

## CHAPTER 9

# Problems of Nomenclature

We do not at present foresee that many changes in nomenclatural procedures will be brought about by numerical taxonomy in the near future. A few constructive proposals will be discussed below, together with suggestions of some of the lines along which nomenclature may develop. The student of numerical taxonomy may require a guide to the application of the present rules of nomenclature, and for his convenience we list in the bibliography the four International Codes of Nomenclature. For zoology a standard text of taxonomy has in the past been that of Schenk and McMasters (1956) or Mayr, Linsley, and Usinger (1953), and for botany Lawrence (1951) or Core (1955). In microbiology and mycology a concise guide is given in Appendix I in Ainsworth and Sneath (1962, pp. 454–463).

It is clear that one effect of numerical taxonomy on nomenclature will be through the taxonomic groups it creates. Obviously changes of rank will mean changes of names; simplification of the taxonomy will mean fewer names; increased detail, or the discovery of new taxa, will mean more names. These would be brought about by any method of taxonomic study that produced changes in established classifications. As Jahn (1962) points out, the use of computers in taxonomy will force some changes of nomenclature; the separation between the plant and animal kingdoms will be abandoned, with consequent alteration of many of the present Rules (especially those allowing homonymy between animals and plants).

## 9.1. NOMENCLATURAL CONSEQUENCES OF NUMERICAL TAXONOMY

### 9.1.1. The phenon nomenclature

One consequence of the application of numerical taxonomy may be the ease with which we are able to recognize small differences in rank. The traditional categories of rank, such as order, family, or genus, may not be numerous enough even when expanded by intermediate ranks produced by prefixing the terms sub-, super-, and infra-. Words such as "supersubfamily" would be ugly and prone to confusion. They could be avoided by citing the value which characterized the rank of the group in a numerical study, as described below.

The groups established by numerical taxonomy may, if desired, be equated with the usual rank categories such as genus, tribe, or family. However, these terms have evolutionary, nomenclatural, and other connotations which one may wish to avoid. We therefore prefer new expressions (Sneath and Sokal, 1962). We call the groups simply *phenons* and preface them with a number indicating the level of affinity at which they are formed. For example, an 80-phenon connotes a group affiliated at no lower than 80 on the similarity scale used in the analysis.

The terms are intended to be general ones to cover the groups produced by any form of cluster analysis or from any form of similarity coefficients. Their numerical values will vary with the coefficient, the type of cluster analysis and the sample of characters employed in the study. They are therefore comparable only within the limits of one analysis.

Phenons are groups which approach natural taxa more or less closely, and like the term taxon they can be of any hierarchic rank or of indeterminate rank. Since they are groups formed by numerical taxonomy, they are not synonymous with taxa; the term "taxon" is retained for its proper function, to indicate any sort of taxonomic group. The term phenon was employed in an entirely different sense by Camp and Gilly (1943). They used it for a division of a biological species which was divided by sterility barriers into phenotypically distinguishable segments. The word does not seem to have come into general usage, and since it appears appropriate, we have employed it in numerical taxonomy in a new sense.

An example of the delimitation of phenons can be seen in Figure 9-1.

