

CHAPTER 5

Taxonomic Evidence: Characters and Taxa

We now proceed to a detailed discussion of the data necessary for obtaining estimates of affinity between taxa. This leads us first into a consideration of the nature of taxonomic characters, to a discussion of problems in character evaluation, such as convergence and weighting, and eventually to an account of the fundamental taxonomic units to be employed in numerical taxonomy.

5.1. PRELIMINARY CHOICE OF SPECIMENS

There is no sharp distinction between the selection of specimens and the selection of characters, since these generally proceed *pari passu*. Both of these are discussed in some detail here and in Section 5.4. Nevertheless, one must first make a preliminary selection of specimens in order to restrict one's study to a manageable taxonomic group. This group is selected by "classification from above," and selection is therefore necessarily based on rather few characters.

A point of some importance is to guard against the exclusion of pertinent material because it does not strictly fulfill all the criteria for the working definition of the taxon to be studied. The danger is that aberrant forms or descriptions of aberrant taxa in the literature may be excluded because they do not possess some character of the taxon which the systematist considers essential or diagnostic. Gross mistakes are unlikely: penguins will not be excluded from a study on birds because they cannot fly, nor will bats be included because they do fly. Yet less extreme cases may well occur, especially in poorly known taxa. We have

seen from the discussion of natural classification in Section 2.2 that it is not possible to be certain that all members of a taxon will possess any given character. It is therefore advisable to spread one's net widely rather than to risk excluding forms because of some aberrance. The inclusion of a small amount of possibly atypical or unsatisfactory material, which may have to be excluded in the final analysis, is a worthwhile insurance against an unrepresentative study. Very similar considerations will apply to material which has been extracted from the literature, though the danger of mistakes is naturally greater.

There is no need to belabor the point, now very well understood, of choosing adequate numbers of specimens for establishing taxa. These could be individuals for establishing species, or species for erecting higher taxa. To some degree the exemplar method proposed in the next chapter (Section 6.4) is in contradiction to this admonition, since by that method only single representatives of given taxa are used in the studies. However, it is clear that such exemplars are only reference points and do not indicate the limits of the taxa which they represent. The special problems attendant on material of different ages, different stages of the life cycle, and fragmentary material are discussed in Sections 5.3.7 and 6.5.

5.2. DEFINITION OF TAXONOMIC CHARACTERS

Procedure in taxonomy, orthodox or numerical, is based on taxonomic characters. The term character has been employed in at least two distinct senses by systematists. Its commonest usage is as a distinguishing feature of taxa—a characteristic (or feature) of one kind of organism which will distinguish it from another kind. Thus, serrated leaves may distinguish one species of plant from another and hence are called a character; similarly, punctate elytra may differentiate between two species of beetles, or resistance to phenol separate two strains of bacteria. This appears to be the sense in which Mayr, Linsley, and Usinger (1953) define a taxonomic character, as “any attribute of an organism or of a group of organisms by which it differs from an organism belonging to a different taxonomic category or resembles an organism belonging to the same category.” Unfortunately such definitions point out again the circular reasoning which may infiltrate taxonomic procedure: if the term character is restricted to differences between *taxa*, the taxa themselves cannot be recognized without the characters themselves being first known.

Another frequent meaning of the term character, which has been

espoused by numerical taxonomists as being the more useful one in their work, is that a character is a property or "feature which varies from one kind of organism to another" (Michener and Sokal, 1957) or "anything that can be considered as a variable independent of any other thing considered at the same time" (Cain and Harrison, 1958; we assume the independence referred to is logical rather than functional or mathematical). Thus, referring to the previous examples, the nature of the margins of the leaves becomes the character, while entire, serrated, undulating, or any other types of margin become different *states* of the character (Michener and Sokal, 1957). Cain and Harrison (1958) would call these the *values* of a character. We prefer the earlier term, since it does not imply a quantitative expression and is thus more suitable in cases of qualitative difference. The word "state" may imply qualitative rather than quantitative subdivision, but in the absence of a more suitable term we employ it to cover both. In the sense in which these terms will be used in this book, the condition with respect to punctuation of the elytral surface and the property of phenol resistance would be characters, while smooth or punctate and resistant or susceptible, respectively, would be character states.

No objection should be raised to defining a character as a feature which varies from one organism to another. However, if we say that it varies between kinds of organisms (or species) then we are ourselves in the same sort of dilemma as a systematist attempting to apply the definition of character in the first paragraph of this section—that is, defining characters on the basis of predefined taxa. Thus we would first have to define our species before we could define the characters. To be extremely critical, therefore, we would have to define characters entirely on the basis of the differences between individuals. Specific characters are, of course, nothing but summaries or abstractions of the characteristics of a large number of individuals.

Sneath (1957b), partly to make his data conform to his methodology, but largely because of the peculiar nature of taxonomic characters in the bacteria, established characters (which he called "features") of two states (which he called positive and negative features or values). In the interest of generality we shall also use the character-state terminology for this type of data.

The general definition of characters established above cannot, however, give much aid to the practicing taxonomist in the process of recognizing and describing individual characters. We shall go into this

problem in the next section, from both the theoretical and practical points of view.

5.3. UNIT CHARACTERS

5.3.1. Theoretical considerations

The person embarking on work in numerical taxonomy is apt to be puzzled by the task of recognizing the basic units of information for the study. We have called these the *unit characters*. Cain and Harrison (1958) have called them "single characters." In trying to define these characters we must first ask ourselves what properties we wish to recognize. Do we wish to recognize genes, or a unit element in selection, or a logical construct? If, as is now clear, genes are themselves complex entities, shall we subdivide them? And if so, to what extent?

Most definitions of a unit character have been too restricted; defining a unit character in terms of morphology, chemistry, genetics or evolution does not allow the broad treatment needed for a general theory of systematics. For this we need to define unit characters in terms of information theory, for in every instance it is information which the characters convey to the taxonomist; this idea is closely linked to the concept of natural taxa as groups with high content of information (see Section 2.2). We may also plausibly interpret the general theory in terms of modern genetic theory. An attempt to introduce the concept of information has been made (Sneath, 1957b), and a unit character (there called a "feature") was defined as an attribute possessed by an organism about which one statement can be made, thus yielding a single piece of information. These attributes are formally logical constructs, since they will change if the technique of observation changes; the definition is therefore an operational one. Where a character can vary continuously, such as the length of an organ, this character of length is broken down into as many steps as the observational method will allow with good reliability. Either each step is counted as a feature, or at least the minimum number of features necessary to account for the existing variation is postulated.

This approach may be carried to its logical conclusion, where each unit character or feature represents an alternative which can be answered as "Yes" or "No," "Possessed" or "Not Possessed"; the information content can then be measured as "bits," as is usual in information theory.

(A "bit" in information theory is a single piece of information conveyed, for example, by one electrical impulse; it can be represented by a binary digit.) Each feature is one bit: a character with x states is composed of no more than x bits. In practice we sometimes have to use a third category "Not Known" or "Uncertain if Possessed," but in principle this is not needed.

Clearly even the most simple organisms contain a great many bits of information. We may plausibly relate this number to the genome in the following way. On the current hypothesis of genetic structure, the genome consists of a series of nucleotides which are paired one-to-one in a double helix of deoxyribonucleic acid (DNA). On the Watson-Crick model of DNA each adenine residue is paired to a thymine residue and vice versa, and each cytosine residue is paired to a guanine residue and vice versa. The genetic information is then postulated to reside in the sequence of the different nucleotides and may be thought of as a code message written in an alphabet of four letters, each letter representing one of the four alternative nucleotides—those containing thymine, cytosine, adenine, or guanine. The genetic code message is then believed to be translated into other codes determining the amino acid sequence of proteins and the structure of other molecules in the cell, and these in turn determine the physiological and morphological properties of the organism (this is excellently discussed by Anfinsen, 1959). We may tentatively identify our taxonomic bits with the genetic code, and since it is believed that all or at least the great bulk of the DNA is functional, we can calculate on this hypothesis the number of nucleotide pairs and the number of taxonomic bits per genome (each nucleotide pair is equivalent to two bits, since there are four alternatives). The weight of a nucleotide pair is about 10^{-21} g and a diploid mammalian nucleus contains on the order of 5×10^{-12} g of DNA, or about 5×10^9 nucleotide pairs; a haploid chromosome set will contain half this. This figure of 2.5×10^9 pairs will comprise 5×10^9 bits, which will mean that there will be $2^{5,000,000,000}$ possible alternative combinations of nucleotide pairs in the haploid genome. Dobzhansky (1959) and Sneath (1963) list some figures for the nucleotide pairs in various organisms, from which we may choose a few examples. The number of bits of information in a haploid set of chromosomes of man and of the horse is about 6×10^9 , of an *Amphiuma* about 1.7×10^{11} , of a crab about 3×10^9 , and of a sponge about 1×10^8 . Some other examples of the number of bits in the haploid state (calculated from data cited by Jacob and Wollman, 1958; Ogur et al., 1951; Frisch-Niggemeyer, 1956; Pontecorvo and Roper,

1956; and Allison and Burke, 1962) are: lily, 1×10^{11} ; *Drosophila*, 1.5×10^8 ; *Aspergillus*, 8×10^7 ; the bacterium *Escherichia coli*, 2×10^7 ; T₂ bacteriophage, 4×10^5 ; most DNA viruses, 3.6×10^5 ; and some RNA viruses, 1.2×10^4 . The number of functional genes is of course much smaller, since each gene is made up of many bits. These figures are only speculative, and do not include any non-chromosomal genetic information or environmental effects, but it is clear that the content of information is much smaller in microorganisms than in higher organisms. The scale of the potential store of information in a nucleus may be judged from the estimate that the Library of Congress contains between 10^{13} and 10^{14} bits (see Good, 1958). It is interesting to note that Elsasser (1958, pp. 100–104) estimated from morphological considerations that the information in man was at least 10^7 bits. It should be noted that the bits are a measure of the potential information content, not of the number of alternative permutations of the information. The latter is 2^x when the number of bits is x , as shown in the example above.

5.3.2. Working definition

Except in those few and simple organisms whose fine chemical structure is gradually being unraveled, the above considerations are premature. A workable definition which the practicing taxonomist can use is needed. We may define a unit character as *a taxonomic character of two or more states, which within the study at hand cannot be subdivided logically, except for subdivision brought about by changes in the method of coding.*

Since we cannot in most cases make genetic inferences from phenetic studies of characters, we shall generally have to use phenotypic characters as our basic information, defining these as narrowly as possible. Our failure to make *logical* subdivisions may rest on ignorance of the finer structure or the causation of a character. Thus presence or absence of a bristle in an insect may be a unit character, if we know nothing of its finer structure and have no way of subdividing it. Even if the general morphogenesis of the structure is known from a representative form, unless morphogenetic differences can be established which can serve as taxonomic characters within the group studied, the presence or absence of the bristle remains the unit character. To consider another character, the same insect may possess or lack DDT-dehydrogenase, which character we are again unable to subdivide further in view of our present knowledge. Thus, the organizational levels of unit characters may differ considerably from character to character and with advances in our

knowledge. The ruling idea is that each character state should contribute one new item of information.

5.3.3. Inadmissible characters

The proper selection of characters is clearly a critical point in the application of numerical taxonomy, as it is in other taxonomies, and misunderstandings have arisen on this score. Full discussion of this subject is intentionally delayed until Section 5.4.4, when more of our general viewpoint will have been presented. There are, however, certain kinds of characters whose nature clearly disqualifies them from employment in a numerical taxonomic study. These are listed in the present section as inadmissible characters.

5.3.3.1. *Meaningless characters*

It is undesirable to use attributes which are not a reflection of the genotypes of the organisms themselves. For this reason taxonomists do not include the names or numbers given to specimens, nor do they employ characters whose response to the environment is so variable that it is not possible to decide what is environmentally and what is genetically determined. The number of leaves on a branch of a tree may be an example of the latter, though if acceptable evidence is forthcoming that this number is relatively constant in a species it might be admissible. This is a matter of scientific judgment, not simply of taxonomic method, and each case must be treated on its merits.

5.3.3.2. *Logically correlated characters*

We must exclude as redundant any property which is a logical consequence of another. We cannot use both hemoglobin and redness of blood if the latter is defined as possession of hemoglobin. Mathematical manipulations which constitute logical consequences should be avoided: for example, we could not employ both the length and half the length of an organ, or the radius and the circumference of a circular structure. Similarly, characters which are tautological—those which are true by definition and those which are based on properties known to be obligatory—should not be included. An example of tautology is to score both length and height of a man. An example of a character which is true by definition is to score “presence of calcium in raphides” after having scored “raphides composed of calcium oxalate”; by definition, this substance contains calcium. To score in this instance “raphides insoluble

in acetic acid” would be scoring a character which is a known and invariable property of calcium oxalate, though it is not part of the definition of that compound; if this is known, the property must be omitted. In making these qualifications for admissibility we are fully aware that many or most of the “inadmissible” characters would be inadmissible on more than one count. Thus, if we use two tautological characters, we would find on examining our data after they had been prepared for machine computation that the two characters are perfectly correlated. According to our rules on empirical correlations (see Section 5.3.3.6) there would be much suspicion about using both of these characters. It is quite likely that we would therefore reject one of them by the empirical correlation criterion.

5.3.3.3. Partial logical correlations

Many cases will arise where the dependence of one character upon another is not total but only partial. Cain and Harrison (1958) illustrate this by an example.

Degree of melanization of the skin in mg./sq. cm. must not be used together with skin colour estimated by some colorimetric method, if the melanin is making a contribution, which is some function of its own density, to the skin colour, unless this contribution can be subtracted from the measurement of skin colour.

We would recommend the following procedure in cases of partial logical correlations. When a character **B** depends in part upon another character **A**, the decision whether to employ **B** as well as **A** will depend on the nature of the factors other than **A** that affect **B**. If, to the best of our knowledge, these factors reflect heritable variations, we would include **B**. But if these factors represent experimental or technical error or are otherwise unaccountable, we would not use character **B**.

5.3.3.4. Invariant characters

We would exclude characters which do not vary within the entire sample of organisms or taxa. To include them would not add any information about affinities within the sample. Employment of invariant characters would simply raise the magnitude of all the coefficients of similarity uniformly and would thus diminish the differences among the coefficients; these differences, however, are the very facts which we would like to bring out and study. There may be instances, notably in

bacteriology, where established techniques prescribe a list of tests to be performed. It may be argued that from the point of view of obtaining standardized results all tests should be included in the computation; however, since all coefficients obtained are only relative quantities, we would recommend that invariant characters not be included.

5.3.3.5. *Experimental error*

One should guard against subdivision of characters on grounds subject to experimental error. Thus, it would not be proper to divide hemoglobin content of mammalian blood into classes 1% wide, since the customary method of determination is not accurate to this figure.

5.3.3.6. *Empirical correlations*

How should we decide if two characters not logically related, but highly correlated empirically, are to be counted as separate unit characters? It is possible to give extreme examples which are absurd. Thus it is observationally true to say that certain avian characters are invariably associated, and likewise certain mammalian characters. Should we attribute this effect to a single character in which birds and mammals differ—a gene, perhaps, which if it mutated would turn a bird into a mammal at one jump? Clearly, we would here prefer to postulate many independent genes, and we would treat these features as independent despite the strictest correlation. In still other instances we would not assume independence so easily. The close correlation between pink eyes and white skin of total albinos in most vertebrates would be counted as a single character, since the total absence of pigment implies lack of retinal pigment.

