforza and Edwards's (1967) model of evolution by genetic drift. A promising new approach is suggested by Farris (1972).

Reticulate Evolution

The cladistic methods discussed so far were based upon the fundamental assumption that lineages may branch but never fuse. As we shall demonstrate, this assumption is quite critical. Fusion of lineages leads to what has been termed reticulate evolution, and introduces difficulties of an entirely different order of magnitude.

Is reticulation an important problem in most biological work? We may ignore sexual reproduction, which involves reticulation on a fine scale, and consider only reproductively isolated lineages with occasional hybridization. Reticulate evolution is usually considered rare in animals, at least in the better known groups (e.g., Mayr, 1963) although this may not be equally well founded in lesser known groups of invertebrates and in some fishes and amphibia (see Dobzhansky, 1970, p. 409 for some instances). In plants, however, hybridization is common between quite distinct species and even between genera, leading to persisting and evolving lineages. Hybridization usually occurs through allopolyploidy in sexually reproducing plants, and in some genera there are a large proportion of species that are presumptive allopolyploids (Davis and Heywood, 1963; Dobzhansky, 1970). There is also increasing evidence for reticulate evolution in microorganisms (e.g., Jones and Sneath, 1970). To the botanist and microbiologist, therefore, the question is not simply academic. The problems introduced by permitting fusion of lineages lie at two levels: first, the number of alternative cladogenies that must be considered

and the quantity of information needed to distinguish the various hypothetical cladograms; second, the type of information that is required for this purpose.

If fusion of lineages is permitted the number of possible cladogenies rises extremely steeply with the number of fusions. This is illustrated in Figure 6-14 where fusions are imposed on very simple branched cladogenies. Fusions can, in theory, occur between any segments of a branched cladogeny, and even with a single permitted fusion there are numerous possibilities. Furthermore, the direction of gene transfer is pertinent, at least if there is transfer of only part of the genome: a fusion from segment i to segment j generally yields a cladogeny different from the fusion from segment j to segment i. Even if one excludes a small number of possibilities on the grounds that they are trivial, that they yield cladograms identical with those from other fusions, or that they are forbidden by the time scale of the cladogeny, there remains nevertheless a great number of possible cladograms. For example, with a single fusion, there are almost n^2 combinations to add to a branched cladogram containing n segments. When there are two fusions we must consider a number of alternatives that is of the order of the square of this, i.e., n^4 (more accurately $n^2(n + 3)^2$, as new segments are erected by the first fusion). The number of possible alternatives thus becomes impracticably large. The number of characters required to distinguish between the cladograms must also be very large.

Information indicative of reticulate evolution is of the following type. A taxon **C** is discovered to have proteins some of which are identical (or almost so) to proteins of another taxon **A**; others are identical to proteins of a third taxon **B**. Evidence of this quality would provide extremely strong support for the hybrid origin of **C**, since in no other way could the properties of **C** be accounted for plausibly. Relative uniformity of individuals of **A** and **B** would suggest that both **A** and **B** are monophyletic. Evolutionary parsimony suggests the hybrid origin of **C**, since convergence in fine details of several proteins would seem very improbable. But the principle of parsimony is used in a new way: one searches for local similarity among the diversity *within* genomes. One of us (Sneath, 1971a) has suggested that this principle of "similarity among diversity" would have to be applied so that some numerical function of the separate evolutionary changes of the parts of the genomes was minimized. The appropriate methodology remains to be developed, but the need for it should be clear. The ultrametric property of proteins that evolve at sufficiently constant rates could be used to detect reticulate evolution, for if some proteins had a longer evolutionary pathway than others this could be detected (Figure 6-15).