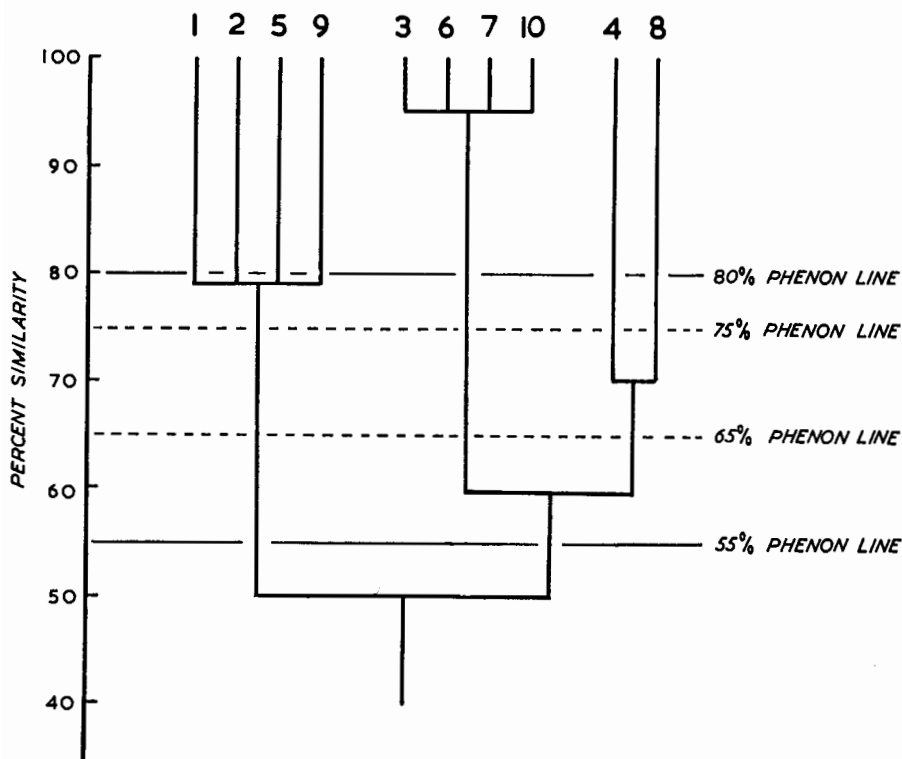


FIGURE 9-1

*Dendrogram to show the formation of phenons (for explanation, see text).*

Drawing a horizontal line across the dendrogram at a similarity value of 75% creates four 75-phenons, 1, 2, 5, 9; 3, 6, 7, 10; 4; and 8. We have found it convenient to refer to a given phenon by its first and last member. In the example above, we have 75-phenon (1 . . . 9). This is an informal system of nomenclature not based on types. The dots (ellipsis) do not indicate that the OTU's are necessarily in sequential, alphabetic, or numerical order, but merely that one or more are included between the terminal units. It is clear, of course, that such a label must refer to a given dendrogram and cannot easily be transferred to another study. If the original taxonomic units had been species, these phenons might be subgenera or genera. A second phenon line at 65% forms three 65-phenons. The advantage of phenons is that it is obvious that they are arbitrary and relative groups. This is not true of the Linnaean nomenclatural scheme. If some investigator felt that the taxa in Figure 9-1 should be divided into two instead of three groups, the phenon line

would have to be drawn at a similarity value between 50% and 60%; or he might feel that the two phenon lines were too close together and did not summarize the main relations very fairly, as a result of which he might draw the first line at the 80% level. The designation of the phenons would then change to 55-phenons or to 80-phenons, respectively; however, the relationships among taxa in the dendrogram are quite unchanged. Phenons are primarily suggested for use with dendrograms but with suitable prefixes could be used for phenetic clusters defined on other than dendritic scales.

### 9.1.2. Stability of nomenclature

Stability is a matter of some practical consequence. It may be argued, as has been done by Gilmour (1961b), that numerical taxonomy may increase instability, and it must be admitted that it may do so, at least during the first studies on a taxonomic group. We believe, however, that numerical taxonomies will in the end be very stable. It is clear that one could, by raising or lowering the phenon level a little, produce considerable changes in the nomenclature. This we believe to be undesirable; the present hierarchic nomenclature cannot profitably employ ranks that are too numerous or finely graded. We would not recommend the rigorous application of phenon lines if this severely disturbed the nomenclature without making any positive taxonomic contribution. For example, if a second study showed that the phenon level of the majority of subgenera of the first study now fell just below the line chosen to indicate genera, we would not rename them all on this account. A third study might well shift them again into the subgeneric level. However, changes which in the opinion of the taxonomist are major and significant should involve renaming. For example, when what had been previously thought to be two aberrant species of one genus are shown by numerical taxonomy to be two monotypic genera of a new subfamily, new names are clearly in order. This is necessary if the nomenclature is to reflect reasonably well the "natural" taxonomic groupings. To do otherwise is to deny biologists the benefits of improved taxonomies. Eventually one would hope that a time would come when name changes were no longer permitted for reasons of priority, and author citations would become unnecessary.

Rohlf (1962) has pointed out that in successive studies the least disturbance of nomenclature would occur if the dendrograms are divided at points where the stems show the widest gaps between successive

branchings. Sometimes these optimal levels would be easy to determine, but the temptation to make the rank lines wander up and down in their course across a single dendrogram would introduce an element of personal prejudice that is at variance with our hope for objective representation of the relationships.

