next few months, for once the Atlas is in use we shall be inundated with questions about the origin of records.

Except for the addition of other symbols the maps are ready for the printers when they come off the machine. The original maps are 14 inches x 12 inches and when reduced to a size of 5 inches x 4,5 inches the black dots made by a heavily inked ribbon are quite satisfactory. The maps have been printed by offset lithography and if, in proof, there are any minor flaws, these can be adjusted on the negative.

One last point which is not strictly data processing but data analysing — we have included in the Atlas 12 transparent overlays, as instructed by the 1950 conference. In our tight little islands, geology and topography change so rapidly that we found it impossible to produce conventional maps on the same scale. Here again we used the 10-kilometre square as a unit and showed by dots and other symbols the distribution of squares having outcrops of limestone, with which the distribution of species can be correlated either in a positive or negative way, and on squares having land over certain altitudes, which gives a reasonable understanding of the distribution of some of our mountain plants.

Since the publication of the Atlas it is perhaps not suprising that British field botanists have given up square-bashing and have been taking a well earned rest playing noughts and crosses.

THE PRINCIPLES AND PRACTICE OF NUMERICAL TAXONOMY *

Robert R. Sokal (Lawrence, Kansas)

The foregoing discussions on electronic data processing will surely have impressed you with the immense potential of these methods for botanical research. The exhilarating prospect is that these methods are at present only in their infancy. Every year brings new and exciting adaptations of existing equipment, and the development of new equipment, to handle tasks which previously were considered drudging chores, advancing at a snail's pace, not only because of inefficient methodologies but also because scientists generally shun such activities and relegate them to a low priority in their plans of work. Tasks routinely carried out today were impossible just a few years ago. In newspapers, magazines and professional journals we constantly hear about the present period of scientific revolution in which we live, so that the term has become a commonplace. In spite of being acclimated to such pronouncements, the individual scientist lifting his eyes for a moment from his desk or work bench and looking about him must stand in awe of the changes in scientific outlook and methodology which have taken place just in the last ten years. As I look at my own field, the application of statistics to biological problems, technological advances in carrying out statistical computations have been of such a nature as to add entire new dimensions to biometric work. Electronic data processing has made it possible for me in just a few years to carry out computations which would have taken a conventional desk calculator operator several lifetimes.

* Contribution No. 1192 from the Department of Entomology, The University of Kansas, Lawrence, Kansas. Preparation of this manuscript was supported by a grant G-21011 from the National Science Foundation.

Presented in a symposium entitled, "The Application of Data-processing Methods to Research in the Biological Sciences", sponsored by the American Society of Plant Taxonomists, the Botanical Society of America, and the Ecological Society of America; Corvallis, Oregon, August 29, 1962.

How is taxonomy to benefit from these developments? Before we answer this question we should perhaps ask a legitimate counterquestion: Does taxonomy need improvement? I believe that the vast majority of objective observers would have to reply with a resounding "yes". Many of the procedures of taxonomy are not much different today than they were at the time of Linnaeus. The forward thrust of biological sciences in this century has largely bypassed the philosophy and practices of taxonomists. To this sweeping indictment there are several exceptions and qualifications. It will be argued that the acceptance of the Darwinian theory of evolution completely revolutionized taxonomy. This is a controversial point. It has been cogently argued by some authors (e.g. Remane, 1956) that changes in post-Darwinian taxonomy as compared with pre-Darwinian practices are largely changes in form and terminology rather than changes in content. On the other hand, the so-called New Systematics has helped us to understand evolution rather than improve taxonomy.

It is also true that a great many new kinds of data have entered taxonomic work in recent years. The physiological, biochemical and behavorial sciences are providing us with ever more data to be evaluated and digested for erecting classifications. In botany striking advances have been made through cytological research, but this research has been primarily at the lower taxonomic ranks and is therefore generally not very helpful at the generic and higher levels.

Before we can discuss how progress in EDP can aid in taxonomic research we shall have to agree on the functions of taxonomy. The terms taxonomy and systematics are frequently used interchangeably. I like to distinguish between them and follow the convention generally adopted, at least by animal biologists, in recent years (as described for example by Simpson, 1961). *Systematics* may be considered as the scientific study of organic diversity, including the description of the organisms, their history and phylogeny, when known, and the study of the evolutionary mechanisms bringing about such diversity. Systematics therefore is a very general term. *Classification* is the ordering of organisms into groups on the basis of their relationships by descent or similarity or by both. *Taxonomy* is the science of classification involving both the theory and practice of classification. In this sense then, biosystematics as commonly practiced in botany is a branch of systematics, but only indirectly related to taxonomy.

