

CHAPTER 4

The Aims and Principles of Numerical Taxonomy

4.1. DEFINITION OF NUMERICAL TAXONOMY

Having shown what we (and many other critics) believe to be the faults of classificatory methods as currently practiced, we shall state in the present chapter the aims and principles of numerical taxonomy. This will be done in a relatively brief manner in order to provide the reader with a summary of our position. A detailed justification of our assumptions and techniques can be found in Chapters 5 through 9.

Before proceeding, it is necessary that we clearly define our use of the term "numerical taxonomy." We mean by it *the numerical evaluation of the affinity or similarity between taxonomic units and the ordering of these units into taxa on the basis of their affinities*. The term may include the drawing of phylogenetic inferences from the data by statistical or other mathematical methods to the extent to which this should prove possible. These methods will almost always involve the conversion of information about taxonomic entities into numerical quantities. We have preferred the term "numerical" to the term "quantitative," since the latter would include other methods (such as serology or paper chromatography) which we do not discuss in detail here but which have been briefly contrasted in Chapter 3. These latter methods, in a purely formal sense, bear a relation to numerical taxonomy similar to that of analog to digital computation.

We do not wish to widen the term "numerical taxonomy" to include every application of statistical or other numerical methods in systematic research. Our approach consists of a variety of multivariate techniques; but such techniques when not applied to problems of classification are not included in numerical taxonomy. Similarly, classificatory studies based on single or few characters are not included since they do not meet the conditions we feel necessary for a valid taxonomy.

The practice of numerical taxonomy involves a number of fundamental assumptions and philosophical attitudes toward taxonomic work, which we shall discuss and defend in detail in the sections that follow. We have already shown that none of the attitudes and assumptions is new. They, as well as isolated attempts at a numerical treatment of taxonomic relationships, date back over 200 years. However, we would prefer to limit numerical taxonomy to the integrated approach of recent years, which deliberately set out to revise taxonomic theory and practice.

4.2. THE AIMS OF NUMERICAL TAXONOMY

Numerical taxonomy would have no claim to the serious attention of biologists unless it could overcome some of the faults found in conventional taxonomic procedure. While we feel that the methods to be discussed below have a number of ancillary advantages, to which we shall draw attention at the appropriate places, the outstanding aims of numerical taxonomy are *repeatability* and *objectivity*.

Although we cannot expect scientists always to agree on interpretations of facts, it is the aim of scientific methodology to reach agreement on the facts themselves through the repeatability of observations. It is in this direction that numerical taxonomy aims. We hope by numerical methods to approach the goal where different scientists working independently will obtain accurate and identical estimates of the resemblance between two forms of organisms, given the same characters on which to base their judgment. Classification must be freed from the inevitable individual biases of the conventional practitioner of taxonomy.

Closely tied up with repeatability is the notion of objectivity. It would hardly seem necessary to stress that, like most rules of scientific methodology, objectivity is a relative concept, seldom fully realized. Yet misunderstandings have arisen on this score (Inger, 1958; Sokal, 1959). By including many characters without previous arbitrary selection or elimination, and by providing standard methods of processing the data

and evaluating the results, we reduce subjective bias and hence increase objectivity. Objective and repeatable procedures should together lead to very stable taxonomies, which are unlikely to be overthrown by later discoveries.

4.3. THE BASIC POSITIONS

Numerical taxonomy is based on the ideas first put forward by Adanson. They may be called Adansonian and are described concisely by the following axioms (modified from Sneath, 1958).

(1) The ideal taxonomy is that in which the taxa have the greatest content of information and which is based on as many characters as possible.

(2) A priori, every character is of equal weight in creating natural taxa.

(3) Overall similarity (or affinity) between any two entities is a function of the similarity of the many characters in which they are being compared.

(4) Distinct taxa can be constructed because of diverse character correlations in the groups under study.

(5) Taxonomy as conceived by us is therefore a strictly empirical science.

(6) Affinity is estimated independently of phylogenetic considerations.

Axioms (1) through (3) are discussed in Section 4.4, "The Estimation of Resemblance"; axiom (4) is treated in Section 4.5, "The Construction of Natural Taxa"; axioms (5) and (6) are dealt with in Section 4.6, "Empirical versus Phylogenetic Taxonomy."