Yet it remains true that we often need to postulate independent characters even in cases such as the albinos, for occasional albinos do have some retinal pigment. The same is true of most other apparently dependent associations. Any exception will suffice to prove that more than one character is involved. Even strictly functional associations are not as dependent as they seem at first sight: the need for the teeth to meet is only true for a species as a whole; aberrant individuals can and do occur. The fact that selection keeps two characters (the position of the top teeth and the position of the bottom teeth) in close correspondence does not necessarily imply unitary causation of these characters. In coding such a species for analysis we would employ two characters in spite of their stringent empirical correlation in the material at hand.

In serially homologous structures such as segments of an annelid or

appendages of an arthropod, or in generally homologous structures such as hairs on the body surface, a character affecting equally all the members of the series could be subdivided into separate characters for each member. However, no new information would be brought about by such a procedure. In such a case we would employ only one character.

In summary, when we have evidence that more than one factor affects two correlated characters within a study, regardless of whether this evidence comes from within the study or from outside, we would include both characters; otherwise we would employ only one. Our position is that we assume at least some independent sources of variation in any empirical correlation, unless we have reason to believe otherwise. This would err in the direction of redundancy, but it would be counterbalanced by the likelihood of obtaining new information.

5.3.4. Operational homology

When scoring the characters of the organisms for comparison, we must decide what is the "same" character and the "same" state in two forms. In many instances there will be no problem. In others, especially if taxa are distantly related, we will be uncertain, and it is tempting to appeal to homology to answer our question. We have seen in Section 2.3 that it is in many cases difficult, if not impossible, to decide if two characters are homologous by the criterion of common descent (or by being traceable to a common ancestral organ or structure). By this criterion the characters are the "same" only in the restricted sense of having the same evolutionary origin, of which we may have little evidence. For example, is the amino acid methionine homologous by descent in bacteria, birds, and begonias? In order to cut this Gordian knot we prefer to call two character states the "same" whenever they are indistinguishable. Similarly, if the abstraction or idea of two characters cannot be distinguished in the taxa, then again we would consider them to be the "same." In practice the worker will divide his organisms into major structures or other such divisions (for example, head, limbs, leaves), and he must first decide whether these are the "same." Then he can proceed to look for differences in their properties which may be used as the bases for setting up the characters and their states. Within each major structure he again looks for subsidiary structures and repeats these procedures. Characters in this sense are synthesized from the states.

For example, consider two species of insects that are both black, while others in the same genus are red. If we had no way of distinguishing the two kinds of black, we would consider them to have the same character

state, "black." Similarly, we would consider "redness" and "blackness" to be states of the same character, "body color," unless we had reason to believe that this color was of a different nature in some of these insects than in others. If, for instance, we found colors due to pigments as well as colors due to optical interference phenomena (such as iridescence) occurring in the group of insects under consideration, we would then subdivide our former character "body color" into two: "pigmental body color" and "structural (interference) body color."

By way of another example, leaves on a given series of plants may be long or short. We first have to decide what a leaf is and whether the structures seen on the separate specimens are in fact leaves—that is, are the "same"—or perhaps are other structures such as modified stems. Having decided that they are leaves, we also have to agree on what we shall call a short leaf—perhaps one of less than 3 cm; if so, we shall call a long leaf one that is longer than 3 cm. In this sense all leaves shorter than 3 cm are homologously short leaves, those longer than that are homologously long. The character will now be called "length of leaf," with two states, "short" and "long." Clearly it could have had more states had one wished to construct it in such a manner. In each leaf other characters, such as the venation, pubescence, and similar properties, could be coded as characters.

We call this approach "operational homology." Our position is largely that of common sense; when we say that two characters are operationally homologous we imply that they are very much alike in general and in particular. If the characters are "not quite the same," then more than one character is involved, and they should be broken into several independent ones; some of these independent characters will then be indistinguishable and will be scored as "the same" character in the two organisms.

Special problems in defining homology arise in the case of rudimentary or degenerate characters. In such cases it would often be impossible to surmise supposed lines of descent, without intermediate forms represented by developmental stages or by fossils. The actual establishment of the homology even in such cases remains on morphological grounds—by comparison of adjacent stages in each ontogenetic or fossil sequence. The procedures involved are therefore not especially different from those of operational homology. When fossil evidence is lacking, homologies of rudimentary organs are frequently established on the basis of a comparative series of recent forms, showing varying expression of the structure in question. Here again a comparative morphological approach is used

between adjacent recent forms, and the evolutionary sequence is then inferred. While a phylogenetic theme underlies all such comparisons, taxonomists are in fact establishing homologies only on phenetic evidence in these cases, just as the vast majority of taxonomic work is based on phenetic classifications, notwithstanding phylogenetic protestations to the contrary. A phylogenetic concept of homology, while simple to define and satisfying to contemplate, is not susceptible to direct proof but only to proof-by-inference. Zangerl (1948) has pointed out quite correctly that any phylogenetic definition of homology robs the concept of its only possible function, namely as a tool, since we do not and cannot know anything a priori about the causality of a given structural relationship between parts of *different* organisms.

Despite this superficial air of simplicity, operational homology does present considerable logical difficulties which have seldom been discussed. We should first mention that this was close to the original meaning of the term "homology" as employed by Richard Owen (see Simpson, 1961, p. 79). Owen used it to mean the "same" organ in two organisms, as opposed to analogy, which was "different" organs fulfilling the same function. Whether an organ was "the same" was judged by its morphology and the correspondence of its parts. Woodger (1945) has discussed this in some detail and has shown the close connection between homology in the sense of morphological correspondence and a common structural plan or *Bauplan*. He points out that in making such morphological correspondences we pair the different parts of the structure in two organisms, with the aim of obtaining the greatest number of one-to-one pairings. For example, if we pair head of cat with head of dog we find numerous subsidiary pairings within this major pairing—for example, eye with eye, brain with brain; further pairings occur within the latter structures—lens with lens, retina with retina, cerebellum with cerebellum, and so on, down to histological levels and farther. A similar comparison between head and leg would show few such correspondences. Woodger illustrates his arguments by the example of the pentadactyl limb. He shows that the pairing of the "same" bone in two forms—as humerus with humerus, radius with radius, and so on—depends on the spatial relations of the bones. For example, the humerus is proximal to the other bones, the radius and ulna are both immediately distal to the the humerus, the ulna being postaxial to the radius, and so on. In a newly studied creature we call a bone the radius if we find that it bears these relations to the other bones, and if they in turn bear their own proper relations to each other. Such matching sets of bones he calls

isomorphic; such sets form a *Bauplan*. He also discusses the difficulties which arise if some bones are atrophied or if their articulations are abnormal, when it may be difficult to recognize which bone corresponds to which.

Woodger correctly says that morphologists pair off organs one with the other intuitively so as to make the greatest number of one-to-one correspondences. We may add that this process is closely analogous to classification of organisms themselves. The aim is to make pairs within which are further pairs, and ultimately to group structures so that there is the greatest number of common properties in the "organ taxa" so set up. This is evident from a consideration of why we pair an eye with another eye and not with an ear. In both alternatives there are some shared properties (both organs are carried on the cranium, both are special sense organs) and some differences (no eye is identical with any other eye), but there is no single property which we can satisfactorily postulate a priori as being essential to the definition of an eye or an ear (since we may be sure that some morphologist will soon find an exception to our rule). Therefore the concept of morphological correspondence is polythetic (see Section 2.2), just as natural taxonomic groups are; hence morphological correspondence can, like natural taxa, only be defined as arrangements by which the groups so formed possess the greatest possible number of common properties. Such reasoning leads us to the threshold of an empirical "numerical homology"—that is, a numerical taxonomy of organic structure.

It is somewhat embarrassing to find that within the concept of natural taxonomic groups there are similar concepts of natural organ groups, and we may question whether there are not yet other concepts (such as natural gene groups) concealed within these, like a nest of Chinese boxes. There must, however, be some limit to this process, even if the limit lies at the fine structure of the genes.

The presence of an enzyme such as β -galactosidase in both an insect and a vertebrate raises an interesting point. Could the enzymatically active area of the enzyme in both animals be derived virtually unchanged from a common ancestor? This might happen if any change in this area led to loss of activity; such change would then be selected against. The rest of the enzyme might become different in the two forms because of the accumulation of changes in the molecule which were advantageous because they assisted the action of the enzyme in the particular conditions prevailing in the two lineages. We can infer that this has happened to some extent, since we know from serological and biochemical evidence

that the same enzymes in closely related genera are usually markedly different in some properties, and yet they are most probably derived from a common ancestor (see, for instance, the work of Vithayathil et al., 1961). We do not know how far we may plausibly trace back the homology. Nevertheless, there does seem to be a possibility of the re-evolution of an enzyme if it has, by an extensive genetic deletion, been totally lost. The sporadic occurrence of enzymes unusual in animals, such as the carbohydrate-splitting enzymes of the snail, argues in favor of the possibility of evolution *de novo*, for it would be difficult to believe that these enzymes in snails are derived from their remote protistan ancestry. It seems very likely, however, that when such re-evolution does occur it is due to reduplication of a gene controlling the production of a similar enzyme followed by change in the enzymatic properties of one of the duplicate enzymes. It would otherwise be difficult to understand how the functional enzyme could be selected for during the period when presumably inactive protein molecules were being synthesized, under the control of the evolving but useless gene. The genetic information required seems to be much too great to be acquired by a single chance mutation, since the probability of a random sequence of amino acids being enzymically active must be exceedingly small. A further possibility is the transfer of genes from very dissimilar organisms. In the bacteria we know of mechanisms by which large amounts of genetic information (equivalent to complete genes and to complex biosynthetic pathways and their enzymes) can be transferred from one bacterium to another without direct cell contact. This can occur by the transfer of DNA either by a bacteriophage or else directly by solution in the medium, and in principle there seems to be no reason why genes should not sometimes be transferred to distantly related bacteria. If animal and plant viruses can in the same way carry genes derived from their hosts, or if direct DNA transfer can occur with successful incorporation into the genome of the recipient organism, we can envisage how a gene might be transferred to distantly related plants or animals. It is therefore not beyond the bounds of possibility that some unusual features of higher organisms are derived by gene transfer from very dissimilar forms of life.

So far as numerical taxonomy is concerned, the position on homology is no worse than that in conventional systematics: in both we have to decide as best we can what characters are "the same" in different organisms, and in the end this entails the application of uniform criteria to individual specimens; what is not distinguishable on these criteria must be considered as "the same" in these respects. Lest it be objected

that one must first achieve a "numerical homology" before proceeding to numerical taxonomy, we may point out that it is possible to formulate a list of characters and their states so that they do not presuppose a prior knowledge of homologies. Such an approach has been successful in constructing phenetic classifications in bacteriology. There is no reason why it should not succeed with other organisms.

5.3.5. Two-state versus multistate characters

Characters of two states (or all-or-none characters) may be recorded as + and - or as 1 and 0. Provision must also be made for entries which cannot be scored, either because the characters are missing, unscorable, or inapplicable. For example, the data may be unknown, or the specimen damaged, or there may be a "logical bar" which prevents scoring of the entry. Such entries have been scored by Sneath (1957b) as NC (standing for "no comparison is to be made with this entry"); a special symbol may be convenient for use with computers, where a blank space may be ambiguous. This straightforward method of coding is referred to by Sneath (1957b) as method *A* and by Beers and Lockhart (1962) and Beers, Fisher, Megraw, and Lockhart (1962) as method 1.

It is usual to record positive characters (or attributes marked as present) as + (or as 1) and negative characters (or attributes marked as absent) as - (or as 0), but in most applications it is immaterial whether characters are scored as + or -. For ease in comprehension, however, it is usual to follow the convention above, in particular where an organ is missing, since the negative sign is a clearer indication that attributes belonging to that organ must be scored NC.

When only two alternatives present themselves to the observer, there is no choice. However, two kinds of multistate characters can occur. The states of *quantitative multistate characters* can each be expressed by a single numerical value; that is, they can be arranged in order of magnitude along a one-dimensional axis. Examples are the amount of a chemical produced by a bacterial strain, length of an animal, or amount of pubescence on a leaf. Characters such as this can almost always be conveniently coded into a number of states (1, 2, 3, ...) corresponding to their range of variation.

Often problems of scale will arise. If the amount of some substance produced is 1, 10, 100, and 10,000 units, respectively, in four taxa, one can code these 0, 1, 2, and 4, using a logarithmic transformation. This indeed is desirable rather than scoring it in its original scale, since the

untransformed variable would exert excessive weight in most coefficients of similarity even if the characters were standardized. Transformations of this type are, of course, standard in statistical procedure. We should emphasize here that no character must be allowed to assume excessive weight; to permit this would make nonsense of our attempt to choose a wide and numerous sample of characters. It is far better to reduce the weight on such a character by transformation or standardization than to run the risk of its swamping the measures of affinity.

Should quantitative multistate characters be turned into several two-state characters, as shown by Sneath (1957b)? There is no practical necessity for this step when the method of similarity estimation is by correlation or distance analysis. These methods allow for a quantitative distinction between dissimilarities arising from two taxa possessing adjacent character states and those arising from taxa exhibiting states near the opposite ends of the scale. However, when association coefficients are to be computed, most of which require two-state characters, we have the choice of recoding the character in some fashion into several two-state characters or arbitrarily dividing the scale into two (not necessarily equally long) parts. The first course has much to recommend it. Quantitative multistate characters are very likely to be caused by more than one genetic factor and several two-state characters may thus be more appropriate. But since we do not in fact know whether one or several factors (genes?) are behind the expression of even two-state characters, use of the second option may be more conservative, while undoubtedly losing some information.

Qualitative multistate characters are those in which the several states cannot be arrayed in some obvious order but still refer to a unit character on logical grounds. An example would be sculpture patterns on the surface of an organism or alternative color patterns of a given structure. Here, ordering of states along a scale has been defended by some more or less hypothetical evolutionary sequence in which the supposedly primitive state is placed at one end of the scale in instances of linear sequences, or toward the middle of the scale in instances of V-shaped evolutionary sequences. The dangers inherent in such a procedure have been stressed in Chapter 2. Ordering of states along a scale might be followed in cases of other sequences: metabolic chains, ecological (successional) stages, and others.

If the investigator wishes—and also in cases where no reliable sequence can be established—an alternative course may be followed. This converts the qualitative multistate character into several new characters. This

step is frequently not an easy task since the recoding has to be done in such a way that a positive score on one of the new characters does not automatically bring about negative scores on all other such characters derived from the same qualitative character. This is the problem of avoiding logically correlated characters, which is discussed from a more general point of view in Section 5.3.3.2.

A detailed exposition of techniques for recoding multistate characters into two-state characters is given in Section 5.3.6, below. The general reader who is not faced with this particular problem can, however, skip that section without losing the continuity of the argument. The problem of coding characters which are variable within the fundamental taxonomic units is discussed further in Section 6.4.

5.3.6. Recoding multistate characters into two-state characters

Persons wishing to employ association coefficients (see Section 6.2.1) may need to convert multistate characters into two-state characters. This procedure is also of interest if the characters are to be handled by information theory techniques as bits. Such recoding can be done in a number of ways, but it presents some logical difficulties, since one must decide whether the attributes comprising a multistate character should be treated as additive or nonadditive.

Additive coding. By this method the multiple character states are scored as a series of + signs. Thus a character with five states could be coded as follows:

		Two-state characters			
		1	2	3	4
Multiple states	0	—	—	—	—
	1	+	—	—	—
	2	+	+	—	—
	3	+	+	+	—
	4	+	+	+	+

In this way a multistate character of x states is turned into $x - 1$ two-state characters. The scoring is termed additive since the state 3, for instance, is expressed as the sum of the effects of the positive states of the two-state characters 1, 2, and 3.