One further point merits attention. It is possible that this principle could be applied to cases where only isolated morphological characters are available, instead of fine structure within sections of the genome or its direct products. In fact P. A. Wells (personal communication) in a cladistic study of hybrid species of *Arctostaphylos* has shown that the introgressed characters show up as parallelisms in a monophyletic cladogram. Yet it would be much more difficult to apply in such

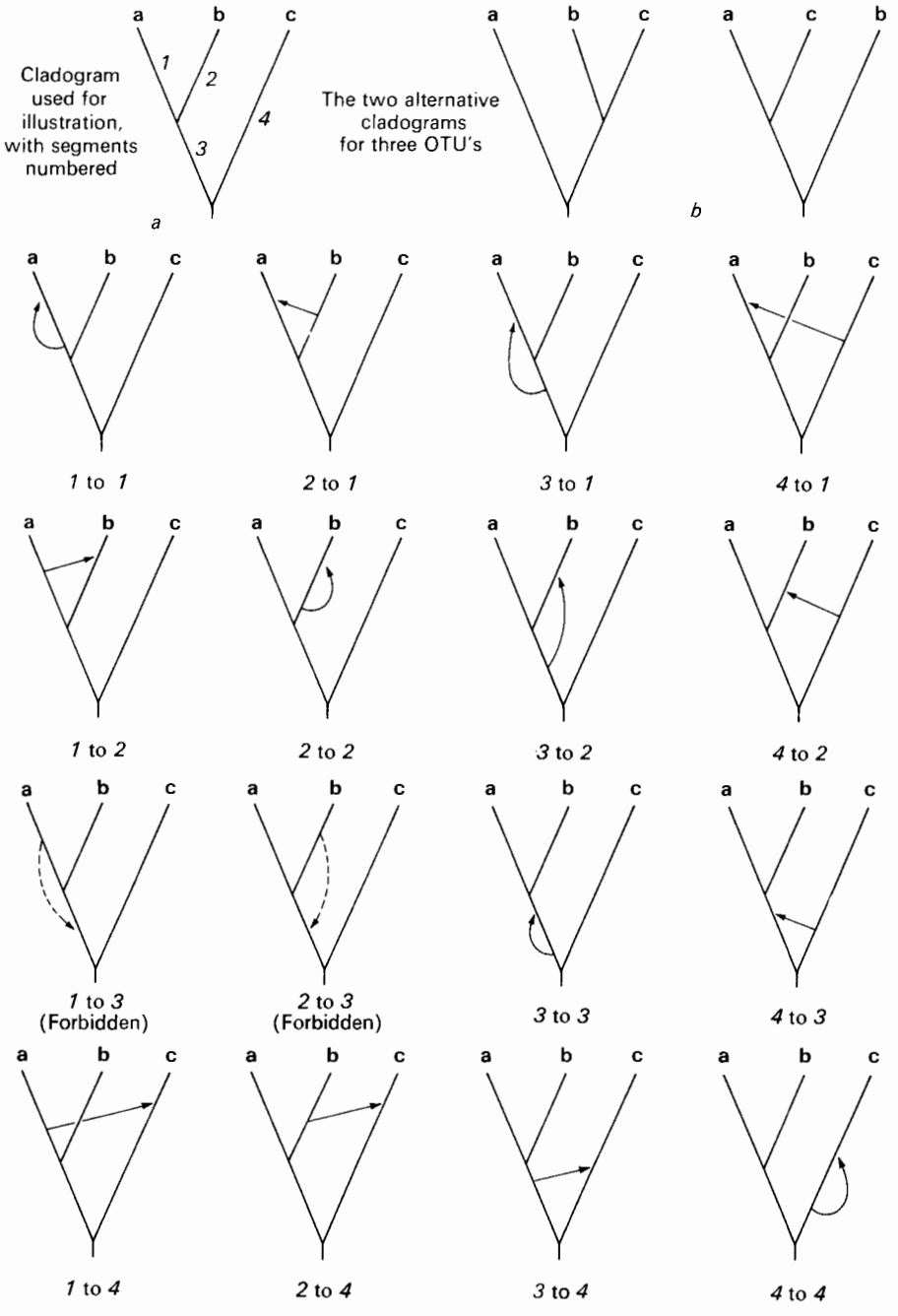
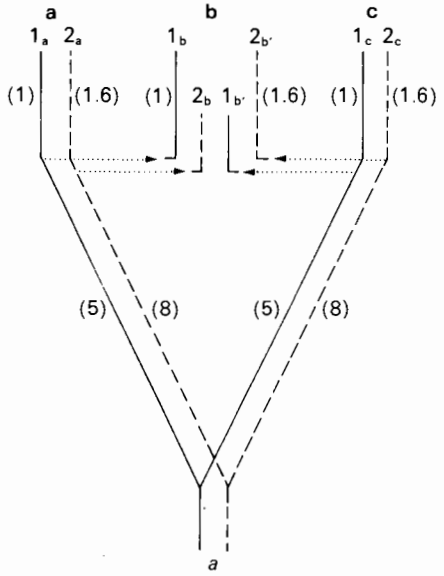