4.4. THE ESTIMATION OF RESEMBLANCE

This is the most important and fundamental step in numerical taxonomy. It commences with the collection of information about characters in the taxonomic group to be studied. This information may already exist and merely require extraction from the literature, or it may have to be discovered entirely or partly *de novo*. In most cases both of these procedures will need to be applied. For the method to be reliable,

many characters are needed. At least sixty seem desirable, and less than forty should never be used. All kinds of characters are equally desirable: morphological, physiological, ethological, and sometimes even distributional ones. We must guard only against introducing bias into our choice of characters and against characters which are not an accurate expression of the properties of the organisms.

We assume characters to be equivalent since we believe that there are no special groups of genes related to single morphological regions but that a random sample of the genotype is best obtained by sampling many and various characters. The general occurrence of pleiotropism, as well as the fact that a given character is usually responsive to more than one locus, confirms us in our position.

From our assertion of the equal taxonomic value of every character (see Sections 5.2 and 5.3 for our definition of a character) it is only a small step to the Adansonian practice of equal weight for every character when using it to evaluate taxonomic relationships. This is a point in direct conflict with traditional taxonomic practice and over which much controversy has raged. We propose to discuss this issue in some detail in the next chapter and would like to mention here only that, granted the desirability of the separation of the measure of resemblance from a study of phylogeny, equal weighting is an almost self-evident logical consequence. We feel reassured in that at least three independent researchers working along somewhat different lines and from different assumptions have all reached the identical conclusion.

The actual computation of a measure of affinity can be done in a variety of ways (Chapter 6). Most methods result in coefficients of similarity ranging between unity and zero, the former for perfect agreement, the latter for none whatever. Except in unusual cases the calculations are likely to be rather tedious, and electronic computation will be needed for any but very minor studies.

The similarity coefficients are then tabulated in matrix form with one coefficient for every pair of taxonomic entities (see Appendix, Table A-6). If a symmetrical (mirror image) matrix is to be tabulated for t entities, a $t \times t$ matrix will result with unity in the principal diagonal (often represented by **X** in the diagonal; see Table A-14 in the Appendix). This matrix can be represented geometrically by points in a space (see Figure 6-2 in Chapter 6). A maximum of t dimensions is needed for a correct representation of the t points (taxonomic entities) in the space. The distances between the points are related to taxonomic distances.

4.5. THE CONSTRUCTION OF NATURAL TAXA

It is our belief (expounded in detail in Chapter 7) that when taxa are established on the basis of an adequate representation of characters, the resulting classification will be natural, in the sense of Gilmour.

Classification in numerical taxonomy is based on a matrix of resemblances, and it consists of various techniques designed to disclose and summarize the structure of the matrix. A rough, graphical representation of the structure can be obtained by differential shading of the elements of the matrix (see Figure 7-3). In this manner the structure of the assemblage of taxonomic entities becomes immediately apparent if they have previously been roughly grouped so that supposedly similar forms are near each other. If, as is methodologically preferable, the entities are placed in the matrix without predetermined order, visual grouping is not easily accomplished without rearrangement. The various computational methods for clustering will process the data equally efficiently whether they are ordered or not. Since computational methods simultaneously provide some numerical evaluation of the taxonomic relationships, they are to be preferred.

These numerical methods, familiar in psychometrics for many years, are collectively called cluster analysis (see Section 7.3.2). By way of a general description, they may be recognized as more or less automatic methods for establishing and defining clusters of mutually high similarity coefficients among the entities in the resemblance matrix. These clusters may be likened to hills and peaks on a topographic chart, and the criteria for establishing the clusters are analogous to the contour lines of such a map. Rigid criteria correspond to high elevation lines which surround isolated high peaks—for example, species groups in a matrix of resemblances between species. As the criteria become more relaxed the clusters grow and become interrelated in the same way that isolated peaks acquire broader bases and become connected to form mountain complexes and eventually chains, with progress from higher to lower level contour lines. It should be emphasized again that the clusters are based on phenetic resemblances only and have no necessary phyletic connotations.