As with all-or-none characters this convention (few + signs for low quantitative values and many + signs for high quantitative values) is

followed for two reasons: it is easier to understand and one must follow one convention consistently to be able to recognize overall size in characters where one can properly talk of size.

Additive scoring may exaggerate dissimilarities due to differences in overall size, although this effect is in general not pronounced, as might be thought. It has the merit of retaining the information on the magnitude of difference in the characters. It does, however, involve some logical redundancy. If, for example, an organism is marked — on character 1, it is by definition also — on characters 2, 3, and 4, and if it is marked + on character 4, it is by definition also + on characters 1, 2, and 3.

Nonadditive coding. Suppose we do not wish to assume that the effect on the phenotype of several small genetic changes is additive. We may then set up the following model. Two organisms **B** and **C**, sharing a multistate character but differing in state, may be regarded as being similar in one respect, *X* (in possessing a detectable value for the character) but also different in one respect, *Y* (in having a different value for the character). The magnitude of this difference is not considered. Their similarity on this character is therefore $\frac{1}{2}$. If they possess the same state they are similar in both respects with similarity $\frac{2}{2}$. However, an organism **A** in which the character is not detectable (has the value zero) differs in the first respect *X* from **B** and **C**, but since it could not by definition manifest any positive value of the character, it is not comparable with respect to *Y* or any subdivision of *Y*. All states other than *X* must therefore be scored NC.

The coding scheme will therefore be as follows:

Organism	Multistate character	Two-state characters			
		<i>X</i>	<i>Y</i> ₁	<i>Y</i> ₂	<i>Y</i> ₃
A	State 0 (character undetectable)	—	NC	NC	NC
B	State 1 (weak positive)	+	+	—	—
C	State 2 (moderate positive)	+	NC	+	—
D	State 3 (strong positive)	+	NC	NC	+

This scheme is that referred to by Sneath (1957b) as method *C* and by Beers and Lockhart (1962) and Beers, Fisher, Megraw, and Lockhart (1962) as method 2. It has been used chiefly in bacteriology.

The details of the scoring system suggested by Sneath (1957b) seem in retrospect to be unnecessarily complicated. They dealt at some length with the logic of coding, since it was felt that with the small number of characters usually available in microorganisms it was important to make

the best use of them. We may omit as redundant any character which is invariant in all the OTU's (operational taxonomic units; Section 5.8), and we can therefore dispense with the scoring system there labeled *B*, as Beers and Lockhart (1962) and Beers, Fisher, Megraw, and Lockhart (1962) have suggested. When many characters are employed the different methods will usually give very similar results, and additive coding appears simple and adequate.

The different coding methods above give somewhat different average weights to each multistate character. When each state of the character is equally frequent among the OTU's and the number of such states is large, the mean similarity between all possible pairs of OTU's (expressed as matches out of the total number of comparisons, and including two minus values as a match) approaches $\frac{2}{3}$ with additive scoring, but $\frac{1}{3}$ on nonadditive scoring. If negative matches are not included, the mean similarities are $\frac{1}{2}$ in both instances.

Binary coding. An interesting alternative which can claim some logical basis is to code characters by employing the binary equivalent of the numerical value of the character state. For example, the binary equivalent of 6 is 110 (or $++-$ in the plus and minus convention) while that of 5 is 101 (or $+ - +$). One here uses the convention that the right-hand binary place indicates the "unit" (which determines its odd or even value), the next indicates the "twos," the next indicates the "fours," and so on, in the same way that in decimal notation the right-hand place indicates the "units," the next place the "tens," the next the "hundreds." One may postulate that as a general rule it is permissible to assume that a multistate character is controlled by a number of attributes, assumed to be genetic, which control the size of the character in a similar way. For example, the state 6 is regarded as being the result of the action of genetic attributes *a* and *b* but not *c*, while state 5 is the result of genetic attributes *a* and *c* but not *b*. Each successive attribute then determines a doubling in the size (when expressed as the number of the character state). It must be emphasized that we do not know if this is true for any character, but we do know that a minimum number of attributes is needed to allow expression of a given number of character states, and this minimum number is the number of binary digits in this convention. It is also the most economical hypothesis: it uses the fewest attributes and the simplest relation between the attributes, the relation of doubling.

We would expect that when large numbers of features are employed this convention will give satisfactory results. It also has the merit of

avoiding the difficulty of overall size. The problem of the size factor arises because in additive scoring the presence of some attributes necessarily involves the presence of many others; for example, in the example on additive coding the presence of two-state character 4 necessarily involves the presence of two-state characters 1, 2, and 3. This is not so with nonadditive scoring, which gives an expected similarity value of $\frac{1}{2}$ between two entities, however great the observed degree of difference is (that is, for entities differing in state, and ignoring negative matches and the complication of the completely negative score). Binary coding strikes a mean between these extremes. The relationship between entities in binary coding is, however, arbitrary in one respect. The difference between 7 and 8 (0111 and 1000) is greater than that between 6 and 7 (0110 and 0111). This, in effect, allocates weights at random to the underlying genetic determinants. When the number of character states is large and all states are equal in frequency among OTU's, the mean value for similarity including negative matches is $\frac{1}{2}$.

5.3.7. Age, growth, and allometry

In much numerical taxonomic work, particularly with fully adult individuals, the absolute size of a quantitative character (or the ratio between it and some standard measure, such as length or weight) can be employed directly by suitable scaling and coding. This may on occasion be an unsafe procedure, since the size of the character may be dependent on factors other than age—for example, the state of nutrition or, in bacteriology, the temperature of growth. The problem is particularly acute with fossil material, where one has no direct knowledge of the age of the individuals at death, and commonly too few specimens are available for an indirect answer to this question. In addition, it is generally found that the ratio of the size of the character to some standard character also varies with age. Therefore another way of expressing the character is desirable, and this may be done by means of one or the other of the allometry formulas. The reader is referred to Huxley (1932) and Medawar (1945) for a general treatment of this subject.

In most cases it is found that a straight line is obtained if the logarithm of the size of a character is plotted against the logarithm of age (or the logarithm of a standard character, such as total body length). This relation will prove adequate for most numerical taxonomic work, particularly if, as we advise, an effort is made to restrict the study to

comparable stages of the life cycle. More complex relations, which are discussed in the works cited above, can be handled on the same principles. The usual allometry formula is

$$\log y = \log a + b \log x,$$

where y is age (or some other standard measure), x is the size of the character under study, and a and b are constants describing, respectively, the value of y when $x = 1$ and the slope of the line. We have adopted the customary symbolism of the regression equation rather than the converse one often applied to the allometry equation, in order to promote uniformity in statistical symbols.

The constants of this formula are normally obtained by plotting a scatter diagram of values of $\log y$ against $\log x$ and fitting the line by the least squares method, employing the usual formulas for the regression of $\log y$ on $\log x$. It is, of course, first necessary to be sure that the scatter diagram does approximate a straight line, failing which some other allometric transformation is required. It is also necessary to be sure that the population is homogeneous and does not, for example, consist of individuals of several different species. These points will commonly be evident from the scatter diagram on inspection, and standard statistical techniques can be used to test them. It is probable that electronic computing techniques will be essential in any large-scale work of this kind, both to obtain the regression lines and to check their significance.

Once obtained, the constants a and b can be regarded as characters themselves, since they express the genetic factors which relate the size of the character under study with the standard character. These two constants can then be scaled and coded in the usual way. The operational taxonomic units will now be the populations and not individuals. The standard character will most commonly be overall length, age, or weight, but more sophisticated parameters (such as the geometric mean of length, breadth, and thickness) may prove to be useful. An additional constant, $s_{y \cdot x}$, expressing the scatter about the regression line, may also be employed, since it may be regarded as an additional attribute of the population, but it may not always be clear how much of it is due to errors of measurement, heterogeneity of the sample, environmental effects on the phenotype, and similar factors. Caution is advised here.

Allometry is a problem related to the effect of environment on characters (age, amount of available nutrition, and others) and to the problem of redundancy and empirical correlation; (that is, the crude measures may depend on a small number of underlying causes). The

orthodox systematist faced with this problem chooses specimens of equal age or size for comparison, intuitively judging these as equivalent. Cases occur, however, where choosing equivalent specimens is exceptionally difficult. For example, if an amphibian larva suffers a delayed metamorphosis it may not be comparable either in age or size to any stage of another amphibian, and where delayed metamorphosis is the rule, as in the axolotl, *Ambystoma mexicanum*, this may be a considerable problem. In such cases it may be necessary to restrict the characters to those which do not show pronounced allometric changes, as is of course the practice in orthodox taxonomy in these cases. As explained in another section, we would expect, if the nonspecificity hypothesis is true, that the resulting taxonomy will be generally satisfactory.

D'Arcy Thompson (1917) pointed out that many organic shapes could

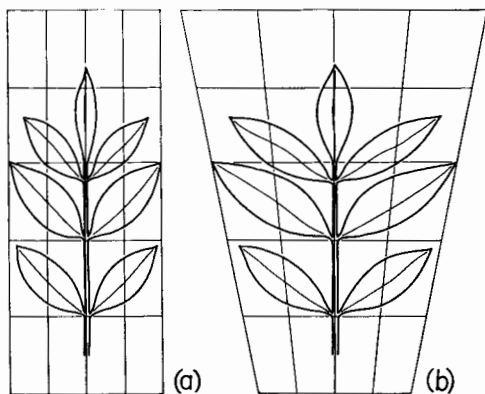


FIGURE 5-1

Transformation grid (after D'Arcy Thompson) applied by us to a hypothetical leaf, showing a regular expansion of the distal part of leaf b as compared with leaf a.

be expressed as simple mathematical transformations of other shapes (further discussed by Woodger, 1945). Raup (1961) has applied this idea to the shape of shells of molluscs. He shows that the shape of many gastropod shells can be specified by the profile of the whorl at one point together with a function expressing the rate at which the whorl increases in size at each complete turn of the helix. Clearly, these two characters—the profile and the rate of increase per turn—would express the morphometrics of the shell more economically than measuring a large number of dimensions of the shell at random.

The simpler of the transformation grids given by D'Arcy Thompson (see Figure 5-1) can be treated in the same way. If there is a straightforward expansion of one part of the grid, this is the only pertinent difference between the forms as represented there; all the other differences in measurements are a consequence of it. One would prefer to use

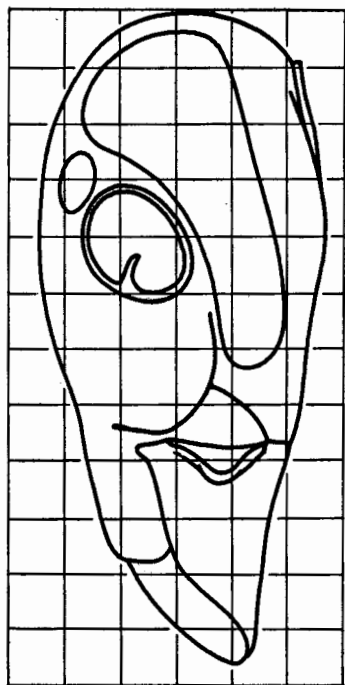
**BLATTA****CHORTHIPPUS**

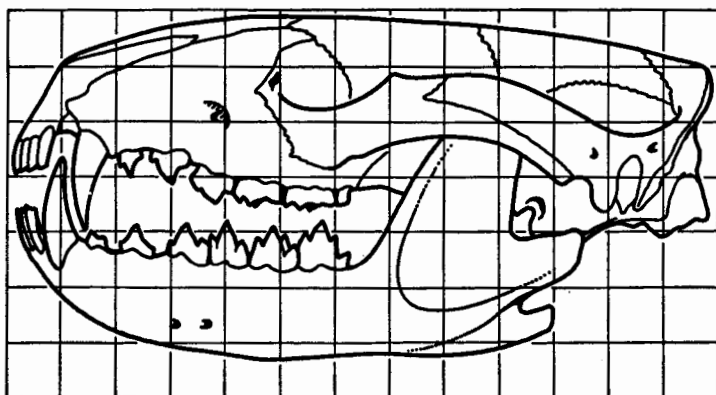
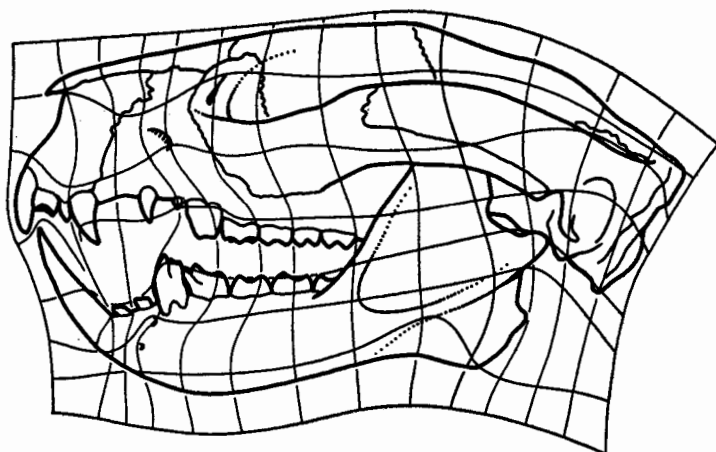
FIGURE 5-2

Transformation grids of complex forms. Above: two orthopteran heads. [Drawn from nature, with the grids added by Mr. A. J. Lee.] Facing page: the transformation between two marsupial skulls. [After Parker and Haswell, and Cuvier.]

for numerical taxonomy a single figure that described this expansion rather than many independent measurements. The latter would indicate a much greater difference than would seem justified; moreover, many of the characters would be redundant, since their values would be logical consequences of the expansion function.

Here again we meet the problem of empirical correlation. The simplest hypothesis is that mentioned above, but it is also possible that the expansion is due to a number of genetic changes, each controlling the expansion of one portion of the body. If we have no evidence bearing on this, we would prefer the simplest hypothesis.

No general and simple methods seem yet to have been developed for extracting the factors responsible for such transformations. While it may be easy to recognize that a figure such as Figure 5-1 is due to a regular

*DASYURUS**PHALANGER*

expansion, it is not easy to see how many separate factors are needed to express more complicated examples, such as the marsupial skulls and insect heads shown in Figure 5-2. Not only are the grid lines deformed in several ways, but the deformation is different in different parts of the skull.

What would be useful would be a way of extracting the minimum number of factors that would account for the difference in form; then these could be employed as characters. It may be that automatic methods can be developed for electronic computers so that diagrams could be compared and these factors extracted mechanically. It would probably be at first necessary to mark operationally homologous points on the diagrams before feeding them into the computer. This would involve superimposing a rectangular grid on one figure and marking on the

other figure the points corresponding to the grid line intersections. Alternatively, the grid coordinates could be recorded for a number of operationally homologous points in the two diagrams.

A related method is that of Smirnov (1927), who resolved the shapes of the elytra of coccinellid beetles into a number of components by a method based on Fourier series.

5.4. KINDS OF CHARACTERS AND HOW TO CHOOSE THEM

Given agreement on what to call a unit character, we next turn to the problem of what kinds of characters to employ as the basic data for a numerical taxonomic study and how to choose these characters. Before we can enter upon a profitable discussion of these points we need to state and amplify two pertinent fundamental hypotheses.