FIGURE 6-14

Alternative cladistic patterns for a single fusion during reticulate evolution in a cladogram of three OTU's. The cladogram at *a* is one of the three possible for OTU's **a**, **b**, and **c**, the other two being shown at *b*. The sixteen theoretical patterns of a single fusion for cladogram *a* are shown at *c*, the arrows showing the direction of transmission of genetic material. Two are forbidden, because they entail time reversal. It is assumed that the pathways contributing to fusion can exist separately for sufficient time for significant evolutionary change to take place (for example, in an isolated geographic area).



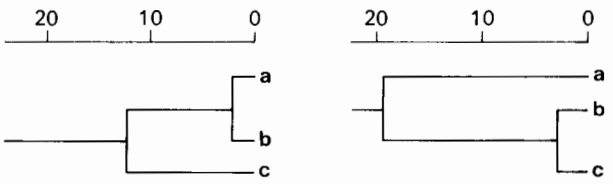
Dissimilarities between proteins

b Initial constitution of hybrid **b** = $1_b1_{b'}2_b2_{b'}$

	1_a	1_b	$1_{b'}$	1_c		2_a	2_b	$2_{b'}$	2_c
1_a	×				2_a	×			
1_b	0	×			2_b	0	×		
$1_{b'}$	10	10	×		$2_{b'}$	16	16	×	
1_c	10	10	0	×	2_c	16	16	0	×

c Later constitution of hybrid **b** = $1_b2_{b'}$, by loss of $1_{b'}$ and 2_b

	1_a	1_b	1_c		2_a	$2_{b'}$	2_c
1_a	×			2_a	×		
1_b	2	×		$2_{b'}$	19.2		
1_c	12	12	×	2_c	19.2	3.2	×



Phenogram from protein 1 Phenogram from protein 2

FIGURE 6-15

Incongruence between different proteins due to hybridization assuming constancy of evolution rates. *a*, Two organisms **a** and **c** have hybridized to form **b**, and two proteins, 1 and 2, are studied. Evolution is steadily divergent and constant in rate for each protein, although protein 2 is evolving 1.6 times as fast as protein 1. The rate of change, in dissimilarity units, is shown as figures in parentheses on the pathways. *b*, Initially the hybrid contains both genomes, and would thus contain two forms of each protein, 1_b and $1_{b'}$ and 2_b and $2_{b'}$, identical with the proteins in the parental organisms. The dissimilarities are shown for each protein. *c*, The hybrid loses one protein from each parent, and now has the constitution $1_b2_{b'}$; after a short period of further evolution it is examined and the dissimilarities found are those of the remaining ones, below which are shown the phenograms for each protein. These illustrate the incongruence.

cases, both because the number of available characters would often be too few to demonstrate statistical significance, and because much of the force of the argument rests on the fact that the similarities are between spatial and functional blocks of the genomes.