Taxonomic theoreticians in recent years can be generally classified into two schools. Those who use reasoning about phylogenetic origins of biological characters in the construction of the classification and those who would use only evidence obtained from comparative study of the specimens without reasoning about their phylogenetic origin. The first school is in a sense aprioristic, while the second school is essentially empirical. The empirical taxonomists view biological classification as merely a special branch of the general theory of classification, while the phylogenetic taxonomist feels that biological material, because of its unique history, must also be classified in a manner reflecting its historical development. To the empirical taxonomists classification is merely a way of arranging objects. As Gilmour (1961) has pointed out, the term classification or arrangement is meaningless unless the purpose for which the classification is to be constructed is stated. There may be specialized classifications of plants for special purposes: plants living in swamps, plants that are insect pollinated, plants without sepals, woody plants, plants with certain medicinal properties, etc. Special classifications while useful for a given purpose may not at all be suitable for another purpose; thus a classification of flowers by color may be of great interest to the student of insect pollination, but is not likely to be of much use for the ecologist studying stratification in a given area. However, biologists from very early times on have felt that there is one classification of greater merit than all others. A *natural* classification is one whose constituent classes have many attributes in common and which is most useful for a wide range of purposes (Gilmour 1961). The recognition of natural groups as entities sharing the largest number of properties is an extremely important concept which we owe to Gilmour. It frees us from the other interpretation of natural groups as representing lines of common descent. While this concept is anyhow not tenable in botany because of hybridization at certain hierarchic ranks, it still has a powerful influence on systematists. The difficulty with a phylogenetic concept of natural groups is that phylogenies are speculative and we are therefore establishing a taxonomy on grounds which are largely hypothetical. Such procedures do not permit classification to be repeatable, quantifiable or objective in any way. Taxonomy therefore should aim for the best possible way of classification, which a number of us nowadays believe is an empirical one based on the *phenetic* evidence, i.e. evaluated on the basis of the resemblances existing in the material at hand (Cain and Harrison, 1960). Empirical taxonomists do not hold that it is the function of taxonomy to elucidate phylogeny or to explain it, or to investigate the mechanism of evolution. The fields of paleontology, speciation and genetics concern themselves with this. It is the function of taxonomy to arrange the organisms in the best possible classification which we can arrive at on the basis of all characters at hand. It will be argued that natural groups could only have arisen because organisms have had phylogenies resulting in gaps between evolutionary lines. The clusters which we observe in nature today are undoubtedly the results of phylogenetic processes and are in fact excellent indirect evidence for the existence of evolution. However, we cannot use speculation on phylogeny to establish the clusters themselves. This point has been made by a number of authors and would hardly need repeating. Yet this is the major bone of contention between taxonomists of the phylogenetic school and that of the empirical school. Gilmour (1961) has recently stressed this point in saying "We must I suggest, draw a distinction, . . . between a natural classification *made possible* by the influence of phylogeny and the classification *based on* phylogeny" (his italics).

I shall now proceed to discuss briefly the principles of numerical taxonomy. These are elaborated in greater detail by Sneath and Sokal (1962) and in a forthcoming book by Sokal and Sneath (1963). A fundamental principle of numerical taxonomy is the strict segregation of phylogenetic speculation from taxonomic procedure. Taxonomic affinity is determined purely by similarities based on observable characters of taxa, so-called *phenetic* similarities. Since there has been for almost a century an intimate connection between taxonomic and phylogenetic concepts, the segregation of these procedures is difficult, yet necessary. Natural taxa cannot be based on phylogenetic information since only a very small part of the phylogenetic history can be taken from the fragmentary fossil record. Homologies, on the basis of which natural taxa are frequently erected, are rarely based on paleontology but are probabilistic evaluations of the independent occurrence of the structures concerned. Discussions of homology very often lead to circular reasoning. The use of customary phylogenetic diagrams or trees is to be deplored, since different authors use these to depict different ideas and since a number of the ideas intended to be represented cannot be shown in these two-dimensional diagrams, for mathematical or logical reasons.

All kinds of characters are equivalent (morphological, physiological or cytological). Taxonomic equivalence of all kinds of characters (which leads directly to the absence of weighting of any character) can be defended on numerous genetic and logical grounds. The simplest demonstration of the illegitimacy of weighting characters is shown in the futility of any effort to define a consistent system for weighting these.