5.4.1. The nexus hypothesis

This assumes that every taxonomic character is likely to be affected by more than one genetic factor and that, conversely, most genes affect more than one character. The result is a complicated nexus of cause and effect. Since numerical taxonomy cannot base its judgments on direct knowledge of genetic factors, it has to rely on information gained from (phenetic) taxonomic characters (as is, of course, done even in genetic research). Any character should give information about several genes. It should be possible in general to pick up the effect of a given gene through any one of several characters.

A detailed justification of the nexus hypothesis would require a long and thorough examination of much genetic evidence, which would be out of place here. We may cite in its support that the multifactorial nature of most characters is well known (Lerner, 1958; Falconer, 1960), and the common occurrence of pleiotropy is even better established (Waddington, 1957). Pleiotropic gene action often involves several organ systems and can be investigated by a number of approaches. Even when a unifying developmental process seems responsible for the pleiotropy, as in the well-known case of cartilage hypertrophy in the rat (Grüneberg, 1938) or in the more recent example of Danforth's short tail in the mouse (Grüneberg, 1958), the characters resulting as intermediate and end products of these processes are quite heterogeneous. While pleiotropic effects have not yet been described for a large number of mutations, this

is probably due to failure to search for them. Even when, as in the ABO blood groups in man, it would appear that the antigens produced are reasonably close to being primary gene products, recent work on their adaptive significance tends to suggest some pleiotropic action (see Sheppard, 1959).

5.4.2. The hypothesis of nonspecificity

Here we assume that there are no distinct large *classes* of genes affecting exclusively one class of characters such as morphological, physiological, or ethological, or affecting special regions of the organism such as head, skeleton, leaves. The emphasis in this statement is on "classes of genes." Although a given gene may have a main effect on one region or kind of character, it is likely to have side effects or chain-reaction effects on other regions and characters.

If this assumption is warranted, then obtaining a disproportionately large number of characters from one body region or of one special kind would not restrict our information to a special class of genes. Furthermore, we would have to conclude that there are no a priori grounds for favoring one character over another. That all kinds of characters, such as bionomic, physiological, morphological, and the like, are of equal taxonomic merit has been quite generally accepted in recent years (Mayr, Linsley, and Usinger, 1953; Michener, 1953). However, certain body regions and organs still hold a considerable mystique for taxonomists as classificatory tools, while others are neglected.

Like the nexus hypothesis, the nonspecificity hypothesis may only hold in part. There may be some genes which affect the characters of one organ only. If such classes of genes are few in number, the nonspecificity hypothesis will still be substantially correct.

5.4.3. The evidence for nonspecificity: taxonomic congruence

Tests of the hypothesis of nonspecificity may be made at a variety of organizational levels in a classification. For instance, we may look for nonspecificity in connection with organs or regions of the body. Thus, if the hypothesis holds strictly, identical classifications should result from characters of the brain and from characters of the intestinal tract; from characters of the epithelium and from those of the connective tissue, or from head versus body or wing characters. All such tests would involve

different portions of the same individuals. Yet there may be dimorphic or polymorphic manifestations of the gene pool of a species. These would lead to tests of congruence between taxonomic systems based on females or males, or based on diverse life history stages, such as insect larvae, pupae, and adults, or on different castes in social insects, or even on different adult forms in cyclomorphic organisms such as aphids. Considered a priori, the hypothesis of nonspecificity would seem more reasonable when applied to organs than to sexual or life history forms. Hennig (1950, p. 151) concurs in this.

What evidence is there for or against the hypothesis of nonspecificity? We may start our discussion with some general observations: it is obvious that the hypothesis must be generally true; on the whole, classifications based on separate sets of characters, be these from organs or life history stages, agree well with each other. It is well known that revisions of established groups based on morphological evidence from a newly investigated organ system frequently confirm previous classifications. If there were no relation at all between classifications established on the basis of larvae or adults, or males or females, or on muscles versus bones, this would have emerged before now. But perfect congruence between such systems has not been shown in the few instances where this has been investigated with any precision.

One must have many characters for measuring congruence. Admittedly, when a life history stage shows very few observable characters, one might be misled. We believe this to be so because the known characters are so few. Should serious incongruences between life history stages occur in a given study, it would raise the question of whether the classification of either stage is reliable. It would be difficult on theoretical grounds to justify any decisions between taxonomies based on one stage and those based on another stage. Where possible, one should choose characters from all stages of the life cycle when performing a numerical analysis.

That the hypothesis of nonspecificity on the whole must hold is shown in vertebrate paleontology and in much of vertebrate systematics where classifications are based on skeletons or even portions of skeletons. Yet consistent classifications can be established from such material. This means that other characters clearly are correlated with the skeletal characters. Thus, in a general way (and for the higher categories), the hypothesis of nonspecificity is supported by paleontology. We are not aware of any studies specifically and quantitatively testing this point.

An argument against the hypothesis rests on cytological evidence that certain genes appear to be active only at specific times in the life history

of an organism (see Clever, 1961). If that is indeed so, then there would be separate classes of genes (at least as regards ontogenetic stages) leading to incongruence between life history stages.

Might the hypothesis of nonspecificity be entirely invalid? If we assumed that each character is affected by one gene only and that any gene affected only one character, then the absurdity of the extreme version of such a belief should be quite evident from the virtually infinite number of possible characters *vis-à-vis* the finite number of genes in an organism, and also from the existence of character correlations within freely interbreeding groups. Furthermore, if separate classifications could be set up by dividing taxonomic characters into logical sets—such as between sexes, between life history stages, head versus body, outside versus inside, nervous system versus circulatory system, and others—the large if not infinite number of independent classifications which could be set up in that way speaks equally for the absurdity of such a view.

It would therefore seem that incongruence is not an all-or-nothing phenomenon but that in different taxonomic groups varying amounts of incongruence will be shown by various sets of characters. It will be one of the tasks of numerical taxonomists in the next few years to investigate quantitatively the extent of such incongruences. Hennig (1950), writing before the development of numerical taxonomy, has clearly seen this problem, although to him the congruence of two classifications based on separate sets of characters proves the phylogenetic validity of the established system. Since he knew of no satisfactory way of quantitatively describing the similarity of two organisms, he felt that incongruences could not be quantitatively evaluated. We hope, however, to have shown in our present work that such comparisons are feasible and legitimate. Hennig (1950) and Remane (1956) both discuss the problem of nonspecificity in some detail, citing the work of Thienemann (1924, 1936) and Thienemann and Krüger (1937) on species and genera of chironomids. Thienemann, aiming at phylogenetic classifications, came to the disturbing conclusion that concordant classifications of larvae, pupae, and adults are impossible to achieve. Remane believes that he can order Thienemann's data into a proper phylogenetic scheme, while Hennig points out that genetic and phylogenetic criteria are necessary to establish species, and consequently the phenetic incongruencies are inconsequential. From our point of view such cases are interesting and challenging for numerical taxonomy to investigate; yet we cannot place much reliance on such data because rarely are they established on known phylogeny or on the basis of adequate numbers of characters studied by

acceptable (quantitative) methods of phenetic comparison. Our confidence in phyletic relationships increases somewhat as we deal with higher categories; however, in these circumstances the hypothesis of nonspecificity is also less likely to be violated. Thus agreements between phenetic and phyletic classifications are more likely at the higher levels, and evidence from such cases would not be critical for taxa at the level investigated by Thienemann.

We think that sex differences will seldom prove troublesome in taxonomy. It seems certain that if one compared within one species all the features of a male with all the features of a female, the overall resemblance in most instances will be very high; the great majority of features will not be appreciably affected by sex. Such strong resemblance may not hold in some highly differentiated groups such as those showing much sexual dimorphism. Good examples would be some birds (ducks, pheasants, birds of paradise), fig insects (*Blastophaga*), Strepsiptera, or *Schistosoma*. In such cases sexual forms should be treated as if they were stages in the life cycle and thus male characters compared only with other male characters, and so on. A similar procedure should be followed in social insects when differentiation among castes is considerable.

We are on less sure ground in some of the most extreme examples of incongruence. It should, for example, be possible in principle to classify tissue cultures or spermatozoa, and we would hope to find their classifications congruent with those based on the whole adult. This has been done to some extent with pollen grains, whose classification was found to be reasonably congruent with orthodox taxonomy (see Erdtman, 1943). Similarly, it should be possible to make classifications of Fungi Imperfecti which are congruent with those based on the perfect forms. Nevertheless, in these cases the numbers of features may well be too few for numerical taxonomic methods. Very similar considerations apply to extremely modified parasites. There is evidence that comparative serology gives great assistance in cases such as these (Wilhelmi, 1940; Perlmann, 1953; Spiegel, 1960), and we believe that in principle numerical taxonomy can do the same.

The general agreement, frequently very close, between comparative serology (Section 3.2) and orthodox taxonomy is another piece of evidence in support of the hypothesis of nonspecificity, though detailed comparisons remain to be made. In studying serological aspects of metamorphosis, Telfer and Williams (1953) found in a silkworm that five out of six proteins were present in all stages examined, while one protein was absent in young larvae and in old adults. An additional

protein was present in female but not male pupae. Perlmann (1953) found that eggs and pluteus larvae of a sea urchin shared ten out of thirteen antigens. The other three could not be detected in the eggs before the gastrula stage. Spermatozoa shared one antigen with the eggs. The relationships among galliform birds inferred from myology (Hudson et al., 1959) were compared to those indicated by serological data of DeFalco (1942) and Mainardi (1959c). There is no close correlation between musculature and serology, but unfortunately only a few forms were studied serologically and the data of DeFalco and Mainardi do not agree well. The evidence is therefore dubious.

Below we cite the few quantitative tests of the nonspecificity hypothesis available so far. Sneath (unpublished) has made a detailed analysis by association coefficients of the data on musculature of galliform birds cited above (Hudson et al., 1959), including also additional characters. Correlations between similarity coefficients based on leg characters and those based on foot characters were $r = 0.720$, while a similar correlation between similarity coefficients based on extensor and flexor muscles yielded an r of 0.785. For comparison, correlations between similarity matrices based on odd- and even-numbered characters (an arbitrary division) yield $r = 0.826$. Since one could not expect on statistical grounds an agreement much better than the last, these correlations are quite good. It may be noted that this is a study of higher ranks—that is, several families of one order.

Haltenorth (1937) found good agreement between affinities based on 44 measures of bone length and 22 measures of the sagittal section of the skulls of eight species of large cats. Agreement between phenetic classifications based on the above sets of characters and 19 measures of head width was not satisfactory. The relatively small number of characters for the sagittal section and head width may make conclusions on nonspecificity tenuous for this study.

The opportunity to test the hypothesis of nonspecificity in a quantitative manner was also provided by a reanalysis of the bee data first published by Michener and Sokal (1957). The 122 characters originally studied were subdivided in two ways: 69 characters found in the female and common to the two sexes were placed in one set, 53 found only in the male bees were placed in a second set. In a second study these characters were divided differently: 60 head characters (both male and female) formed one set, and the 62 others were placed into the "non-head" set. When correlations between all pairs of species were computed (Michener and Sokal, 1963), they turned out to be similar to a consider-

able degree, regardless of the set of characters on which they had been based. The correlation coefficient between male and female classifications (based on product-moment r 's between pairs of species) was 0.71, and the analogous coefficient for head versus nonhead classification was 0.61. The magnitude of this correlation was partly a matter of scale. The analogous correlation between males and females on the distance scale was 0.35, while that between head and nonhead was 0.33. Similar discrepancies between various scales of measurement have been noted by Rohlf (1962). When the data were subjected to closer scrutiny, a number of differences among dendrograms prepared from them were noted.

Rohlf (1962), in a study of the hypothesis of nonspecificity, compared phenetic relations among 48 species of the mosquito genus *Aedes*, based upon independent analyses of characters of adults and larvae. Seventy-one characters were taken from the larvae and 77 from the adults. He found that on the whole the relationships were similar in the two stages but that there were many individual cases (somewhat more than would be expected due to chance alone) in which there were statistically significant differences between adult and larval interrelationships. The correlation between adult and larval correlations was found to be 0.29, and the correlation using distances was 0.59. Both correlations are, of course, highly significant, since there are 1,126 degrees of freedom.

In a quantitative study of two cyclomorphic adult forms (stem mothers and alates of the gall-producing aphid, *Pemphigus populi-transversus*), Sokal and Thomas (1963) were able to resolve the pattern of geographic variation into three factors for alates and two factors for the stem mothers. Only one of the factors determining variation corresponded in the two morphotypes. These data thus do not support the hypothesis of nonspecificity. However, the taxonomic rank investigated is near the lower limit—the local populations within a portion of the range of a species. It appears both from a priori principles as well as from the evidence so far collected that the hypothesis of nonspecificity will hold better, the higher the rank we consider. Thus, for example, there is little doubt that the orders of insects would be as faithfully mirrored in their larvae as by their adults, or that the classes of vertebrates would be recognizable from their skulls, their pelvic girdles, or their circulatory systems, and that these systems would yield roughly the same classifications. However, at lower taxonomic ranks this may no longer follow, and convergence resulting from adaptation to particular environmental

circumstances may confuse the issue, yielding phenetic classifications which differ when based on different sets of characters.

In comparing classifications such as those based on females and those based on males it is not the absolute scale of resemblances that matters but the relative structure to be found in the two similarity matrices. Thus females may resemble each other more than males do; yet as long as the relative resemblances of the females among themselves are similar to those among the males, congruence would still obtain. Correlations among matrices of correlation coefficients, as carried out in tests of the hypothesis of nonspecificity by Michener and Sokal (1963) and Rohlf (1962), and the cophenetic correlation method of Sokal and Rohlf (1962) allow for unequal absolute scales of resemblance among dendrograms to be compared.

In summary of the above, we may conclude that the hypothesis of nonspecificity does not hold absolutely either for parts of the body or for stages of life history. It is more likely to be true for organs than for life history stages and at higher rather than at lower taxonomic levels. The extent of its validity will have to be examined by work in numerical taxonomy to be undertaken on a variety of groups. It should be added that only numerical taxonomy will be able to evaluate the degree to which the hypothesis holds because, as Hennig (1950) correctly and repeatedly points out, problems of this sort can only be solved when an exact method of estimating resemblances has been devised.

What if it can be shown that incongruences among different life history stages are the rule rather than the exception? The procedures of numerical taxonomy would not be invalid, but samples from all the available character sets would then have to be taken. This will have to remain an open question until more evidence has been collected. We would therefore recommend that until such a time undue reliance should not be placed on the hypothesis of nonspecificity and that characters should be distributed as widely as possible over the organisms to be studied.

5.4.4. The choice of characters: Some general considerations

It may be argued by the philosophically inclined that it is not possible to make absolute measures of resemblance because this would involve an arbitrary selection among the endless array of attributes which could in some sense be called characters of the organisms. Nevertheless, mean-

ingful estimates of resemblance can be made once there is agreement on what characters are to be admitted as relevant in taxonomy. It is generally considered that only genetically determined characters should be used in orthodox taxonomy, and with this we concur. However, a study to investigate the influence of environment using numerical taxonomic methods could legitimately include environmentally determined characters.