We have dwelt at some length upon the problem of reticulation for three reasons. First, as cladistic methods are developed and refined, workers will start to employ them in groups other than those where there are strong grounds for believing that reticulation is unimportant. Indeed, this is happening already in the field of human biology; it seems most unsafe to assume that racial and tribal lineages are without significant fusions (e.g., Fitch and Neel, 1969). Workers should therefore be aware of the new problems they may face. Second, it would be a mistake to commence building an extensive corpus of numerical cladistic work in biology upon unsound foundations. The whole subject of phylogeny has greatly suffered in the past for this reason. Third, the statement of the problems may lead to attempts to devise more appropriate numerical strategies to solve them, as has been briefly indicated above.

6.5 NUMERICAL TAXONOMY IN PALEONTOLOGY

In paleontological studies, the importance of exact, numerical methods is even greater than it is with living material. In extinct forms there can be no appeal to genetic analysis, the material may be scanty and incomplete, and unsuspected heterogeneities may complicate what at first sight seem to be single phyletic lines. Reviews of numerical taxonomy in paleontology are given by Rowell (1970) and Kaesler (1970b). Reyment (1963) has discussed the application of multivariate analysis to fossil material, and Van Valen (1969) has reviewed intraspecific variation in animal fossils.

The most obvious application of numerical taxonomy in paleontology is to fairly complete, well-preserved fossil material, in which the hypothesis of non-specificity is likely to hold well enough to obtain reasonably good estimates of overall phenetic similarity. Examples of such studies are those of Rowell (1967) on brachiopods, Kaesler (1969b) on ostracods, Cheetham (1968) on bryozoa, and Lange, Stenhouse, and Offler (1965) on conifers. Ordination has also been used (e.g., Reyment, 1965; Pitcher, 1966; Gould, 1967). Some studies have been primarily directed toward discriminant analysis of previously defined groups (e.g., Reyment and Naidin, 1962). Kesling and Sigler (1969) used Wagner networks (Section 6.4) to reconstruct the cladogeny of certain crinoid genera, and found good agreement with independent geological data.

Fossil material poses several special problems, many of them stemming from inadequacies in the primary data. Specimens may be rare, fragmentary, or not found in critical parts of the geologic record. They may have been sorted by environmental effects, or distorted (though transformation grids might prove useful

then—see Sneath, 1967a). In fossils of simple shape it may be difficult to obtain more than a few characters. It is especially unfortunate that many well documented lineages are based on material of this kind (e.g., ammonites, oysters, sea urchins). The application of the allometric equation where possible (Section 4.9), is important in paleontological work because one may have no way of knowing (or estimating) the age of the specimens at death, and the crude character sizes or ratios will sometimes be partly dependent on age and other factors. This has been well discussed by Joysey (1956) and in relation to numerical taxonomy by Fry (1964) and Anstey and Perry (1970). In many instances a fossil form is known from only a single specimen, which may represent a young or an old individual, and it may then be very difficult to know how to code its characters. The same problem, of course, occurs in orthodox taxonomic studies.

Problems of homology (Section 3.4) may be greater with fossils than with living forms because of the lack of supporting evidence from embryology or soft parts. Jardine (1969c) has given an illuminating account of this subject as illustrated by skulls of fossil fish. Benson (1967) has suggested a novel technique for studying shape that may be pertinent here. Nevertheless, most of the studies reported above appeared to be generally satisfactory. One may note, however (as Kaesler, 1967, has pointed out), that cluster analysis may not always be well adapted to analyzing the variation pattern of fossils if evolving sequences are present. Ordination methods (Section 5.6) may often be more appropriate and particularly graphs and trees (Section 5.7), because these are well fitted to finding branched sequences.

A special application of numerical taxonomy that, though it resembles ecological studies, is directed toward a taxonomic problem, is well illustrated by the paper of Kohut (1969). This is the use of clustering methods to group together isolated specimens (in this case conodonts) or fossil fragments that are thought to come from a single organism, but which have been scattered in the sediments. One hopes in this way to distinguish the assemblages derived from different species.