The most important single procedure in numerical taxonomy is the computation of similarities, which is done on the basis of many characters (often as many as 100) of the taxonomic groups to be processed. Three types of coefficients have been employed in the computation of similarities or affinities. The first and simplest are the so-called coefficients of association. These employ largely, but not necessarily, those characters subdivided into only two character states or attributes, which are frequent in microbiology (see Table 1). Such character states are coded as "plus" and "minus" or "zero" and "one". A number of coefficients of association have been developed in psychology and ecology and can be applied in numerical taxonomy. They express some measure of the number of character matches or character agreements out of the total number of characters available or employed in the study. These coefficients lie generally between zero and one. Sneath (1957 *a, b*), working with bacteria, was probably the first to employ this method for classificatory purposes und numerous bacteriologists have followed his example. The similarity ratio of Rogers and Tanimoto (1960), which has been applied to plants, falls into the category of association coefficients. It is mathematically related to the coefficient of Sneath and also to a simpler coefficient proposed by Sokal and Michener (1958).

Table 1. A Data Matrix

	Taxa (Operational Taxonomic Units or OTU's)				
	1	2	3	4	5
Characters					
1	0	1	0	1	0
2	0	1	0	0	0
3	1	0	1	0	1
4	1	0	1	1	1
5	NC	1	0	1	0
6	NC	1	NC	1	1
7	0	1	1	1	0
8	1	0	1	1	0

Columns represent taxa and rows characters. In this hypothetical example data are coded as "all-or-none" characters. In a real study more characters would have to be used. NC stands for "no comparison". It is a code employed where no record is available for a given character of a taxon.

A second method of computing similarities is based on Pearson product-moment correlation coefficients between various taxa. In this method the characters are free to include an arbitrary number of characters state classes as shown in Table 2. Correlation coefficients are free to vary from -1 to $+1$.

Table 2. A Data Matrix

	Taxa (Operational Taxonomic Units or OTU's)				
	1	2	3	4	5
Characters					
1	1	8	1	7	2
2	1	6	1	6	1
3	6	1	5	1	4
4	1	0	1	0	1
5	NC	6	3	6	NC
6	NC	2	NC	3	1
7	8	2	7	2	5
8	1	6	1	6	3

Data matrix as in Table 1, but here more than two states have been recorded for all characters except character 4.

A third method determines the similarity between two taxa as the distance in an n-dimensional space whose coordinates are the characters (Sokal, 1961). The distance can be calculated by a simple formula from analytical geometry. Figure 1 shows the distances among four taxa in a space defined by three characters. If the state codes of the characters have previously been standardized, the squared distance approaches Pearson's coefficient of racial likeness used in physical anthropology (Pearson, 1926). However, in physical anthropology we work with few taxa and few characters per taxon but numerous individual replicates, while in numerical taxonomy we use a typical single representative of a taxon but many characters and many taxa.

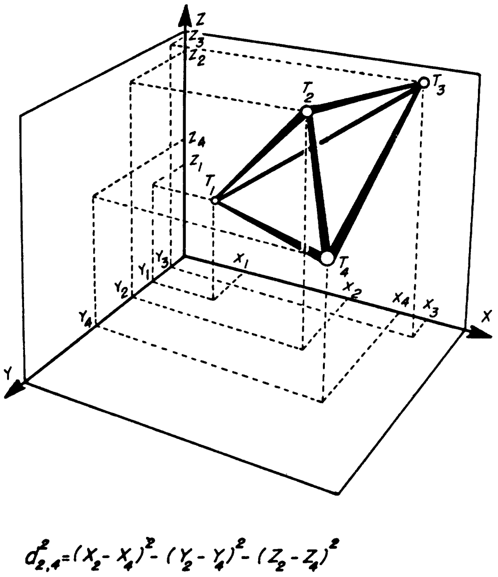

Figure 1. Representation of four OTU's, T_1, T_2, T_3, T_4, as points in a three-dimensional solid determined by the character states for three characters, X, Y and Z. Each character is represented by a dimension. The distance, d, between a pair of OTU's is given by the formula shown. (Redrawn from Sokal and Sneath, 1963).

Regardless of which of these methods of evaluating affinities is used, as long as the characters are properly identified and defined, so that every researcher can classify a given taxon into the appropriate character state class, and if the method of finding a coefficient of similarity is reached through a mutually agreed on series of computational steps, then all three types of methods are objective and repeatable in a first simple sense. However, will such studies agree if another researcher with partially or entirely different characters would engage in a study of the affinity of the same taxa? If two taxa are cómpared on the basis of a sample of the characters, their affinity can be expressed as that proportion of all the characters whose character state codes agree. We can assume that the similarity between two taxa is a parametric proportion of character state agreements which is estimated by the sample of characters. The more characters we sample, the more reliable the similarity coefficient will be, and finally a point is reached in which a further study of characters will no longer be worth-while. This hypothesis seems plausible to us at the moment and agrees with our experiences to date.