## 9.2. NUMERICAL TAXONOMY AND NOMENCLATORIAL PROBLEMS

### 9.2.1. Limits of taxa

It is now generally recognized that modern nomenclature does not concern itself with the limits of taxa but only with reference points to the taxonomic names. What is to be included in a taxon is left to the decision of the taxonomist. Bradley (1939) expressed this as follows. "Nomenclature is concerned with the nuclei of groups, never with their limits. Taxonomy is concerned with the limits of groups, not their nuclei. The limits are debatable, subjective, forever changeable, not amenable to decision by authority. The nuclei can be fixed by common consent, for they are objective, utilitarian, permanent." Numerical taxonomy may here prove of value, for it will in principle be possible to delimit the boundaries of taxa by exact estimation of affinities, so that what organisms should be placed in a taxon will no longer be simply a matter of opinion. The limits then, as well as the nuclei, may also be objective, utilitarian, permanent, and fixed by common consent.

### 9.2.2. Terminology of intermediate forms

If numerical taxonomy reveals intermediate forms between taxa we can employ mathematical methods of the type of discriminant functions to show where it is best to draw the dividing line, and the nomenclature of the groups will follow this decision. The forms close to the dividing line may need a special terminology similar to that already used for hybrids and for intermediate forms in phylogenies, such as " $X.y - X.z$  intermediates," or " $X.y$  inter.  $X.z$ ." This might even take numerical form when an entity  $E$  is at a known taxonomic distance between two other entities, such as,  $20 X.y - 80 X.z$  for an entity which in taxonomic hyperspace is 20% of the way from  $X.y$  to  $X.z$ . To avoid confusion, the standard conventions for hybrids,

$$X. y \times X. z \quad \text{or} \quad \frac{X. y}{X. z}$$

should not be used. This simple situation occurs only when the intermediate form  $E$  lies on what may be envisaged as the direct line in taxonomic space between  $X. y$  and  $X. z$ . If it lies off this line, so to speak, the sum of the distances  $X. y$  to  $E$  and  $X. z$  to  $E$  will be greater than the distance  $X. y$  to  $X. z$ . An intermediate could then be a "50  $X. y$  - 70  $X. z$  intermediate," where  $X. y$  to  $X. z$  is 100, and  $X. y$  to  $E$  is 50, and  $X. z$  to  $E$  is 70 (see Figure 9-2). The

excess over 100 of the sum  $X. y$  to  $E$  and  $X. z$  to  $E$  (50 plus 70 in this instance) gives an idea of how far  $E$  deviates from the straight line joining  $X. y$  to  $X. z$  (it does not, of course, indicate in which direction  $E$  lies off the line). This raises the question of whether it should be considered an intermediate at all. Possibly  $X. y$ ,  $X. z$ , and  $E$  should be

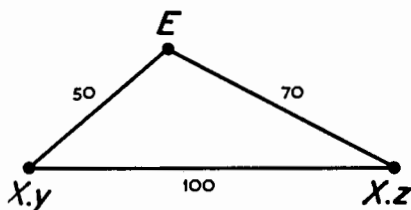


FIGURE 9-2

*Nomenclature of intermediate forms (for explanation, see text).*

combined into one taxon or  $E$  should be given independent existence as a taxon. A rule-of-thumb would be that the sum of the distances  $E$  to  $X. y$  and  $E$  to  $X. z$  should not be more than 1.5 times the distance  $X. y$  to  $X. z$ , if the convention is to be meaningful. It should also be noted that  $X. y$  and  $X. z$  may be represented either by their most central or most typical members or by their nomenclatural types (which may not be central or typical), and this must be made clear.

Similar occasions may occur in phylogenetic studies, and it may be possible to devise a terminology to suit this. One might, for example, record either the estimated cladistic or phenetic position of a form intermediate between two established and named points of a lineage, such as "40  $X. y$  - 60  $X. z$ ." While cladistic relations must total 100%, the phenetic relations need not do so. Phyletic relations would be conveniently distinguished by using the term "transient" instead of "intermediate" for the form between the named points.