Numerical cladistics and phyletics have been reviewed in earlier sections of this chapter. It remains here to take up some points particularly pertinent in paleontology. The reconstruction of phyletic lineages is largely based on connecting the forms that are most similar phenetically into sequences showing an even progression in characters. Certain restrictions are imposed by the stratigraphy of the fossils, but this is often of little practical value; gaps in the fossil record may make it difficult to know whether a given form had already evolved at a given time, so it may be uncertain whether it might have been an ancestor or a descendant of some other form. The criteria for determining whether a lineage has branched require, in addition, methods of deciding upon the distinctness of the resulting lineages and, usually, some consideration of the principle of evolutionary parsimony (see Sections 6.3 and 6.4). Methods of ordination should be particularly useful in determining sequences in lineages. The techniques employed by archaeologists, including multidimensional scaling (see Cowgill, 1968 and Section 5.6) would also

repay closer study. Single-linkage clustering may occasionally be more useful than average-linkage methods because of its tendency to find chains of closely related OTU's.

Besides its importance for studying rates of evolution (Section 6.2), fossil material can contribute to the study of patterns of evolution. Speciation and branching is the most obvious, but there are other patterns that merit investigation. Some points may be seen in Figure 6-16, which is taken from a pilot study (Sneath, 1961) on the fossil fish *Knightia* based on data of Olson and Miller (1958). The four samples are thought to represent one phyletic line, and the individuals have been displayed in two major phenetic dimensions and in time. It may be noted that the comparison of individuals considered as geometric shapes is fairly straightforward. However, when one wishes to draw conclusions about the populations to which the fishes belonged, much more data are required, including allometric transformations to account for possible age differences in the individuals, and also evidence that only one population is represented in each sample. The study showed that the phenetic means did not show a regular displacement with time. This may represent some degree of reversal of evolution, or what Henningsmoen (1964) has described as zig-

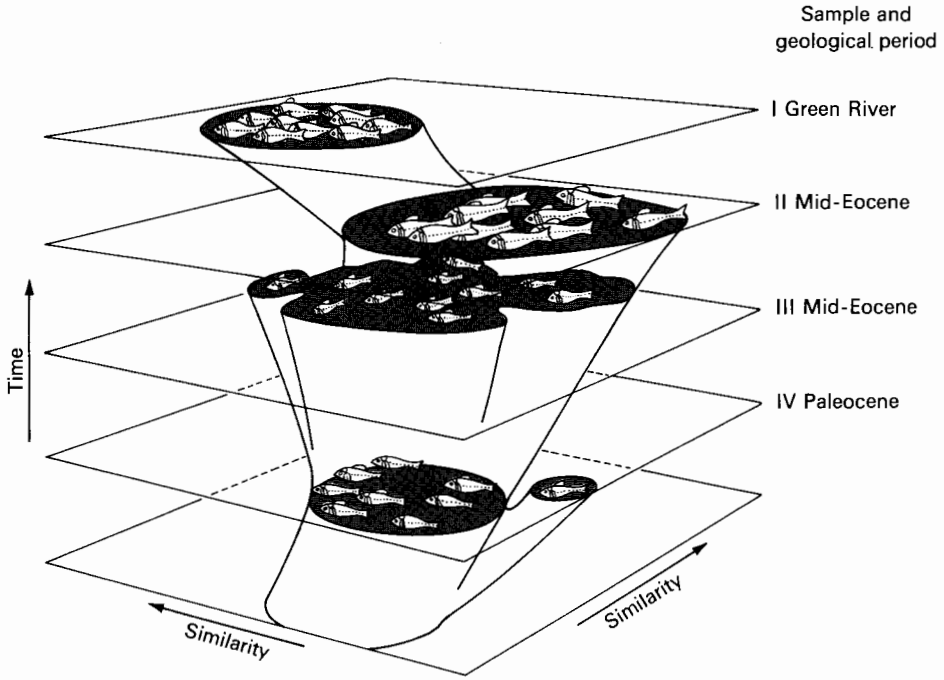


FIGURE 6-16

Schematic and speculative diagram of phylogeny in *Knightia*. The position in the horizontal plane indicates phenetic relationships among individual specimens. The time cuts are equally spaced although the actual time intervals are not equal. The axes labelled "similarity" approximate principal component factor axes. They should properly be labelled "dissimilarity". [From Sneath (1961).]

zag evolution. Alternatively, the ancestors of each group might have been an atypical section of their contemporaries, so that perhaps repeated burgeonings of forms well adapted to the prevailing conditions had occurred, rapidly dying out and contributing little to the succeeding part of the phyletic line.

