9

Implications for Nomenclature

Although it is not yet possible to foresee the changes in nomenclatural procedures to be brought about by numerical taxonomy, some implications for nomenclature are evident now. A few constructive proposals will be discussed below and suggestions will be made about 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; for zoology, standard texts are Blackwelder (1967a) and Mayr (1969a) and for botany Lawrence (1951) or Core (1955). In microbiology and mycology a concise guide is given in Appendix I of Ainsworth and Sneath (1962, pp. 454–463). Savory (1962) has provided a useful discussion and comparison of the several International Codes of Nomenclature.

A new development in nomenclature is the use of punched cards and computers to handle the "book-keeping" of taxonomic names, keys, bibliographies, and so on. Although this is not part of numerical taxonomy as treated here, it is a parallel development that merits notice.

Jahn (1961) points out that schemes of classification and nomenclature are being increasingly developed in many branches of science as an aid to efficient automatic processing of information. Once established, such schemes are difficult to alter; therefore it 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 that have been lumped together. Jahn (1962) has also noted that such schemes may well force some changes also in the codes of nomenclature; the separation between plant and animal kingdoms may be abandoned, with consequent alteration of many of the present rules (especially those allowing homonymy between animals and plants).

9.1 SOME GENERAL CONSIDERATIONS

An excellent discussion of the problems of nomenclature is that of Simpson (1961, pp. 28–34). Systems of nomenclature have three major objectives: to provide for names that are (1) universally applied, (2) unambiguous, and (3) stable. Other considerations, such as that names should indicate taxonomic rank, position, or relationship, or that they should be descriptive are of secondary importance. The codes that regulate taxonomic nomenclature (but not necessarily other codes) also suppress unnecessary names (e.g., synonyms) by the application of the rules of priority and the nomenclatural type method. Since these different requirements are often in conflict, the codes offer an uneasy compromise; present nomenclature attempts to serve many functions but does none of them very well.

Numerical taxonomic techniques have implications in particular for stability, which is a matter of some practical consequence. It may be argued, as has been done by Gilmour (1961), 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. Instability may be of at least three kinds, referring to (a) OTU's to be included in a taxon, (b) the rank to be accorded to such a taxon, and (c) the name this taxon should be given. We believe, however, that numerical taxonomies will in the end prove 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; in common with others (e.g., Walters, 1965; Watson, Williams, and Lance, 1966) 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 that in the opinion of the taxonomist are major and significant should result in renaming. 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 for a time when the International Commissions on Nomenclature would no longer permit name changes 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 phenograms 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 let the rank lines wander up and down in their course across a single phenogram would introduce an element of subjectivity that is at variance with our hope for objective representation of the relationships. The phenon nomenclature described in Section 5.11 is suggested as a means of expressing finer details of taxonomic relationships without having to force them into a rigid and formal system of nomenclature.

A number of new proposals for nomenclature have been made in recent years, even to the extent of questioning the value of any rules at all (Oldroyd, 1966; Cowan, 1970). Some of these, such as proposals for new starting dates based on recent monographs (Howden, Evans, and Wilson, 1968) can be accommodated within the framework of the present codes, and may become feasible as advances in data processing make available complete and annotated lists of names that will lighten much of the present labor of taxonomists. Bacteriologists are already moving in this direction (Lessel, 1971). Other proposals, such as systems of virus nomenclature, are so radical that they involve intense debate (e.g., Gibbs et al., 1966).

Many of the new proposals revive the question of the extent to which names of taxa should be indicators of taxonomic relations rather than be simply labels for taxonomic groups. The former was an original intention of Linnean binominal nomenclature, though now it plays a minor function because most genus and species names are familiar to only a few specialists. Attempts to continue using names to indicate relationships bring disadvantages. They may cause, in particular, instability of names due to later taxonomic revisions. Biologists have given up the attempt to make names descriptive or to make nomenclatural types typical in the usual sense; it is therefore not surprising that there is increasing discussion about whether biological names can successfully serve any other function than as labels for taxa. It is not so much the indication of taxonomic rank that causes difficulty; the use of uniform endings for names of each rank category above genus is becoming increasingly fashionable, even though it leads to some changes of familiar names. Rather, problems are engendered by attempts to make a name carry an indication of the position of the taxon in classification, e.g., the class, order, and family to which it belongs.

The trend towards names as pure labels rather than as vehicles for taxonomic information is seen in proposals for uninominal nomenclature put forward by Cain (1959b) and Michener (1963, 1964). The genus and species name (taken from the most recent revision) would be hyphenated to make a uninomen that would thereafter never be broken up. On transfer to a new genus it would remain unchanged. A genus would then contain species whose names were entirely different; there would be no common generic part. Genera and higher taxa would receive the name of one of the contained species in Michener's scheme. Thus the Hymenoptera might be called Order *Apis-mellifera*, and would contain Family *Apis-mellifera* (for the Family Apidae) and Genus *Apis-mellifera* containing the species *Apis-mellifera.* The inconvenience of this plan could be ameliorated by adding uniform endings for each rank, and the resulting nomenclature would then be very similar to the present one if it were the custom to give names to no taxon below the level of genus. It would indeed offer numerous attractions if one required a nomenclature for entirely new groups of organisms that had no existing names hallowed by long usage, e.g., viruses, or as Michener mentions, creatures on another planet. It may be noted, however, that the functions (though not the form) of the names in such a system are essentially those given in current nomenclature by the original name (basionym).