A new development would be a nomenclature based on the volume occupied by a taxon in taxonomic hyperspace. Whether it would have advantages remains to be seen, but the principle would be to define a volume of a certain size as a generic volume, and so on, and to name the taxa within the corresponding volumes accordingly.

### 9.2.3. Choice of new nomenclatural types

Numerical taxonomy will sometimes be applied to groups where there is no significant earlier taxonomy, or it may cause extensive revision of an existing taxonomy. In such cases it may be necessary to set up types for the names of the new taxa. Such types need not be phenetically typical of the taxon. Their function is expressed better by the term "nomenifer," or name bearer, suggested by Schopf (1960) than by the term "type," implying typicality (Simpson, 1940, has suggested "onomatophore"). Nevertheless, there are advantages in choosing a nomenifer which is also reasonably typical, and the taxonomist can choose a typical specimen for a species from the results of numerical analysis. Similarly, a typical species can be chosen as the type of a genus or higher category. In general we require an OTU which is central in a geometrical sense in a cluster of OTU's in taxonomic space; however, there may be practical considerations indicating the adoption of a noncentral OTU as the type of a taxon.

One simple method is to list for each character the character state which is the commonest among the OTU's of the phenon in question. An OTU which possesses these states, or almost all of them, will be acceptably close to the geometric center of the cluster. The typical member of a cluster is that OTU having the highest loading on a factor representing the cluster, when a factor analysis of the Q-matrix has been carried out (see Rohlf and Sokal, 1962). Another method, suggested by Silvestri et al. (1962), is to employ the technique of Rogers and Tanimoto (1960), and for each OTU in turn to calculate the product of the similarity coefficients between that OTU and all others in a cluster. The OTU for which this product is greatest will be the most central (the centrotpe of Silvestri et al., 1962). Rogers and Tanimoto employ the  $-\log_2$  of the similarity coefficients and take the OTU for which the sum is smallest.

These methods presuppose that the cluster is reasonably homogeneous and is a representative sample of the group in nature. Of course, if a number of OTU's are equidistant from the geometric center, the choice among them is arbitrary. The central form thus selected bears a certain relation to the artificial types of Smirnov and to Heincke's "average individual" (see Sokal, 1962b, where some other methods for delimiting phenetic types are also discussed).

In summary we may say that, all other considerations being equal, we ought to employ type OTU's near a central point for a taxon. It may

be asked why one should go to all this trouble for a purely nomenclatural device; however, the computation of typical values can usually be easily obtained as by-products of the computer routines of numerical taxonomy.

#### 9.2.4. Data-processing machines

A new development in nomenclature is the use of card-sorting machines and computers to handle the "book-keeping" of taxonomic names, keys, bibliographies, and so on, discussed in detail by Gould (1958) and Jahn (1961). Although this is not part of numerical taxonomy as treated here, it is a parallel development which merits notice.

Jahn (1961) points out that classificatory schemes are being increasingly developed in many branches of science as an aid to efficient automatic processing of information. Such schemes when once established are difficult to alter; it therefore behooves taxonomists to see that these are of the sort they want, lest they find themselves faced with a *fait accompli*. In such applications, the use of "unnatural" taxa may have damaging consequences; once the schemes are compiled it is difficult to disentangle the information pertaining to the different entities which have been lumped together.

Several schemes for codifying names are now being developed. In taxonomy the most ambitious of these is the International Plant Index (IPIx) at the Connecticut Agricultural Experiment Station, New Haven, Conn. Its outlines are described by Gould (1962). The generic names of plants are being given code numbers, and bibliographic, biographic, and distributional data are being collected and coded. The Index hopes to issue in the near future mechanically prepared indices that are more comprehensive than any available hitherto. Other schemes have been suggested by Mullins and Nickerson (1951) and Denmark, Weems, and Taylor (1958). The Chemical-Biological Coordinating Center has developed an elaborate coding system (Seitner, 1960; Seitner, Livingston, and Williams, 1960).

These developments should be welcomed. The use of code numbers while highly convenient, may perhaps be modified by treating the name itself as a code number written in a twenty-six-valued notation (one value for each letter of the alphabet) to save the trouble of allocating code numbers and to permit automatic checking of homonyms.