Once the coefficients of similarity have been computed they are tabulated in matrix form, one coefficient for each pair of taxa. A symmetrical matrix for t taxa consists of t

Table 3. A Similarity Matrix

		OTU's					
		A	B	C	D	E	F
OTU's	A	X					
	B	0.033	X				
	C	0.995	0.028	X			
	D	0.034	0.973	0.030	X		
	E	0.927	0.090	0.941	0.058	X	
	F	0.331	0.450	0.394	0.400	0.452	X

The six operational taxonomic units are labelled *A* through *F*. The coefficients of this hypothetical matrix are Pearson product-moment correlation coefficients. Each coefficient expresses the phenetic similarity between the two OTU's which define its row and column position. Thus the similarity between OTU's *B* and *D* is 0.973.

rows and *t* columns with unity (or *X*'s) in the principal diagonal (see Table 3). The classificatory process consists of several methods which reveal and summarize the inherent structure of the matrix. One may obtain an approximate representation of the structure of the matrix, when its elements are shaded according to the magnitude of the coefficients, which they represent. If the related taxa have been placed roughly adjacent to each other in advance, the taxa can be recognized immediately. However, on methodological principles it is better if the taxa are completely randomly assorted at the beginning. There are search methods, so-called cluster analysis, which will go systematically through these matrices and arrange them in a reasonable hierarchic system. The groups which we obtain as a result of this analysis can be analogized with a topographic chart, and the criteria for the delimitation of a group are analogous to the contour lines of such a chart. Strict criteria are analogous to high contour lines surrounding single high mountain tops, as for example species in a group of species. When criteria are loosened, the groups begin to grow and run into each other, just as single mountain tops would get broad bases and unite with each other to form mountain chains as the contour lines were lowered. The differences between the various methods of cluster analysis rest mainly on the definition of the clusters. It is particularly important to determine the principles for evaluating similarities between higher categories. How is the resemblance of a single taxon to another group containing several taxa to be computed? Are similarity coefficients between the higher taxa to be based on averages of the coefficients of the taxa contained therein or on some different function of the coefficients? These are technical questions which we cannot take time to discuss here. I would like, however, to say with some satisfaction that all these methods are statistically robust. By this we mean that the results do not vary very much in spite of rather great differences in methodology.

In such a manner we can get groups of any desired rank or order whose delimitation can be done in quite an objective manner. The results can be put in the form of a dendrogram which is clearly not a phylogenetic tree, but simply shows magnitude of phenetic relationships with the nearest stem (see Fig. 2).

Cross sections at any given level enclose all taxa born by one stem into one larger taxon. The number and the levels of such cross-sections must, of course, follow a predetermined system, which depends on the size of the matrix of similarities. Too many cross-sections would give too fine a classification. Too few would not reveal the entire structure. Since the aim and the purpose of the systematist must be contained within the basic rules of the classificatory process, once the rules have been set they cannot be changed in the middle of the analysis. The application of different criteria in different parts of a scheme of relationships would vitiate our entire efforts to bring systematics onto a strictly quantitative basis. This is, of course, in crass contradiction to present day practices in taxonomy.

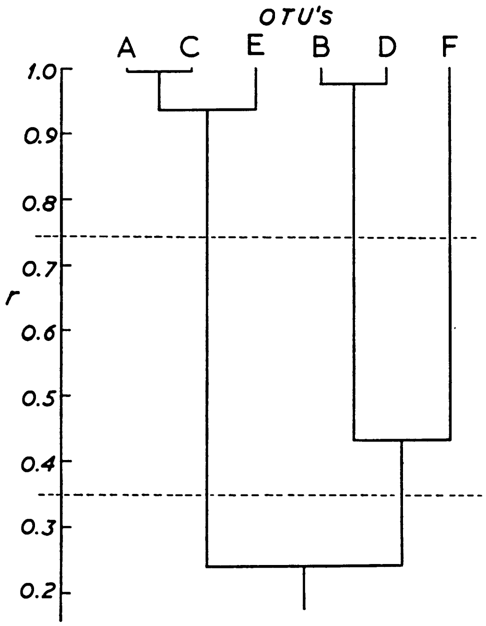

Figure 2. Dendrogram of the relations among the 6 OTU's whose correlations are given in Table 3. The dendrogram is based on the weighted variable group clustering procedure (Sokal and Michener, 1958). The ordinate shown at the left is in correlation coefficient scale, *r*. Phenon lines have been drawn in at *r* values of 0.75 and 0.35 in order to illustrate the delimitation of taxa.