A recent, especially thorough study of phenetic relationships in a time dimension has been carried out by Rowell (1970) on ten species of Upper Cambrian pterocephaliid trilobites conventionally assigned to three genera, whose stratigraphic position had been fairly well established by previous work. In addition to examining the problem of applying numerical taxonomy to paleontological work in great detail (this paper should be recommended reading to all paleontologists who wish to become acquainted with numerical taxonomy), Rowell elaborates a technique for ordinating the species in a three-dimensional space in which the two horizontal dimensions are the best phenetic representation in a 2-space and the vertical axis represents time (Figure 6-17). The OTU's are also connected by a shortest spanning

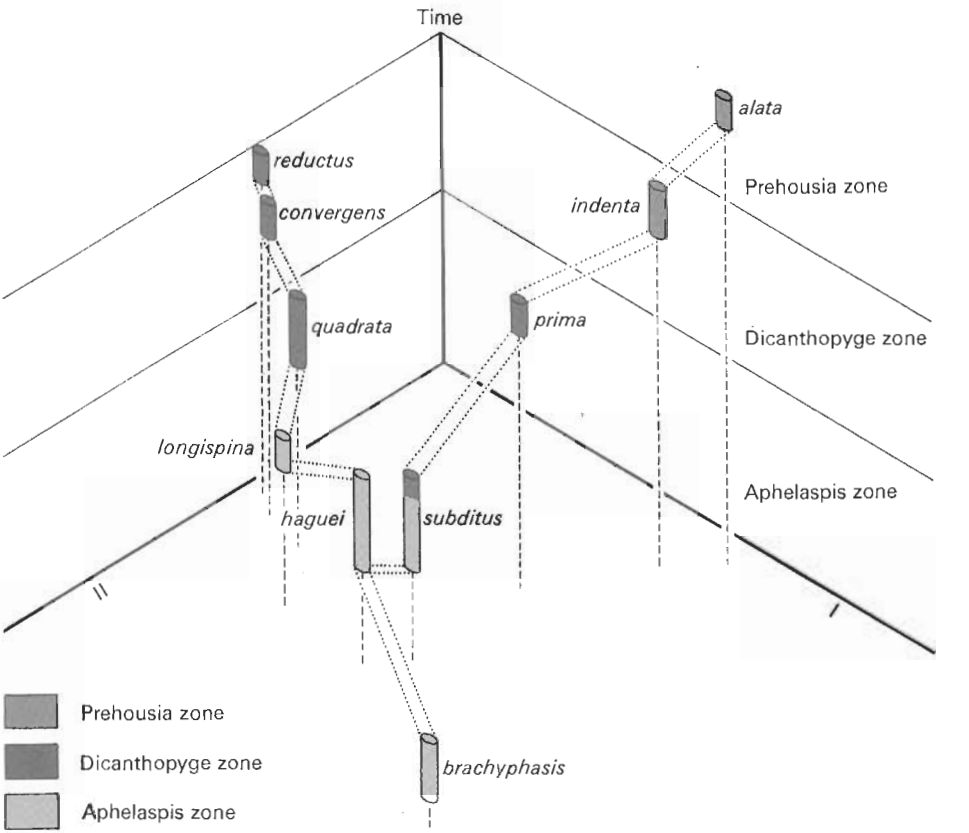


FIGURE 6-17
Inferred phylogenetic relationships between ten species of trilobites. The three axes represent two phenetic dimensions (principal axes) and approximate age. Dotted lines are the shortest spanning tree in the full *n*-space modified by some paleontological considerations. [From Rowell (1970).]

tree that shows the considerable congruence between the presumed phyletic relationships and phenetically closest OTU's. Rowell raises questions about the phyletic relationship of the earliest forms and reviews the evidence in favor of alternative hypotheses. From the figure shown in 6-17 Rowell develops a generalized diagram (Figure 6-18) that is in agreement with known phenetic and chronistic facts. He finds that such a model has high heuristic value, depicting the main features of the phylogenetic history of the group, readily displaying the regions where overall convergence has occurred, and providing information on relative rates of evolution. Although the model does not define the limits of generic taxa it will assist systematists in making such decisions.