Differences in methods of clustering refer mainly to definition of the cluster, whether by lowest or highest or average resemblance. Of particular importance are the principles upon which the relationships among the relatively higher categories are to be based. What weight is to be given to isolated single entities when their resemblance to a group

containing several other entities is to be evaluated? Are resemblances among the higher units to be based on averages of the resemblances of individual entities or are they to be computed in some other manner? Various approaches have been taken so far; their relative merits are discussed in Section 7.3.2.6. The differences among them are, however, largely matters of technical detail.

The important common aspect of all these methods is that they permit the delimitation of taxonomic groups in an entirely objective manner, providing that the coefficients of relationship have been properly based on many characters chosen without bias and correctly computed. This is best illustrated by using a diagram of relationships of the type shown in Figure 7-10. Horizontal transects at any level will include in a taxon all those entities branching off a single stem crossed by the transect. If this is the first (and finest) grouping, there will be no upper limit to the closeness of the relationships within a taxon, but the transect will constitute the lower limit of affinity. Thus objectively definable and—what is even more important—exactly comparable limits can be drawn for all taxonomic groups within a particular study. Similar transects at progressively lower levels of affinity will create taxa of higher and higher taxonomic rank.

The number and position of transects should follow some prearranged system. Clearly they will depend on the size of the matrix: too many transects would provide too fine a classification; too few would leave much structure unrevealed. The aims of the investigator, conventions in the particular group, and questions of convenience and esthetics would all affect the placing of the transects. While Mayr, Linsley, and Usinger (1953) would permit considerations of convenience to affect both the number and size of taxa of a given rank, we consider the number of taxa to be established at any rank a relatively unimportant and arbitrary detail. But once a transect has been established, the structure within a taxon (and hence the number of lower ranked taxa contained therein) depends entirely on the resemblance values of entities and stems and is not subject to the manipulations of the investigator. In other words, *the position and number of transects is arbitrary, but they must be straight and horizontal lines*. Thus within one study the criteria for one taxonomic level must be identical.

We view monotypic taxa or very numerous ones with equanimity. Their occurrence does not lead us, respectively, into lumping or splitting. In taking such a position we are motivated by an effort to supply taxa with some objective and definable criteria. While conventional system-

atics purports that its taxonomic ranks represent affinity levels as well as monophyletic groups, they are in reality dubious vehicles for either concept. Numerical taxonomy restricts its criteria to phenetic affinity only but attempts to apply its criteria with consistency. In erecting taxa, some criterion for intragroup cohesion has to be established. This will largely depend on the method of cluster formation (Section 7.3.2.).

Biologists who use the results of taxonomic research are much concerned with the stability of a classification. The stability of a scheme based on numerical taxonomy may be affected in two ways.

(1) More information (in the form of new characters) may accumulate. If the initial evaluation of resemblances has been based on an adequate sample of characters it is our contention that their (relative) values would change very little on the addition of further characters (Section 5.6).

(2) New taxonomic entities may be included in subsequent studies. Although this will not change resemblance values among the old entities, application of the previous criteria for levels and number of transects may result in new and different taxa. Agreement will have to be reached by practitioners of numerical taxonomy on whether the established system should be rearranged to suit the new results or whether the new data should be judged by standards of relationship already established—that is, whether the transects should be continued at the level at which they were drawn in the first study. The pros and cons of this issue are discussed in Section 7.8.

In establishing a series of nested categories, the question of naming them inevitably arises. What level in the hierarchy are we to call a subgenus, a genus, a family? Have these terms any significance on their own other than as indications of the relative levels of the transects? It is generally accepted in conventional systematics that genera (and other categories) in such diverse groups as insects, birds, and flowering plants do not represent taxa of equivalent affinity values. Is it possible for numerical taxonomy to set up such equivalent categories, although these would be based on entirely different groups of characters (discussion in Section 7.5)? It would appear preferable to employ a new series of terms for the hierarchic system established by numerical taxonomy, which would include in the terms a quantitative estimate of the affinity of the group. The term 80-phenon (or 60-phenon) is suggested to connote groups affiliated at levels no lower than 80% (or 60%) of the scale used

in the analysis (see Section 9.1.1). Though the use of unfamiliar terms has some disadvantages, it would obviate the use of the imprecise older terms with their semantic and emotional encumbrances.