It is, of course, quite impossible to give an adequate catalogue of all the taxonomic characters that can be used in various groups. Such a catalogue would comprise nothing less than a description of organized nature. Only specialists in the various groups will be in a position to define and describe unit taxonomic characters in the organisms they are studying. In their search for characters they ought to follow two guide lines, one of which at least is not included in customary taxonomic practice. The first is that all kinds of characters from all parts of the body and from all the stages of the life cycle should be used. When the hypothesis of nonspecificity is interpreted very strictly, a stratification of our sample from many body regions would not appear necessary. Any one part, such as the head, and any one kind of character, such as skeletal characters, should suffice. Many paleontological classifications are based, *faute de mieux*, on data of this nature. But this position can be reduced to the absurd or at least the questionable. If, on the basis of the hypothesis of nonspecificity, the skeleton suffices for a valid classification, so must the skull (as much of vertebrate and particularly primate paleontology has indeed inferentially claimed). But if the skull suffices, so must the jaw, or even a sliver or section of the jaw, reaching eventually the cellular level. A purist defending the hypothesis of nonspecificity may argue that jaws are indeed sufficient to establish *overall* phenetic relationships between taxa. The difficulty lies in finding and evaluating enough characters of the jaw. These may need special, refined techniques which are either not available to the paleontologist or (probably) have not even been invented. A conservative position would appear to be the employment of all available characters rather than restriction to a given region or kind.

The second guide line is to use all characters varying within the group studied, not merely conventional diagnostic characters. The latter are likely to be constant within the members of a given category. The exclusive use of such characters would prejudge the very issue—the establishment of taxonomic categories free from subjective bias—which numerical taxonomy wishes to solve. If the studies are based to a very

large degree on characters previously described in the literature, there is some danger that diagnostic characters will be favored inordinately, since there is a historical weight in favor of diagnostic characters in the published literature. A number of critics have used such a possibility in order to question the validity of results obtained by Michener and Sokal (1957). The critics claim that since the characters employed in the study were those previously used to separate the groups concerned, it was not surprising to find that a classificatory technique utilizing these characters would yield a classification quite similar to the one described previously by conventional taxonomic methods. However, previous comparisons generally did not consider all characters simultaneously, so that the various diagnostic characters were each used only for portions of the systematic group studied. Thus the diagnostic value, or lack thereof, of any given character was rarely tested in any but a small section of the classificatory scheme.

Taxonomic characters can be roughly grouped into

- (a) morphological characters (external, internal, microscopic, including cytological and developmental characters),
- (b) physiological characters,
- (c) behavioral characters,
- (d) ecological and distributional characters (habitats, food, hosts, parasites, population dynamics, geographical distribution).

This list is far from being exhaustive. Readers interested in greater detail may wish to consult Mayr, Linsley, and Usinger (1953, p. 108 ff.), who have prepared a more complete classification and discuss each category in detail. Moll (1934) and Lawrence (1955) may be consulted on characters in botany. We have omitted serological characters; as these are analogous to similarity coefficients it would be underweighting them to use them as single characters. But we have as yet no criterion for applying weights to them in combining them with coefficients of similarity obtained by numerical methods.

Ecological and distributional factors, which have become much emphasized in recent years, require some comment in connection with their use for numerical taxonomy. While they are regularly reported, when known, they are not too frequently employed for classificatory purposes. Some difficulty may be encountered in coding them for numerical taxonomy. When an ecological character expresses some sort of gradient such as life zones in mountainous areas, depth in soil, temperature

maxima, and the like, multistate coding is straightforward. However, how are we to code phytogeographic or zoogeographic distributional characters, host-plant preferences, or parasitic fauna? With distributional data, a two-dimensional breakdown into two characters is sometimes possible. In the other cases the information may have to be partitioned into a number of two-state characters. Thus, where several species or genera of hosts occur for a group of parasites under study, each of the former may have to be a single two-state character marked "present" when parasitized and "absent" when not. We may, however, wish to express systematic relations among host plants by appropriate coding. If a group of parasites live on four hosts, species *a* and *b* of genus *X* and species *c* and *d* of genus *Y*, we could have one character for genus *X* or *Y* and one for species (*a* or *b* and *c* or *d*) for each of the two genera.

Geographic distributions are characters which need special caution. In most cases it is not possible to be sure that they represent any character in the genotype. We may, for instance, have no evidence that there are genes conferring adaptation to life in North America as against life in Europe, and the distribution may reflect accidents of migration rather than any genetic attributes. In such instances the characters should not contribute toward estimation of phenetic relations, however useful they may be as pointers to the phylogeny. Similar care is needed with many ecological, behavioral, and parasitological observations. Ecological characters may sometimes be due to the recent invasion of a new habitat, and at first there may be no genetic differences between the population and its ancestral population. Behavioral characters may be acquired in youth by some accident, as for example song peculiarities of some birds, which may be learned by the young from adult birds of their own or different species (Lanyon, 1960). Parasitological characters may sometimes depend on chance infestation and not on the genotype of the host or of the parasite. Attempts to use the Mallophaga in classifying birds, for example, are made difficult by doubts as to whether some mallophagans are stragglers from other birds or have quite recently become established on their hosts through cross-infestation (for example, see Clay, 1949).

Characters such as host specificity also pose another difficulty, that of deciding what weight they should be given. Suppose several species of a genus of aphids live upon trees of several species, genera, and families. We wish to score the aphids so that the scores reflect those genetic attributes of the insects that restrict them to certain of the trees. It seems likely that wide host ranges involve greater genetic differences in the

insects than narrow host ranges, and we may therefore require an estimate of the taxonomic resemblance of the trees, so that we may set up characters for wider and narrower host specificity. It would not be justifiable to carry this to extreme lengths, since the taxonomy of the trees themselves may be uncertain. In virus classification this problem is acute. Some viruses are restricted to one species of mammal, while others attack both birds and mammals. One might score class specificity, ordinal specificity, familial specificity, generic specificity, and so on (possibly introducing subclasses and tribes as well), and one might consider that higher categories should contribute more weight than lower ones. Since we have little detailed information on this problem, and hence no satisfactory method of allocating weights, we would suggest that only a few such characters be used and that each should be given equal weight. Although this may reduce the information it will avoid introducing bias.

Little need be said about avoidance of bias in choosing the characters. It is clear that when we use only a set of characters known to show affinity between certain groups, the similarity coefficients which will result from this study will reflect that choice. In an extreme case, by choosing only those characters which were the same in two organisms, one would obtain perfect but spurious resemblance between them. A systematic survey of all known characters or inclusion of all characters the investigator has been able to observe should avoid bias of this sort.

5.5. PHENETIC VERSUS PHYLETIC CLASSIFICATIONS

5.5.1. General considerations

Our major criticisms of the phylogenetic approach to taxonomy have been discussed in Section 2.3, and our reasons for preferring phenetic classification have been briefly stated in Section 4.6. However, there are some instances where the procedures of numerical taxonomy would yield erroneous results by the currently accepted criteria of phylogenetic taxonomy. These problems would all arise from the possibility of more or less serious discrepancy between relationships based on similarity and those based on descent. Issues such as monophyly, convergence, and parallelism come to mind immediately.

It is almost a truism that an intimate relation must exist between phenetic evidence and the degree of relation by ancestry. It should also

be obvious that, while the two kinds of taxonomy are equally valid for their own purposes, they stand in a peculiar relationship to each other: if knowledge about phyletic relations is required, it must be obtained by conjecture from phenetic evidence; phenetic relations, however, should be deduced not from phyletic hypotheses but from the specimens themselves. A ready simile offers itself here: we may estimate the similarity between geometric objects on the basis of the nature and size of their dimensions, without any implication as to their past history or how they were developed or constructed by geometers. Yet under some conditions we may *deduce* from their geometrical form certain probabilities about their past history. We may, for example, suggest that a regular octahedron was developed by geometers subsequent to the development of the square. Inger (1958) has misinterpreted this point in referring to a hypothetical classification of nuts, bolts, and screws, which, he says, implies no historical relationship among the objects classified. On the contrary, as we must emphasize (see also Sokal, 1959), it does *imply* historical relationships without necessarily being based on these relationships. Any collection of such objects will show unmistakable evidence of their history: there will be nuts, bolts, and screws of simple design and crude workmanship, representing the early attempts at manufacture of these objects; there will be evidence that the simple slot across the head of the screw was an earlier invention than the four-armed crossed slots; there will be evidence of selection, since the most useful patterns will tend to be selected and manufactured by succeeding generations, a close parallel with natural selection and survival of the fittest. To comment that such evolutionary deductions are hazardous is to state the obvious: they are just as hazardous, or more so, in biology, and for the same reasons. We cannot be sure without dated "fossil" examples which form was the earlier and which was the later. It is also clear that, unless we had historical records of their development, we could only deduce the evolutionary relations from the objects themselves; without such records we could not make a genealogical classification directly but only by inference. The classification could not then be based on the evolution, or at most it would only be based on guesses about the evolution; the interpretation of the evolution would be based, rather, on the classification. In any event, it is plain that a classification of nuts, bolts, and screws (and equally, classifications of organisms) would not be fruitless nor lie outside the province of engineering (or biology), even if it took no account of evolution.

Let us take a closer look at the reasons for wishing to develop tax-

onomies which are in accord with phylogeny. This ideal is expressed mainly by the dictum that taxa should be monophyletic groups. There will therefore be difficulty only when the taxa given by numerical taxonomy are not monophyletic but polyphyletic. We believe that numerical taxonomy will in general give monophyletic taxa because we believe that phenetic groups are usually monophyletic. Before we can criticize numerical taxonomy on this score, therefore, we must have evidence that phenetic groups, created by adequate and acceptable numerical techniques, are not monophyletic. A clear example would be where convergence has occurred to an extent that causes confusion. But we must first show that this degree of convergence has indeed occurred. Figure 5-3 (p. 98) shows that there are several interpretations of what is at first sight a simple problem. There may indeed have been convergence, so that organisms **A** and **B** are more similar phenetically than their ancestors **A'** and **B'**, and **B** is convergent on **A**, though by ancestry related most closely to organism **C**; see Figure 5-3(a). But this conclusion may be uncertain. The evolution may have occurred as in Figure 5-3(b), where there is no convergence but divergence at different rates, with the phyletic line **B** evolving slowly (bradytely). Or, **B** may in fact have descended from **A'** so that **A**, **A'**, and **B** are monophyletic; see Figure 5-3(c). Finally, we may have been mistaken in thinking that **B** was convergent on **A**, for a careful estimate of phenetic affinity may show, taking all their attributes into consideration, that **B** is more similar to **C** than to **A**; here we had been misled by some striking but superficial or restricted set of features, as in Figure 5-3(d). It is clear that from a consideration of the organisms **A**, **B**, and **C** alone (without the evidence of the fossil forms **A'**, **B'**, **C'**, and **X**), we cannot distinguish between these alternatives except to recognize the last of them. Indeed, we have as yet no acceptable evidence that convergence of this kind—that is, overall convergence or convergence in phenetic resemblance—does take place to any marked extent. If it did, it would be exceedingly difficult to prove, for we would have to have an excellent series of fossils to be certain that we had not made any mistake in reconstructing the phylogenetic tree. The known examples of convergence are all open to the objection that relatively very few characters are affected.

Discrepancies between phenetic and phyletic taxonomies can also occur without convergence—for example by divergence at different rates, as in Figure 5-3(b), where **A** and **B** are more similar phenetically than either is to **C**, although cladistically they are equally related.

Even if we grant that overall convergence can occur (as, for example,

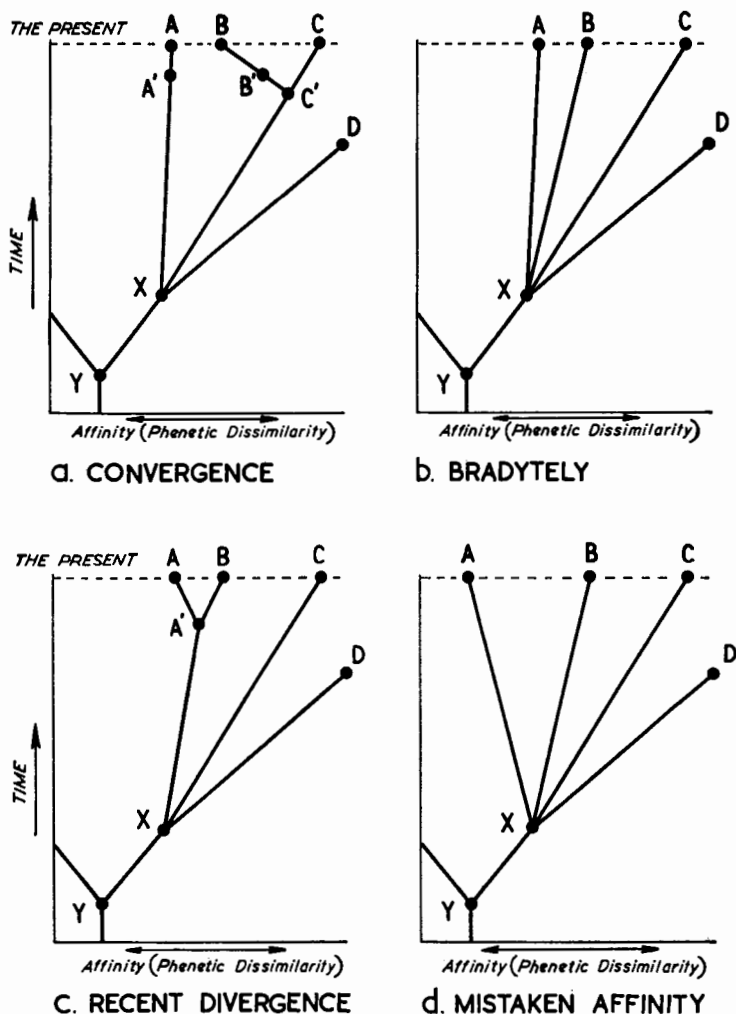


FIGURE 5-3

Alternative interpretations of a case of apparent convergence (for explanation, see text).

has been suggested for some groups of birds), we must ask ourselves why we should wish to make taxonomies based on monophyletic groups. Suppose the convergence had become so absurdly extreme that the two forms are almost indistinguishable and can readily and successfully hybridize: what is the purpose of separating them on grounds of ancestry when in all other attributes they are virtually the same? The purpose cannot be to emphasize minor dissimilarities, nor to serve as a guide to

their behavior with respect to genetic or any other class of properties. The purpose is presumably, therefore, to show that this convergence had occurred, a fact which could be expressed in simple terms without any need for setting up the whole apparatus of formal systematics, and, as is discussed above, the classification would thus be a special one for this special purpose and not a general classification. Where independent phyletic lines fuse into one, the whole problem becomes thoroughly confused, whether we know this has happened or not, since there are so many alternative ways of dividing the network of phyletic lines. This is a common occurrence in plants through the mechanism of allopolyploidy.

Simpson (1961, p. 120) has pointed out quite correctly that most definitions of monophyly are nonoperational because "they are so vague that they provide insufficient criteria for separating one from the other [monophyly from polyphyly] by analysis of evidence." We can ignore the naive statements of earlier authors defining monophyly as descent from a single pair of progenitors, statements made in ignorance of contributions of modern evolutionary theory. Hennig (1950, p. 307) defines as monophyletic groups of species which in the final analysis can be referred back to a common ancestral species (*Stammart*). He points out the sometimes neglected fact that the monophyletic group has to include *all* species derived from this ancestral species, not only those which the taxonomist wishes to classify. As Simpson correctly states, the problem with a definition of this sort is that it is difficult to know how far back one has to trace separate stems in order to arrive at the common stem form. For example, by Hennig's definition the mammals could be made monophyletic only by tracing them back to an unknown early reptile stem.