Proposals in the other direction, to add information on the taxonomic position of a taxon, have a long history. Mayr (1969a, p. 345) discusses some early suggestions for supplementing the generic name with letters indicating the higher taxa. Thus *Papilio* would become *Ylpapilia* (*Y* for Insecta, *l* for Lepidoptera and *-a* for Invertebrata). Amadon (1966) has put forward an idea intended to strengthen the indication of taxonomic position without causing much instability: the rank of the category that receives the first name in a binominal would be raised from the genus to the family level (or thereabouts), and within each family no two species would have the same specific epithet. Taxonomic revisions resulting in change of family would be quite infrequent. The usual custom of adding the class and order in parentheses after first mention of a name is an informal scheme of a similar kind, and it would seem that no one has made the suggestion that it should be formalized and made obligatory, so that the higher taxon names would become part of the name of the species.

There have also been proposals to supplement taxonomic names and perhaps eventually supplant them by a system of numbers—a "numericlature" (Little. 1964). The reason for using numbers is twofold, to make the system readily handled by data-processing machinery, and to avoid difficulties with names owing to their associations, pronouncability, and so on. But it should be mentioned (as Hull. 1968b; Randal and Scott, 1967, and others have pointed out) that computers can handle names as easily as numbers. Also, names are easier to remember and to check for accuracy by eye. It is the ease with which numbers can serve as labels, as indicators of taxonomic rank or position, or even descriptors, that makes them attractive. For such purposes the numbers must have a system of rules, what Hull refers to as a syntax, and which in due course taxonomists will have to learn. It seems unlikely though that these rules will ever be as complicated as the present rules of nomenclature!

Suggested schemes of numericlature have been numerous. A pioneering paper is that of Gould (1958), who was one of the first to appreciate the potential of data-processing in taxonomy. He proposed that taxa should have a number to indicate taxonomic position that we may refer to as the *classification number*. Others (e.g., Michener, 1963; Little, 1964; Rivas, 1965) suggested adding a unique number (the *reference number*) as a label, and Hull (1966) proposes in addition a number indicating cladistic position (if known).

Stability is achieved by the reference number, which would never be changed, although union or division of taxa would require cross-references, and possibly supplementary numbers, to clarify the situation. The setting up of a system of reference numbers is simply a conversion of existing basionyms, because a basionym with its full citation is effectively a unique label in alphabetic characters.

The classification number would contain groups of digits for phylum, class, order, family, genus, and species. Taxonomists would soon become familiar with the main outlines of the system, as well as with the numbers for the taxa they specialized upon. Change in taxonomic position would involve change of the classification number only. Again suitable cross-references would be needed from time to time. Provision can be made for uncertainty of taxonomic position. Systems of numericlature can be applied at all taxonomic ranks, though they are often principally intended for species. The numbers would of course have to be international, and allocated from a central source. In a comprehensive system it would be easy to prevent homonyms and to identify many synonyms.

Finally, Jahn (1961) has suggested that modern data-processing equipment could allocate new names to newly discovered organisms, and we have noted (Sokal and Sneath, 1966) that the completely automatic renaming of taxa would be feasible.

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, Connecticut. Its outlines are described by Gould (1962). A comprehensive review of recent projects is provided by Crovello and MacDonald (1970).

The danger that the advent of electronic data processing may rigidify classification through its effects on nomenclature does not seem too serious. Any new nomenclatural system developed today should, of course, allow for automated information storage and retrieval, but taxonomic flexibility can also be provided. An important consideration of a nomenclatural system is the question of how information about the organisms is to be retrieved. Among the problems of document retrieval being discussed currently is the following: what is the optimum system of classification for a series of documents so that a document can be retrieved with a minimum of searches? Documents must be indexed under those headings that will most frequently be employed, and storage should be arranged so that access to the more frequently required documents is easier than to those less often needed. These problems have clear relevance to taxonomy. The cross-indexing of taxa and their more salient properties is desirable. Research in this field is urgently needed.

9.2 NUMERICAL TAXONOMY AND NOMENCLATURAL PROBLEMS

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 will change this position, for it will be possible to determine the centers and boundaries of taxa by exact estimation of resemblances, 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.