Mention should be made of alternative ways of obtaining structure from a matrix of resemblances. Sokal has suggested the use of multiple factor analysis for matrices based on correlation coefficients (Sokal, 1958; Rohlf and Sokal, 1962). This method is also described in Section 7.3.3.

4.6. EMPIRICAL VERSUS PHYLOGENETIC TAXONOMY

It should by now be apparent that the taxonomic procedures proposed above are of a strictly empirical nature. As such they are related to procedures used by some typologists in the past. This, however, is not an automatic disqualification of our views (as Simpson, 1961, would imply), since there are certain typological viewpoints which are more defensible than the corresponding phylogenetic views (Sokal, 1962b). The fundamental test of the validity of empiricism in taxonomy must be whether it can be used as a consequential and consistent method for arranging organized nature. We hope to convince the reader that it can be so used and that it is the only reasonable approach.

A basic (and very controversial) attitude of the proponents of numerical taxonomy is the strict separation of phylogenetic speculation from taxonomic procedure. Taxonomic relationships between taxa are to be evaluated purely on the basis of the resemblances existing *now* in the material at hand. The relationships are thus *static* (Michener, 1957) or *phenetic*, as we now prefer to call them. They do not take into account the mode of origin of the resemblances found nor the rate at which resemblances may have increased or decreased in the past. This attitude is taken for two distinct reasons.

(1) Until and unless methods are developed for objectively assessing and quantifying the phylogenetic significance of character differences or affinities, the consideration of such information is incompatible with our stated aim of objectivity and repeatability for the taxonomic process.

(2) We have mentioned in Section 2.3 that we do not at the moment possess (nor can we currently conceive of) a classificatory scheme, graphic or otherwise, able simultaneously to yield information on degree of resemblance, descent, and rate of evolutionary progress. Any scheme attempting to combine these approaches must of necessity involve

compromises and inconsistencies, to which any observer of our present-day *systema naturae* must testify. We therefore treat these aspects of taxonomy separately, establishing the phenetic resemblances of taxonomic entities and basing our classification on this information alone. Any phylogenetic deductions are then made on the basis of the phenetic classification but are not involved in the classificatory process.

The separation of phenetic and phyletic considerations in taxonomy is a very drastic step, to which we have not come lightly. Unavoidably, misunderstandings with our colleagues are frequent. It is difficult for all of us to abandon patterns of thought acquired with our early training. Numerical taxonomy is accused of being anti-evolutionary and of dragging taxonomy back into its typological, pre-evolutionary period. The practitioners of the new methods are lumped with the few surviving true anti-evolutionists because the latter also propose classification of organisms on the basis of their visible characters and without phylogenetic considerations. As it happens, all the proponents of numerical taxonomy are evolutionary biologists in their own right. They are firmly convinced that phylogeny is responsible for the existence and structure of the natural system. They are criticizing not evolution or the study of phylogeny but speculation passed off as fact. We believe that numerical taxonomy offers new opportunities for measuring and studying evolution in quantitative terms, as discussed in Chapter 8.

Not only do we insist on the separation of phenetic from phyletic considerations in taxonomic procedure, but we also feel that only phenetic evidence can be used to establish a satisfactory classification. We hold this belief for two reasons.

(1) The available fossil record is so fragmentary that the phylogeny of the vast majority of taxa is unknown; taxonomy as a subject can scarcely be restricted to a few favored groups simply to satisfy the dogma that to be natural a classification must be phylogenetic.

(2) Even when fossil evidence is available, this evidence itself must first be interpreted in a strictly phenetic manner—with the exception that a time scale is given in addition, which may restrict certain choices of interpretation of the phylogeny—since the criteria for choosing the ancestral forms in a phylogeny are phenetic criteria and are based on the phenetic relationship between putative ancestor and descendant. Any attempt to decide the phylogeny on one set of characters, in particular those believed to be homologous (derived from a common

ancestor, by the common definition of the term), or to decide the lines of descent without resorting to phenetic criteria, or to decide a priori which characters are important or are reliable guides to the phylogeny, soon leads to a tangle of circular arguments from which there is no escape. Even Simpson (1961), who strongly supports a phylogenetically based taxonomy, is aware of and points out the *circulus vitiosus* of this procedure.