Other evolutionary patterns that can only be properly studied by numerical methods are convergence and parallelism, as noted in Section 2.4. Eldredge (1968) has reported a case of convergence in fossil snails by using D^2, though rather few characters were employed. The investigation of mosaic evolution requires techniques for measuring congruence, as noted in Section 6.2. It is possible, too, that numerical methods can be applied to hypotheses of neoteny and macroevolution (de Beer, 1951).

Many of these points are illustrated in a paper by Cheetham (1968) on fossil bryozoa. The cladograms constructed from phenetic and stratigraphic data showed some evidence of zig-zag evolution, and there were also signs that a small degree

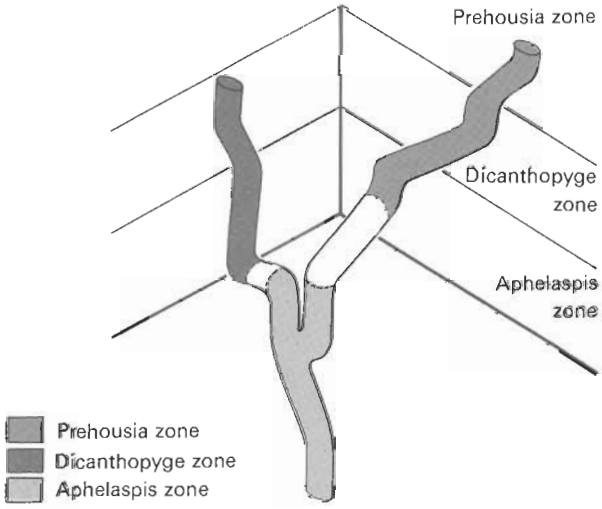


FIGURE 6-18
Inferred lineage of some Upper Cambrian trilobites based on ten species. Phenetic relationships are shown by projection on the first two principal components, relative age is proportional to distance along the third, vertical axis. Limits of conventionally recognized taxa of genetic rank shown by shading. [From Rowell (1970).]

of overall convergence, as well as marked parallel evolution of certain characters, had occurred in parts of the phylogeny. Also notable were changes in overall evolution rates; periods of slow evolution without branching (stasigenesis) alternated with bursts of speciation when lineages diverged rapidly (cladigenesis). Cheetham's conclusions are much firmer than those of most studies of this kind, because of their strong quantitative basis, although in some places he departs somewhat from the phenetic findings when they do not entirely fit the presumed cladistic relationships.

Stratigraphy is discussed in Section 11.4. However, the comparison of fossils in different strata would be based on numerical taxonomy of the fossils, even though the results of the taxonomy may be intended to enable geologists to identify the strata in different localities, across geological faults, and so on. Such work is akin to ecological studies in that many different kinds of organisms may be included as components of one OTU (in stratigraphy the OTU's would be strata instead of organisms). Such studies therefore incorporate a new element: strata are similar or dissimilar not only in the number of species of fossils that they share but also in the degree to which the fossils of one higher taxon are similar in the two strata. For example, pertinent evidence on the degree of similarity between the strata **a**, **b**, and **c** might be obtained from the degree of similarity between three species of a given genus, species **x**, **y**, and **z**, each characteristic, respectively, of strata **a**, **b**, and **c**. There is an increasing awareness (e.g., Shaw, 1969; Hazel, 1970; Hughes, 1971) that finer details of stratigraphy can be obtained by numerical taxonomic means such as this than by the traditional methods.