Remane (1956) solves the monophyly-polyphyly problem more drastically by not accepting as natural any groups with characters which do not conform to his criteria of homology. Thus, by not recognizing polythetic taxa, he decreases the probability of a taxonomist's having to recognize polyphyletic groups. Remane nevertheless does have to consider cases of reputed polyphyly—for example, the well-known case of the origin of the mammals, in which he attempts to show that polyphyly (or polygony, as he prefers to call this phenomenon) is not present. However, the evidence considered by him is rather outdated, and the interested reader is referred to newer papers (see Olson, 1959; Reed, 1960; Van Valen, 1960; Simpson, 1960). Indeed, in numerical taxonomy this problem is of no real consequence, except so far as it may lead to interesting conclusions about the parallel evolution of separate

lines of mammal-like reptiles. To admit the existence of polyphyly would be quite fatal to Remane's system, since he relies upon a closed circle of reasoning from monophyly to a natural system, to homologous structures, and back to monophyly.

Bigelow (1956) has pointed out that in all supposedly monophyletic classifications overall similarities and differences are usually not disregarded. Even in those cases where the ideal of a monophyletic classification could be attained, it often is disregarded in favor of a phenetic classification by supposedly phylogenetic taxonomists. Bigelow feels that "if classification is to correspond with evolution, it must be based on the extent of overall difference, not on time."

The redefinition of monophyly (Simpson, 1960, 1961) is not free from ambiguity in its practical application: "Monophyly is the derivation of a taxon through one or more lineages (temporal successions of ancestral-descendant populations) from one immediately ancestral taxon of the same or lower rank" (Simpson, 1961, p. 124). We therefore need to know what is meant by the term taxon in each instance, and we need to decide the relative rank of the taxa. The difficulties can be illustrated by Figure 5-4. Presumably the descendant taxon, **X**, can be a phenetic group (of a rank indicated by its diversity *A*), and need not be itself a strictly monophyletic group, or the amended definition is superfluous. If the ancestral population, **Y**, is also a phenetic taxon and not a strictly monophyletic group, it can itself be a phenetic group (of rank indicated by its diversity, *B*). This would allow it to be composed of convergent lineages. If such a condition were not permitted (and possibly this is what is implied by "one immediately ancestral taxon"), the term taxon would be used in two different senses in the definition. It is improbable that the lineages ancestral to taxon **Y** would not have had some side branches, and if these had a phenetic rank *C* less than *A* (taxon **Z** in the figure), then taxon **X** would be monophyletic with respect to rank *A* under the amended definition. Yet it would be considered polyphyletic by most workers. There seems little point in redefining monophyly simply to retain for polyphyletic or phenetic taxa the word "monophyletic." One must also specify the rank with respect to which a taxon should be monophyletic. Simpson does not give clear criteria for deciding this rank.

We would not ourselves be concerned directly with the problem of monophyly but would divide the lineages so as to yield phenetic groups as given by the vertical projection of the lines *D* and *E* (Figure 5-4), and taxon **Y** would remain with taxon **X**. This is not because it would put

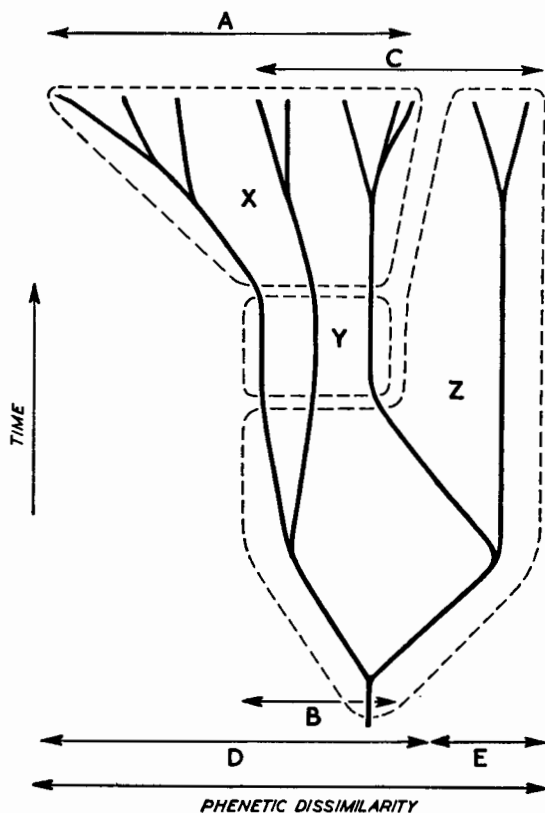


FIGURE 5-4

Diagram of a phylogenetic tree to illustrate the problem of defining monophyly (for explanation, see text).

into different higher ranks some species or genera that were very similar to one another, for this problem occurs wherever one divides the lineages, but because **X** and **Y** are phenetically similar.

Because of the difficulty of obtaining clear decisions on monophyly, Simpson (1961) describes four separate approaches for establishing classifications. The first of these is the empirical approach, advocated in this book but rejected out of hand by Simpson, although much of his discussion of this particular problem (particularly on p. 123) fairly cries out for an empirical solution. The other three approaches follow various types of phylogenetic reasoning, the details of which need not concern us here, in view of the general fallacy of the method. We need only quote from Simpson himself (1961, p. 122): "Choice among them is not a matter of right and wrong but of artistic judgment in each individual case. . . ." It seems to us that such an approach would exclude taxonomy from critical scientific inquiry.

The problem of phenetic *versus* phyletic classification arises also when

considering the problem of vertical *versus* horizontal classifications. By a horizontal classification is meant relationships "among contemporaneous taxa of more or less distant common origin," while a vertical classification is based on relations "among successive taxa in an ancestral-descendant lineage" (Simpson, 1961, p. 129). It is not clear from his account whether vertical relations are entirely phyletic, nor to what degree phenetic considerations enter into horizontal classification. This problem is of most concern to the paleontologist. When based on neontological data entirely, the uncertainties of the relationships involved make a vertical classification indeed risky. Figure 5-5, modified after Simpson (1961), shows the dilemma faced by taxonomists in deciding between vertical and horizontal classifications and the relative simplicity with which a phenetic and empirical taxonomy such as numerical taxonomy would deal with this problem. The problems of grades and clades (Simpson, 1961) can be similarly treated by numerical taxonomy.

We may therefore here give a concise answer to the question: should taxa in orthodox taxonomy be in general monophyletic groups (clades) or phenetic groups? We believe that they should be phenetic groups, for the following reasons.

(1) Phenetic and phyletic relations are taxonomic dimensions formally independent of one another and must be so treated. In the great majority of cases we must deduce the cladal nature of the phenetic taxa by assuming that the phenetic resemblance does reflect the phyletic relations. Since we do not usually have any independent phyletic evidence, there can be no question of disagreement between the phyletic and phenetic data. The occurrence of scattered and incomplete fossil data does not materially alter the situation in such cases, since we have usually to decide whether or not these fossil forms are ancestral by employing largely phenetic criteria. Therefore, in such cases (the great majority) the classifications are *de facto* phenetic ones, and they are phyletic by lip service and wishful thinking only.

(2) We expect that in the great majority of instances the phenetic groups will indeed be clades, though we may not be able to prove this. This follows from our assumption that close phenetic similarity is usually due to close relationship by ancestry. It is only in cases where this appears to be untrue that any difficulty will arise. In such cases we must first verify that the phenetic and phyletic relations are indeed discrepant. This involves comparing affinities from a numerical taxonomic study with a phylogenetic tree obtained from independent fossil evidence (such

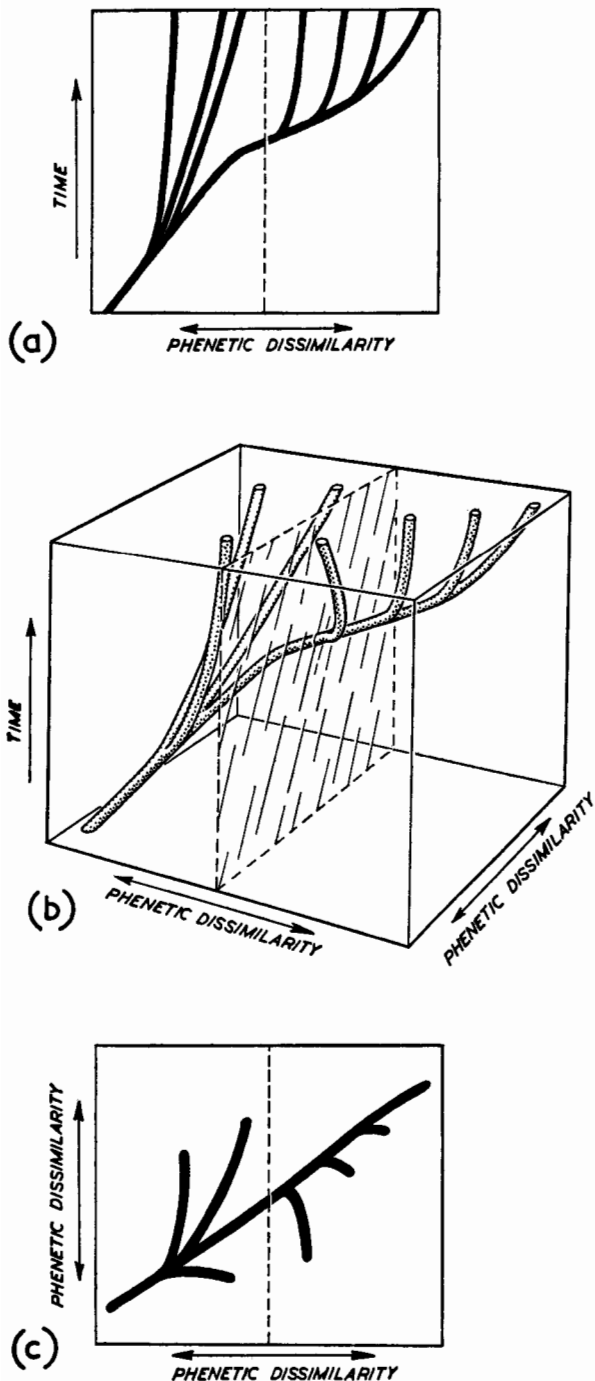


FIGURE 5-5

Horizontal and vertical classification. (a) Diagram of a phylogenetic tree as commonly drawn, with a time dimension and one phenetic dimension. Vertical distances indicate temporal relationships and horizontal distances indicate phenetic relationships. (b) A more realistic representation with an added dimension of phenetic dissimilarity. In (a) is shown a horizontal projection of this three-dimensional tree, but a different horizontal projection might show very different horizontal relationships to that in (a). (c) The projection of (b) onto the base. The relations are now wholly phenetic. The relations between the tips of the branches are relationships among contemporaneous taxa, and hence are horizontal relationships in Simpson's usage.

It should be noted that even two phenetic dimensions are generally inadequate to represent phenetic relationships.

as close sequences of fossil forms in which there can be no reasonable doubt as to the nature and branching of the phyletic lineages). No such studies have yet been made by acceptable numerical taxonomic methods, so the question is to this extent premature. A recent discussion on the evolution and taxonomy of mammals, summarized by Simpson (1960), illustrates this point well.

(3) If (and only if) we have acceptable evidence of disagreement between the phenetic and phyletic classifications, we may discuss which is the better classification to employ. The answer depends on our purpose. If our purpose is to study evolution, there may be advantages in employing clades, though as discussed below we may restrict the scope of our inquiries if we restrict our taxa to cladal taxa. If our classification is for "general purposes," then it is inescapable that we must prefer phenetic taxa to cladal taxa. To decide otherwise would be to imply that an evolutionary classification is a general classification, which by definition it is not. The most important point is that we must be prepared to recognize the disagreement if it exists and to recognize that biological data may admit many interpretations other than evolutionary ones.

To what extent should taxa be phylogenetic entities if the phylogeny is fairly certain? An example of current interest is the classification of the mammals. It is at present generally believed that the mammals are a polyphyletic group. Bigelow (1961), in discussing the implications of this, concludes that a horizontal (phenetic) classification is more satisfactory than a vertical (phyletic) classification. In the former the Mammalia would be a grade, while in the latter they would be a clade. Reed (1960) has advocated a cladal classification as being the more logical but admits that it is less convenient, even for paleontologists. Simpson (1960) prefers to retain the present horizontal classification, but his position is not very clear in view of his advocacy of cladal classification as a preferred general principle of taxonomy (Simpson, 1961, pp. 120–125, 189–191).

The evidence in the case of the Mammalia is itself not clear or easy to interpret. It is generally agreed among paleontologists that a minimum of two groups of reptiles (and possibly four or five) independently crossed the arbitrary line separating reptiles from mammals. The evidence for this (based on jaw and ear structure) is summarized by Simpson (1959) and by Olson (1959). Yet, as Reed (1960) points out, this implies an independent origin of an impressively large number of mammalian characters. This argument is of especial cogency in the case of the monotremes, which are thought by some to be one of the groups

which crossed the line independently, although there is no pertinent fossil evidence of their origin (Simpson, 1945, 1960). The application of monophyly as a criterion would at the least require the inclusion in the mammals of the therapsid mammal-like reptiles, as Van Valen (1960) suggests. A consideration of the phyletic trees given by Olson (1959) and Reed (1960) shows a number of uncertainties about the origin of all the mammalian groups, and the authors cited by them and by Simpson (1959) are evidently not in agreement on the details. It may therefore be premature to discuss their taxonomic position, though it may be noted that the Multituberculata are the group which are hardest to explain on a monophyletic basis of the origin of mammals. The others might plausibly have arisen from a single therapsid stock, and, by excluding the Multituberculata from the mammals, the problem would be avoided (Simpson, 1960). Despite the uncertainty in this instance, there are some well-documented lineages in the evolution of mammals which pose the same problem (for example, see Simpson, 1951, pp. 204, 211, 280), though a critical examination by numerical taxonomic methods remains to be made.

However, it should be emphasized that even if the entire phylogeny of a group of organisms were known so that they could be arrayed in a dendrogram accurately representing descent (that is, who came from whom), we would still have a very limited representation of the organisms. In order to make a reasonable classification, even the most ardent phylogenist would wish to know something about the similarity between the organisms in order to make a meaningful classification. Unless we knew how much evolution had occurred and how this affects the phenotype of the organisms, it would be futile to construct a classification. Thus there is a place for numerical taxonomy even in the rare instances of completely known phylogenies. For this reason the application of numerical taxonomy to paleontological as well as neontological work is indicated (see Sneath, 1961).

5.5.2. Convergence

Convergence and parallelism are terms over which there is considerable confusion (see Haas and Simpson, 1946, for a full discussion). Convergence or parallelism may mean convergence or parallelism in one organ (or in one character complex) or of the entire phenotype. Many authors do not specify which they mean. In this book we use the terms with respect to overall, phenetic similarity, whether the changes are contemporary (isochronous) or not (heterochronous). The occurrence

of convergence restricted to a small part of the phenotype would not, of course, produce convergence in overall similarity; for example, the bats are convergent on the birds with respect to flight, but in the remainder of their phenotype they are divergent from birds, when compared with the common reptilian ancestor.

A comment that is voiced more often than any other criticism of numerical taxonomy is that the frequent occurrence of convergence would confuse and invalidate its results. This question has been considered from the first (Michener and Sokal, 1957), and it is the position of numerical taxonomists that most cases of convergence would affect coefficients of similarity only slightly, since they would not involve more than a small proportion of the characters.

Considered strictly by the criteria of this book, convergence is no problem at all. So long as we are concerned with phenetic affinities the similarity value obtained by numerical methods will be truly representative. It is only when we wish to draw *phylogenetic* conclusions that convergence may confuse the issue. We hope to show in the following sections that even phylogenetic deductions will not be seriously affected by the inclusion of convergent characters in the computation of similarity coefficients.