Numerical taxonomy will sometimes be applied to groups in which 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, or lectotypes or neotypes for old ones. 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. Nevertheless, there are advantages in choosing a nomenifer that is also reasonably typical, and the taxonomist can choose a typical specimen for a species from the results of numerical classification. Similarly, a typical species can be chosen as the type of a genus or higher ranked taxon. In general we require an OTU that is central in a geometrical sense in a cluster of OTU's in A-space; however, there may sometimes be practical considerations indicating the adoption of a noncentral OTU as the type of a taxon. Measures of the center of taxa have been discussed in Section 5.2.

The OTU nearest to the centroid (usually the centrotype) may be chosen as the most typical OTU, and in microbiology the organism closest to the hypothetical median organism is commonly selected. Of course, if a number of OTU's are equidistant from the geometric center the choice is arbitrary. Similarly, boundaries of taxa can be set by techniques discussed in Sections 5.2 and 5.11, and Niemalä and Gyllenberg (1968) have even suggested that several types per taxon could serve the same purpose. Descriptions and diagnoses, with precise indications of character variation, could also be produced automatically by numerical techniques.

There are a number of potential applications of numerical taxonomy to variation patterns not envisaged by orthodox nomenclature. One of these is terminology for intermediate forms. They may need a special terminology similar to that already used for hybrids and for intermediate forms in phylogenies, such as "**X** – **Y** intermediates," or "**X** inter. **Y**." This might even take numerical form. The intermediate form **I**(**X**, **Y**) could lie on what may, as a manner of speaking, be envisaged

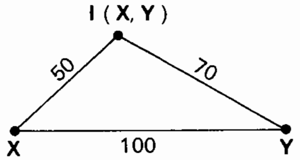

FIGURE 9-1
Nomenclature of intermediate forms. **X** and **Y** are taxa, **I(X, Y)** is the intermediate. Numbers along sides of triangle indicate distances between taxa. For further explanation, see text.

as the direct line in some appropriate taxonomic space between taxa **X** and **Y**. If it lies off this line, the sum of the distances $d_{\mathbf{X},\mathbf{I}(\mathbf{X},\mathbf{Y})}$ and $d_{\mathbf{Y},\mathbf{I}(\mathbf{X},\mathbf{Y})}$ will be greater than the distance $d_{\mathbf{X},\mathbf{Y}}$. An intermediate could then be a "50 **X** – 70 **Y** intermediate," where $d_{\mathbf{X},\mathbf{Y}} = 100$, $d_{\mathbf{X},\mathbf{I}(\mathbf{X},\mathbf{Y})} = 50$, and $d_{\mathbf{Y},\mathbf{I}(\mathbf{X},\mathbf{Y})} = 70$, as shown in Figure 9-1. The difference $|d_{\mathbf{X},\mathbf{Y}} - (d_{\mathbf{X},\mathbf{I}(\mathbf{X},\mathbf{Y})} + d_{\mathbf{Y},\mathbf{I}(\mathbf{X},\mathbf{Y})})|$ gives an idea of how far **I(X, Y)** deviates from the straight line joining **X** and **Y**—an indication of epistasis or overdominance—regardless of the direction of the deviation. It should be noted that **X** and **Y** 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 we have already given methods for the construction of HTU's and their location in A-space.

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. This would lead directly to a nomenclature or numericlature based upon the coordinates of taxa in a suitable phenetic space by giving each taxon a measure of location and a measure of dispersion. It would probably be applied to ordinations to make the numericlature more concise, although some simple arithmetic would be needed when using it with actual specimens. Du Praw (1964, 1965a,b) has given some examples on these lines.

Should overlapping taxa prove useful, present nomenclature, which is by nature hierarchic, is not well suited to these (see Michener, 1963). Uninominals and numericlature would be easy to apply however, as a species could be listed in two genera and readily cross-referenced by code numbers. Hierarchic nomenclature is also ill adapted to some other variation patterns (see Section 5.14), particularly the pattern in which dense clusters of OTU's are embedded in a sparse scattering of single OTU's. It is likely that this pattern is not uncommon at low taxonomic ranks in apomictic groups, where the OTU's are individuals or clones, but nomenclatural treatment of apomicts is not at all uniform (see Davis and Heywood, 1963). If the clusters are named as species (in this case taxospecies) how should the scattered OTU's be named? Would each one be named as a species on the grounds that a little search would surely turn up a tight cluster around it? This would often be quite impracticable because of their number, and these aberrant OTU's might indeed be unique. If one raised the species level to include all the OTU's then the clusters could be named as subspecies, which might be the simplest solution. But any course adopted might place considerable strain on Linnean nomenclature, partly because the variation could not be well represented as a hierarchy, and partly because of different definitions of the species category.

The requirement that every organism (with few exceptions) must have a binominal name is also a disadvantage in naming apomicts; it would often be more convenient to refer some forms simply to a genus or section of a genus rather than force them into the nearest species. This practice is finding increasingly favor in botany and microbiology (see Hill, 1959; Davis and Heywood, 1963). Whether similar problems occur with variation patterns at higher ranks is not yet clear, but the increasing number of phenetic studies will improve the knowledge of variation patterns at all levels and enable taxonomists to develop more appropriate systems of names or numbers.