To ignore phylogenetic considerations while evaluating taxonomic relationships is not an easy mental task for a biologist of this day and age. For almost a century there has been an intimate conceptual association between taxonomic and phylogenetic reasoning, so that terms such as "specialized," "primitive," "homologous," and many others have assumed double meanings whose distinction is rarely attempted. We ourselves find difficulty in keeping apart phyletic and phenetic implications of the terms we use. We have, however, made every effort to do so in this book.

4.7. THE RECOGNITION OF PHYLETIC RELATIONSHIPS

Much of our discussion of phylogeny has been destructive criticism, unavoidably so since it is useless to build classifications upon phylogenetic speculations. Our present aim is to make classificatorily stable phenetic groups. Nevertheless, we believe that numerical taxonomy offers constructive suggestions for the study of evolution (Chapter 8). We believe that quantified measures of phenetic affinity between organisms of different periods will afford the phylogeneticist with objective information. Such information may serve several purposes. It may indicate which of a number of forms is most likely to have been the ancestor or the descendant of a given organism, thus assisting in the construction of phylogenetic trees. It may indicate the rates and direction of evolution, or it may assist in the solution of some stratigraphic problems.

Even when organisms from only one period of time are studied, their phenetic relations may be of phylogenetic interest. At the very least they will serve as a sounder basis for phylogenetic speculation than has hitherto been available. Much useful information may also be obtained for studies of speciation by comparing phenetic differences with genetic or geographical data.

4.8. IDENTIFICATION OF SPECIMENS

One of the tasks of the taxonomist revising a group is to set up a scheme for the simple and rapid identification of undetermined specimens by persons not necessarily familiar with the group in question. This is usually accomplished by the preparation of a taxonomic key (for a recent discussion see Metcalf, 1954). Such a key may be based as much as possible on the natural system of classification, so that following the various dichotomies in sequence one traces along the branches of the dendrogram from high-ranking taxa, through intermediate ones, to the ultimate entities of the system. Other keys are entirely artificial and utilitarian, their arrangement ensuring the simplest possible discrimination of the forms and having no relation to the taxonomic structure of the entities. For purposes of identification alone, both types of keys may be equally suitable. What aid, if any, can numerical taxonomy give in the construction of taxonomic keys?

All the various methods of estimating resemblances in numerical taxonomy abstract or summarize information based on many characters. They thus submerge the very data used by taxonomists in making keys. This does not mean, however, that practitioners of numerical taxonomy will be unable to construct keys. The basic information on the characters and their states in each taxonomic entity which is fed into computers is eminently well suited for the construction of keys. The various machines available in the average computation laboratory are able—better and faster than the most practiced taxonomist—to sift through the mass of information available.

4.9. NOMENCLATURE

It does not seem to us that numerical taxonomy is likely to cause very great changes in the present system of nomenclature (see Chapter 9). The increased differentiation in phenetic affinities will mean that finer differences in rank can be distinguished. To take advantage of this we have suggested the phenon system of nomenclature. This seems to us preferable to attempting on every occasion to force categories of rank upon the affinity scale. Such a terminology would, we think, assist the stability of the present nomenclature. However, if a numerical study showed considerable changes from the orthodox nomenclature, the author has no choice but to alter the classification and to make such

changes of name as are then required. We must do this if we are to provide other biologists with the benefits of improved classifications.

The present dependence of the system of nomenclature upon types may be somewhat changed by the advent of numerical taxonomy, for it will be in principle possible to determine the limits of taxonomic groups, which up to the present has been a matter of individual opinion. Numerical nomenclatures for intermediate forms, whether phenotypic or phyletic, may also prove of some use. New nomenclatural types may also be chosen from numerical taxonomies. All of these developments would be best coordinated with recent methods of handling data by modern data-processing machinery.