It is necessary to specify in what respects lines are convergent. It is quite possible for two lines to converge in respect to one organ and to diverge in respect to others. The only kind of convergence which is pertinent to our present argument is that where the lines converge in so many respects (that is, characters) that it causes an increase in the overall similarity of the two lines. This, which can be called "overall convergence" to distinguish it from convergence in a few respects, might cause serious discrepancies between the taxonomy yielded by phenetic methods and that implied by reliable phyletic studies. Convergence in a few respects ("organ convergence") will not do this, since these few respects will have little effect on the affinity values of the many non-convergent attributes included in the analysis. There is of course no sharp line between overall convergence and organ convergence.

The pertinent question, then, is: Does marked overall convergence ever occur? And it is one which urgently needs study. There are many examples in which numerical taxonomy could be readily employed to test this question: *Canis* and the thylacine marsupial wolf, the marsupial and eutherian moles, the seals and sirenians, and some xerophytic or parasitic plants could be compared. We believe that the overall similarity of pairs of this kind is not high. Indeed, if this were not so, it is uncertain

how they were recognized as “convergent pairs” and not just close relatives. The very obvious and striking similarities in appearance will, we believe, account for very few of the total features analyzed. Any reasonably random and unbiased selection would, we think, include far more features which did not show convergence. And even in habitus one finds many differences due to different modes of achieving the same function with diverse anatomical parts. Striking cases of overall convergence where the phylogeny is thoroughly known are not common, and none has yet been examined by acceptable numerical taxonomic methods.

Sneath (1961) has pointed out the ridiculous implications of total overall convergence in higher animals, though it may not be so ridiculous in viruses (see Section 8.1.2). It is perhaps worth noting that when considering many characters there is every expectation that evolutionary processes will *overall* be divergent. This is a consequence of variation, which has a strong element of randomness. In order to obtain convergence, the possible kinds of variation must be restricted; that is, there must be more change in differential features than in common features. We have no reason to believe that, in general, natural selection will have this effect (except possibly in situations like that described below), since it will act largely upon random mutations and in many different directions.

What we have said above applies to the higher levels of taxonomic rank. At lower levels, at the genus and below, there is a possibility of some degree of overall convergence. For example, the introduction of a new food plant into an archipelago possessing several island races of a fruit-eating bird (which had diverged slightly from one another over the course of time) might produce such a strong selection pressure—in the same direction and in all the islands simultaneously—that these races would rapidly evolve toward adaptation to feeding on the new plant, and this might outweigh the slow accumulation of genetic differences which had been continuously occurring in each race. The overall similarity between these races might then increase somewhat, and there would then have been some degree of overall convergence. In the absence of knowledge of the past, it is difficult to see how any systematic procedure would elucidate the case, and numerical taxonomy is in no worse position than others. It is possible that convergence on this level and of this degree may be frequent but undetectable, at least by any of the procedures known today.

Finally, we may emphasize an obvious but often forgotten point. If we

do indeed wish to study convergence, we can only do this by comparing a phylogenetic arrangement with a phenetic arrangement. In no other way can we detect the process of convergence, and any attempt to restrict taxonomy to phylogenetic arrangements will then defeat its own ends.

5.5.3. Parallel evolution

Parallel evolution occurs when two genetically isolated stocks evolve so as to maintain constant the difference in those attributes which are under consideration. Parallel evolution seems to us to be similar to convergence and subject to analogous reservations of definition, but to a lesser degree. The same problems and dilemmas arise, and again there is very little clear evidence in favor of extensive parallel evolution if *all* the features are included in the taxonomic analysis. Possibly the best examples of what may be *overall* parallel evolution (the inclusion of the word "overall" implies a constancy in the overall similarity, as with convergence discussed above) are certain ferns (Holttum, 1949) and certain ammonites. The apparently parallel trends in the degree of convolution and ornamentation of the shells of ammonites represent very few characters out of the many present during life. In addition to this, these characters are likely to be selected by the environment in the same way; for example, it is possible that certain forms of the shell may have protected many different species of ammonite from a particular predator. Another example of this may be the repeated evolution of increased curvature of the shell in lineages of oysters, leading to the *Gryphea* phenotype, which has been interpreted as a recurrent adaptation to a muddy sea bottom (discussed by Joysey, 1959). This may well have involved only a few characters, since we do not know what changes occurred in the soft parts of these molluscs.

The reviews of Trueman (1930), Swinnerton (1932), George (1933), and Joysey (1959) may be consulted for some of the better-known instances of this phenomenon; these authors all emphasize the difficulty in deciding whether the apparently parallel lineages are indeed independent phyletic lines, or whether the forms found in any given stratum should be grouped together in a monophyletic taxon. If the latter procedure is correct, then parallel evolution is simulated by similar adaptive radiations in successive taxa (stages) of an evolutionary line.

As with convergence, we feel that most apparent examples of parallel evolution are due to parallel trends in a few characters. Again we

urgently need quantitative studies of this problem. Even if overall parallel evolution does occur, it will be no easy matter to prove convincingly the validity of the phyletic lineages concerned, and the phenomenon will probably be of small degree.

Readers of the above two sections may feel at variance with our ideas because our definitions of convergence and parallelism, while not alien to the usage of systematists, are probably not the most common ones; hence our comments may be thought not relevant to the central ideas of convergence and parallelism as customarily conceived. While the terms convergence and parallelism permeate evolutionary literature, it is difficult to find definitions for them. Remane (1956), for example, uses but does not define convergence and parallelism. Rensch (1947) considers convergence to be simply nonparallel evolution. Simpson (1961, p. 78) has attempted to coin precise definitions for these terms. He defines parallelism as "the development of similar characters separately in two or more lineages of common ancestry and on the basis of, or channelled by, characteristics of that ancestry." Convergence he defines as "the development of similar characters separately in two or more lineages without a common ancestry pertinent to the similarity, but involving adaptations to similar ecological status." The inferences that are customarily made from such a definition, and that are also discussed by Simpson (1961, pp. 103 ff.), are that convergence occurs between forms that are relatively far apart, while parallelism occurs only among lines that are relatively closely related. Simpson mentions the occasional literal interpretation of convergence as the narrowing of differences between lines and the parallel change of these lines as constituting parallelism, but he does not consider that to be a particularly useful taxonomic concept.

It seems to us, on the contrary, that in any taxonomy based on a phenetic system (in fact any taxonomy that wishes to base itself on measurable quantities), the definition of parallelism and convergence should be entirely based on the parallel or convergent nature of the differences between the lines. We hold these beliefs for several reasons.

(1) It seems to us that no fundamental and useful distinction can be established between convergence and parallelism. To say that convergence takes place only between distantly related forms would prejudge the issue of relationship completely. We would no longer be able to use cases of convergence or parallelism in our classificatory schemes because relationships would have been predetermined before any deci-

sion could be taken on whether a convergence or parallelism is at hand.

(2) The definitions of Simpson (1961) and his subsequent discussions lead one to believe that by convergence is meant the construction of similar structures based on different genetic systems, while by parallelism is meant the construction of similar structures based on similar genetic systems. By inference one may conclude that different genetic systems occur in more distantly related organisms and similar genetic systems in more closely related organisms. This concept is, however, partly erroneous and of little utility. It is erroneous, because it has been shown in genetics, theoretically as well as by experimental evidence, that presumably identical populations when faced with an environmental challenge will not necessarily meet it by identical responses. For example, the challenge of insecticides to insect populations has been met by the same species in a number of different ways (Crow, 1957). Therefore it is not necessarily true that similar genetic systems will produce similar responses to a given environmental challenge. The converse point is granted, namely, that it is not likely that distantly related lines will develop similar genetic systems in response to similar environmental challenges.

(3) In the absence of knowledge about genetic homologies, we are faced in the main with phenetic changes which may be divergent, parallel, or convergent. It would therefore appear that the only useful distinction is whether changes are parallel or convergent, or better still whether there is relatively more or less divergence, since often less divergence between some members of two taxa than between the majority of the members of the taxa may be considered as convergence or at least parallelism. Of course, *sensu strictu*, this can only be done if fossil series of organisms are available. Where only living end points of evolutionary change are known, parallelism or convergence cannot really be demonstrated but must be inferred from recent evidence.

5.6. THE REQUISITE NUMBER OF CHARACTERS

A problem of immediate and urgent importance in any numerical taxonomic study is the number of characters required to obtain reliable results. No general answers can as yet be given; however, there are some general considerations which must form the basis of this problem and which may be useful in guiding attempts at a solution. It is obvious that the problem of the number of characters relates to the hypothesis of nonspecificity. If the hypothesis holds rigorously, sampling a relatively

small number of characters should still give an adequate idea of the classification. But if the hypothesis does not hold well in a given group, and separate characters exist in separate regions of the body representing different factors, then a relatively larger sample of characters must be employed. Two conceptual approaches were employed in obtaining an estimate of the number of characters required. The first, the hypothesis of the factor asymptote, did not prove fruitful; however, it is discussed here to give some insight into the reasoning we used. The reasoning eventually employed—the hypothesis of the matches asymptote—is given in the subsequent section. Any discussion of characters and the number necessary for the estimation of resemblance is inextricably linked to the question of what the taxonomic units are whose affinities are being investigated. We shall defer a consideration of this problem until Section 5.8. In the discussion that follows below we shall use the term operational taxonomic unit (OTU) for the unit that is to be classified without defining it further here.

5.6.1. The hypothesis of the factor asymptote

The following three statements are based on the nexus and non-specificity hypotheses. We label them collectively the hypothesis of the factor asymptote. This hypothesis concerns itself with information obtained about a single organism or operational taxonomic unit by studying its characters. First of all, it appears obvious that the more characters we study, the more information we will accumulate. We deliberately use “information” in a general, ill-defined sense here. For a recent discussion of the information content of biological classifications from the point of view of information theory, see Rescigno and Maccacaro (1960).

Second, a random sample of the characters should represent a random sample of the genome of the operational taxonomic unit. Our use of the term “random” is not in the strict probabilistic sense but rather in the sense of being haphazard or unbiased.

Third, as we include more and more characters, the gain in information with each additional character, while large at first, should decrease fairly rapidly and, after a certain number (n_a) characters have been recorded, should approach close enough to an asymptotic value to make further inclusion of characters in the study unprofitable. This last aspect of the hypothesis is most important for numerical taxonomy, since it would permit placing a limit to a theoretically boundless task. It is

important to note whether the rate of gain of information slows down considerably before an appreciable proportion of the genetic factors of the operational taxonomic units have been described.

A simple model may help us to understand the process, although an even mildly realistic interpretation of the situation would have to be much more elaborate. Imagine an organism with 10,000 gene loci. Let us assume that each character yields information on 12 loci and that random choice of characters is therefore equivalent to random samples (with replacement) of 12 loci each time from the total of 10,000. The first character would present us with information on 12 loci. The

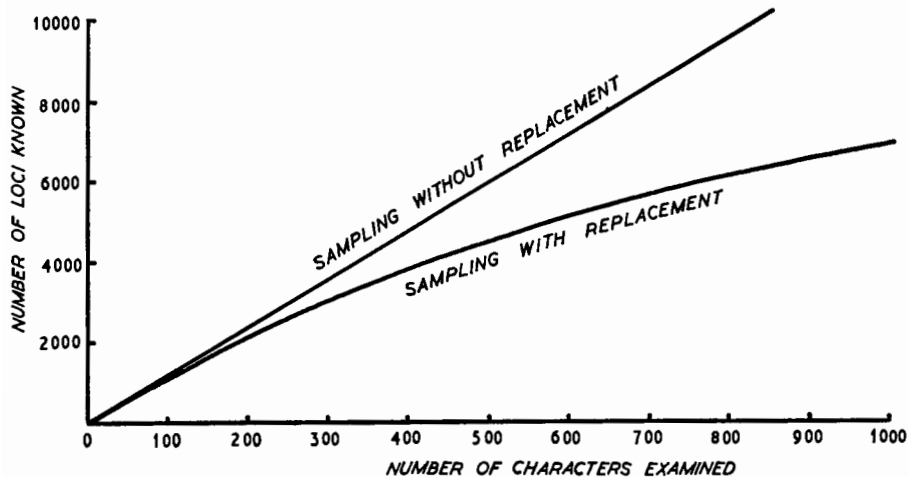


FIGURE 5-6

The expected number of gene loci on which information will be given if each character provides information about 12 loci (for explanation, see text).

probability that the next character will contain 12 *new* loci is extremely close to 1 (approximately 0.98567). Using an approximate method sufficiently accurate for our purposes, we can calculate the expected proportion of genes which will have been sampled by the formula $1 - (0.9938)^n$. In Figure 5-6 we see the results of such a computation graphed for studies involving 1 through 1000 characters, each assumed to be a sample of 12 independently occurring gene loci. If each character yielded information about different gene loci, the expected number of genes known would be a simple linear function of the number of characters studied. In view of the possibility that different characters will represent previously sampled genes, the expected number of genes known

on the basis of a given number of characters is always less than the maximum possible number. While it is clear that the rate of actual recovery of new information (or new gene loci) is slowing down, loci are still being recovered at a fairly rapid rate after 200 characters have been studied. Thus not only do we not reach an asymptote of information recovery by the time an appreciable number of characters has been measured, but using what appears to us to be a reasonably realistic model we find that, after measuring 200 characters, information on only a little over one-fifth of the gene loci has been obtained.

To obtain information on 50% of the genes (5000 loci) we would have to study 579 characters, a procedure which would only be approximately 72% efficient, since that number of characters could yield a maximum of 6948 loci, if each sampled different loci.

A more realistic approach would be to permit the number of genes which a character represents to vary according to a plausible distribution, perhaps in Poisson fashion. While this would widen the confidence band of the graph, it would still not alter the general relations. Further realism could be added by considering the correlations among some of the characters sampled. This would reduce the rate at which the genome of the organism would be sampled.

The interesting conclusion from this model is that we are unlikely to learn about an appreciable proportion of the genotype (say >15%) unless the average number of major gene effects on each character is considerably higher than 50 or unless an exceedingly large number of characters is studied. Thus a character survey of a taxon would fall short of the aim of obtaining a genetic characterization of it unless an impractically large number of characters is used.

A fundamental question may be raised at this point. If one cannot obtain adequate coverage of all genetic factors of a taxon from a reasonable number of characters (this was apparently not realized by any of the early writers on the subject), does this invalidate the practice of numerical taxonomy? The answer would appear to be "no," for two reasons. First, from the point of view of recognition and discrimination of taxa, the number of character states available when even a moderate number (say 100) characters are measured is far greater than is necessary. Thus, while our character sample may not be adequate to reveal the complete properties of a taxon they are likely to be quite adequate for its description and discrimination.

The second reason is that the coefficients of similarity on which numerical taxonomy is based do not rest directly on the absolute

magnitudes and qualitative nature of the characters sampled but on the relative degrees of difference between the characters of the operational taxonomic units being compared, as will be shown below. This point of view leads us directly into the next hypothesis.

5.6.2. The hypothesis of the matches asymptote

When we compare two operational taxonomic units on the basis of a sample of their characters, we can express their resemblance as a proportion of characters agreeing (matching) out of the total number being compared and assume that the similarity between the two operational taxonomic units is some parametric proportion of character matches which we are estimating with a sample of characters. In other words, we are taking a random sample from a very large number of characters which we could in theory sample and which would yield us a single, definite proportion of matches if we were able to sample all the characters. This might, for example, be the matches in the nucleotide sequence of the DNA of the genotype. The hypothesis, then, simply assumes that, as the number of characters sampled increases, the value of the similarity coefficient becomes more stable; eventually a further increase in the number of characters is not warranted by the corresponding mild decrease in the width of the confidence band of the coefficient.

The justification of this hypothesis rests on ordinary sampling theory in statistics and needs no special defense, if we can conceptualize a parametric value of matches among all possible characters and if character sampling as actually practiced can be considered close enough to being random in a statistical sense. The fact that in correlational and distance techniques the magnitude of mismatches between characters is being taken into account should not affect the hypothesis, since the only consequence will be the automatic and quite appropriate weighting of mismatches by their magnitude. Suppose we think of the comparison of each character in two taxa as a sample of, say, 12 loci being matched, where comparison between the entire genomes of two taxa would yield a proportion p of mismatches and $1 - p$ matches. Then a random sample of characters should provide an estimate of the value of p based on simple binomial theory of sampling with replacement.

The hypothesis of a matches asymptote can be tested as follows. For a group classified on an adequate number of characters we may, by a systematic search, uncover another group of characters. If the hypothesis

holds, the changes brought about in the matrix of similarity coefficients by the inclusion of the new characters should be no more than expected by sampling theory. We have some misgivings on this score. Sneath (1957b) has pointed out that a change of emphasis or technique on the part of the investigator may lead him, perhaps unconsciously, to look for matching characters in preference to nonmatching ones (or vice versa) and thus introduce bias.

5.6.3. How many characters?

We have as yet no unequivocal answer to this direct question. In the published literature on numerical taxonomy only Michener and Sokal (1957) have ventured to make a suggestion—not less than sixty characters. Their idea was based on the statistical consideration that the confidence limits of the correlation coefficients became too wide below that sample size. As we shall see, there are indeed two separate considerations involved.

First of all, there is the purely statistical problem of the reliability of the estimates, if we regard similarity as a parametric value which we are sampling. Figure 5-7 shows the changes in the 95% confidence band for two correlation coefficients as sample size (number of characters) increases. If association coefficients are considered simply as a proportion of matched character states out of a total of all possible matches, we can apply the standard error of the binomial to them ($s = \sqrt{pq/n}$; if we

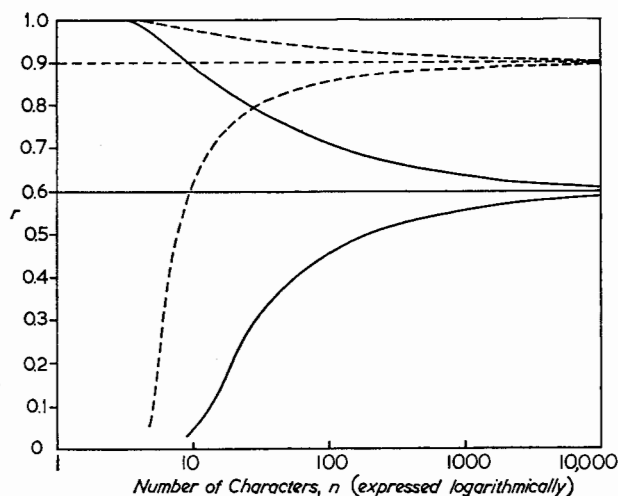


FIGURE 5-7

The 95% confidence limits for two correlation coefficients, $r = 0.9$ and $r = 0.6$, plotted against the number of characters, n , on which the coefficients are based.

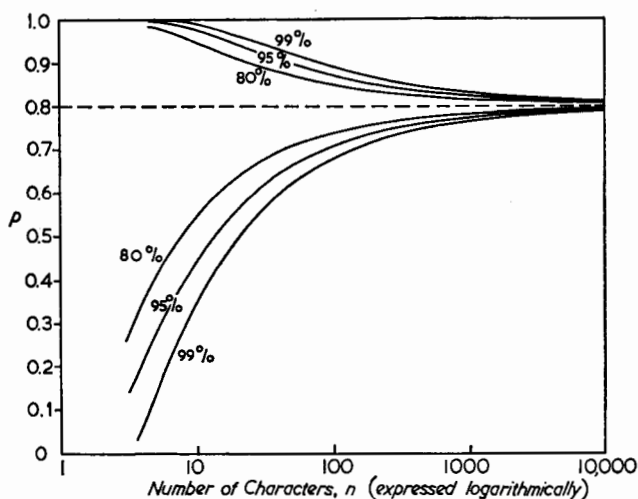


FIGURE 5-8

The 80%, 95%, and 99% confidence limits of a proportion, such as the coefficient of association, S_{SM} , at a value of 0.8 plotted against the number of characters used to obtain the coefficient.

assume the maximum error at $p = q = .5$, then $s = \sqrt{.25/n}$. Where p is close to zero or unity and n is relatively small, the binomial probability distribution $(p + q)^n$ should be expanded to obtain expected confidence limits for a given probability p . Figure A-4 in the Appendix provides a handy graph for so doing. Figure 5-8 shows the 80%, 95%, and 99% confidence bands for an association coefficient of 0.8. Figure 5-9 gives the expected value and 95% confidence band of a distance coefficient calculated as explained in Section 6.2.3.2.

We can see that as sample size increases the rate of decrease of the confidence bands becomes greatly reduced in all three instances. Thus the narrowing of the bands due to the increase of sample size from 50 to 100 is considerably greater than the subsequent reduction of the band between sample sizes 100 and 200. Therefore, after approximately $n = 100$, gains become unappreciable in general.

The above values should give some guide lines to the required sample size. The allowable width of the confidence band depends partly on the refinement with which one wishes to analyze the similarity matrix. If a rough classification is adequate, wide limits are permitted. But if fine structure is investigated and a detailed, nested classification is desired, the limits have to be sufficiently narrow that sampling error is not likely to be mistaken for real differences in affinity.

The second consideration is to think of each character as a sample of the genome and to inquire when a required percentage of the genome has been sampled at a required confidence level. We have seen in our discussion of the hypothesis of the factor asymptote that this approach is not very fruitful. Under plausible assumptions any reasonable number of characters would not represent more than a relatively small percentage of the genome. Thus we cannot use this consideration as a criterion for the necessary number.

Strictly speaking, the two ways of considering the effects of sample size would in any case be incompatible, since the required heterogeneity of the samples on which the second estimate is based invalidates the assumptions on whose basis the confidence bands have been computed. At best we can only claim the estimates of this section as indicators of the trends to be expected. As experience with numerical taxonomy of various groups begins to accumulate, we may learn empirically how an increase in the number of characters changes coefficients of similarity.

As a general rule we may stress that it is better to employ more, rather than fewer, characters. Certainly this should be the case whenever data on the characters are already available or can be easily obtained. By present methods of computation, limitations are generally set on the

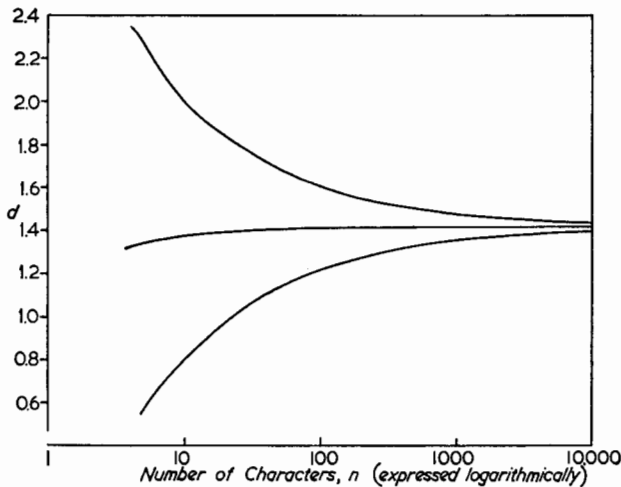


FIGURE 5-9

Expected value of d and 95% confidence limits of the distance coefficient on the assumption that the observations on which the operational taxonomic units are based (that is, the n standardized characters used) are independent and normally distributed with a mean of zero and unit variance.

number of operational taxonomic units rather than the number of characters, since computer operating time increases in general as the square of the number of taxa, but only in direct proportion to the number of characters.

5.7. THE PROBLEM OF WEIGHTING

Numerical taxonomy gives each feature equal weight when creating taxonomic groups (Sections 4.3, 4.4). This is unorthodox, but we will discuss its justification. We should emphasize that we are not here discussing the use of characters in identification. After a manner of speaking, "weighting" is used in such a procedure, and properly so. However, the construction of taxonomic keys and the identification of specimens belong to a late stage of taxonomic procedure and must not enter into the formation of the taxa concerned.

That every feature should be given equal weight is implicit in the work of Adanson and the writings of Gilmour (1937, 1940, 1951) and Cain and Harrison (1958), and it was stated explicitly by Sneath (1957a) and Michener and Sokal (1957). Verheyen (1960) holds the same opinion. Sneath reached this conclusion on considerations stemming from Gilmour's work on epistemology. It was argued that since natural taxa ideally contain the greatest possible content of implied information, this can only be measured in the number of statements which can be made about its members, which is independent of how important we may think any statement is. This argument has been developed at some length in connection with the "general" nature of natural classifications (see Sneath, 1958).

Michener and Sokal (1957) concluded that, even if desirable, there is no rational way of allocating weight to features, and therefore one must in practice give them all equal weight. Even if the entire genetic constitution of a form were known, it would be impossible to find a basis for weighting the genetic units, for these have no fixed adaptive, ontogenetic, or evolutionary significance. To mention only one possible criterion, the significance of the genetic units depends on their environment, which is always changing. In addition, when many characters are employed, the statistical analysis of similarity is only slightly affected by weighting some characters (unless this weighting is extreme). It is perhaps worth re-emphasizing this point, for if in practice the measures of overall similarity when based on many characters do yield substantially

the same results, whatever the weighting (within reason) of individual characters, it would seem unnecessary to argue the point further.

The arguments in favor of equal weighting fall under seven headings.

(1) If we cannot decide how to weight the features, we must give them equal weight—unless we propose to allocate weight on irrational grounds.

(2) If we are to create taxonomic groups we must first decide how to weight the features which are to be employed for classification. We can therefore use no criterion which presupposes the existence of these taxa. For example, we cannot choose the constant features—to know if they *are* constant we must first set up taxonomic groups, and these we do not yet have. This was implied by Adanson's arguments, when he correctly asserted that one cannot make a priori judgments of the importance of characters.

(3) The concept of taxonomic importance has no exact meaning. If "importance" means "importance to me because I am interested in it," this is only special pleading. If "importance" means basic or fundamental, this can only mean that it sums up a number of other characters: if they are unknown, they are hypothetical; if known, the character is not single but multiple. If "importance" means essential to survival, the taxonomy can estimate viability, but not resemblance. If "importance" means "correlation with other features," then the added weight is due to these other features; where we observe the correlation to break down we do not regard the feature as important.

(4) If we admit differential weighting, we must give exact rules for estimating it. We must know whether the weight to be given to the possession of feathers is twice or twenty or two hundred times that given to possession of claws, and why. We do not know of any method for estimating this, and even if such a method were to be developed we doubt if any systematist would have the patience to use it because of the hundreds of characters he would need.

(5) The nature of a taxonomy depends upon its purpose: we could arrange living creatures in many ways, but we choose one way because we think it is the best for some purpose. We hold the view that a "natural" or "orthodox" taxonomy is a general arrangement intended for general use by all kinds of scientists (Gilmour, 1937; Sneath, 1958). It cannot therefore give greater weight to features of one sort, or it ceases to be a general arrangement. It can only be made if scientists are

willing to forego the claim that their own interests should be specially favored. It is not always clear to them that it is in their own interests to do this, but they need this general arrangement against which to compare their own special findings. Being general, it is best for general purposes but is perfect for none.

(6) The property of "naturalness" is, we believe, due to the high content of implied information which is possessed by a natural group. A group such as the Mammalia at once tells us much about its members with a high degree of certainty. A group such as "black animals" tells us nothing more than that they are all black. The content of information is measured by the number of statements which can be made about its members: each statement has unit value, and whether we think them important or not is irrelevant.

(7) As noted above and in Chapter 6, the use of many characters greatly evens out the effective weight which each character contributes to the affinity coefficient. In practice, therefore, the methods are equal-weight methods.

Equal weighting can therefore be defended on several independent grounds: it is the only practical solution, it and only it can give the sort of natural taxonomy which we want, and it will appear automatically during the mathematical manipulations. Singly, these arguments are cogent; taken together, we feel that they are overwhelming.

5.8. THE OPERATIONAL TAXONOMIC UNITS

What taxonomic units can be classified by numerical taxonomy? The logical fundamental unit in a large majority of instances is the individual organism. While this is usually an unambiguous entity, it is generally not feasible to use numerous conspecific individuals of each of several apparent taxonomic groups to compute a classificatory hierarchy. Matrices of excessive size would have to be processed—a formidable enterprise even by the standards of the most modern computers. Furthermore, such studies would throw light largely on resemblances among intraspecific variants and would not be likely to offer much scope for comparisons at the subgeneric, generic, and higher levels. Thus, except for special studies aiming at intraspecific classification, the most customary unit in zoology and botany will be the species (strictly speaking, the taxonomic unit with a binominal name, the *binom* of Camp, 1951, which is believed to correspond to one or other of the biological units which

are given the name of species). Since the hierarchic level of the taxonomic unit employed in numerical studies will differ, we cannot speak of fundamental taxonomic units but shall refer to them as *operational taxonomic units* (OTU's).

Should numerical taxonomy rely on the validity of species erected by conventional methods? We believe (probably in agreement with most present-day biologists) that of the categories established in the present system of nature the different entities which have been called species have more reliability than any others, with the possible exception of the very highest taxa. So far as numerical taxonomy is concerned, it does not matter that there are many species concepts and species definitions, so long as these are clearly understood and used fairly consistently within a study. Some kinds of species, furthermore, can be defined by biological (analytical and experimental) criteria (see Mayr, Linsley, and Usinger, 1953, for some definitions) and differ in this respect from the higher and lower taxa, which are defined on more ambiguous bases. However, it should be pointed out that the criteria of specific difference actually applied by most systematists to most taxa are phenetic and do not satisfy the genetic definition of specific status (see also, Blackwelder, 1962, and Sokal, 1962b, 1963). Although taxonomists make objective tests of the criteria for specific distinction in only a small minority of cases, their judgment in this matter generally inspires confidence (certain difficult taxa and indifferent taxonomists excepted).

Zoologists in recent years have turned to a genetic definition of species (*sensu* Mayr, 1942). However, there are many groups—including, for example, the bacteria—where such a species definition seems at present inapplicable. Since a single bacterium can seldom be studied sufficiently, taxonomy is based on pure cultures of strains. In plants there are many genera in which the individuals are apomictic—vegetatively propagated, parthenogenetic, or self-fertilized—and here it is generally the clone which is taken as the unit of classification. Yet clones may throw off mutant forms, which may or may not be considered as distinct units. The important point is to avoid prejudice in choosing the OTU's and if need be to explore by preliminary analyses the phenetic relations of the specimens which are to form them. In both plants and animals there may be a choice between stages in the life cycle, or between different castes in social insects. The special problems raised by these possibilities have been discussed in Section 5.4.3.

Problems may arise if a taxon used as an OTU proves to be variable for one or more characters. This brings up the question of whether we

can use as OTU's higher taxa (such as genera, families, and orders) in which the majority of characters within a taxon will of course vary. Such taxa can be used, in principle, for the reasons discussed in Section 6.4.

A second equally serious problem is the low degree of relevance of most lists of characters (see Section 6.5.5). Since these considerations have not yet been discussed, we shall postpone an evaluation of the analysis of higher taxa until Chapter 6.