Another method of finding structure in a matrix of similarity coefficients is factor analysis. We employ centroid factor analysis after Thurstone, using analytical methods of rotation to simple structure. The results which we have obtained by this method are very encouraging (Rohlf and Sokal, 1962).

Turning now more specifically to the actual computer operations involved in a study of numerical taxonomy, I have found by experience and in talking with other taxonomists that a number of misapprehensions are current regarding the relative roles of taxonomist and computer in such procedures and also the extent to which the function of the taxonomist is supplemented or even replaced by that of the computer. It should be obvious even after brief consideration that the study of the pertinent systematic material, the collection of data, the evaluation of characters as to whether they are admissible in a study of numerical taxonomy, the coding of such data and similar procedures necessarily and unequivocally remain within the realm of the practicing taxonomist. The computer performs only two steps: the computation of the similarity matrix and its structuring into a hierarchic taxonomic system. The first of these procedures must be done by a computer, since by the principles of numerical taxonomy a large number of characters must be processed in order to get an adequate estimate of the affinity between organisms. Such estimates of affinity are, of course, performed by the conventional work and thought processes of the taxonomist. However, conventional taxonomists never use as many characters as are employed in numerical taxonomy and certainly do not consider all characters simultaneously. They also do not necessarily employ the same characters for different taxa within a study. It is humanly impossible to consider affinities based on a large number of characters, say 60, although overall affinity judgements by taxonomists, such as that "a tiger is more like a cat than a dog", are abstractions of the same kind of procedure as is used routinely by the computer is making similarity judgements. The finding of structure in a similarity matrix is a procedure that could be performed by the taxonomist on a numerical basis, but is generally a rather subjective procedure. Here again the computer is able to handle the operation in a minimum of time, consistently and objectively. Given the same basic information the computer must inevitably arrive at the same classificatory scheme. This is not necessarily true of conventional taxonomic procedure! The uses to which the taxonomist puts the information presented to him by the computational program are of course various and again subject to his wishes and control. He may wish merely to publish a classification based on the affinities or phenetic similarities of the taxa, or on the other hand he may wish to use these data in erecting a presumed phylogenetic system of classification. He is, of course, entirely free to do so. It should therefore clearly be understood that the machine program is not in any way trying to replace any of the essential functions of the taxonomist. It will merely try to speed up and improve several of the operations which he has to perform in the execution of his work. It is merely a tool, a most efficient tool indeed, but as a tool is no different in its way from the microscope which upon application has opened up an entirely new dimension in establishing adequate and meaningful taxonomies for various groups of organisms.

Some estimates of time and costs involved should interest this audience. When we first undertook our study we had only the simplest kind of tabulating equipment plus desk calculators available. No wonder our initial study on bees (Michener and Sokal, 1957) took in the vicinity of six weeks computation time alone. On a medium speed computer this work can now be done in less than a day for probably less than $ 500. This sum may seem high to you, yet it would be quite impossible for a taxonomist working full time to classify this material in one month, not to mention the fact that we claim he could not possibly classify it as well and as accurately as the computer did. With a more modern computer the total computation time could probably be compressed into one hour's time with computational charges of not much more than $ 150.

The application of numerical taxonomy to botany has so far been limited to four studies. Morishima and Oka (1960) working with rice (*Oryza* species) found reasonable agreement of the numerical results with previous conventional classifications and with genetic data. Soria and Heiser (1961) studied species of *Solanum* and found rather close agreement with the earlier taxonomy and also some agreement with the relations as judged by the ease of hybridization between the species. Hamann (1961), in a numerical taxonomic study of the thirteen families of the "Farinosae" of Engler together with four other monocotyledonous families, found the resulting similarity coefficients agreeing well with newer opinions on relationships of these families. Rogers and Tanimoto (1960) carried out a preliminary numerical taxonomic study of manioc plants. In the microorganisms these methods have been employed very widely (see Sokal and Sneath, 1963, for a review of this extensive literature).

In conclusion I might mention some special problems of numerical taxonomy in relation to botany. One of the most obvious is the scarcity of characters. It has already been pointed out that many characters, at least 40 but usually many more than that, are used in obtaining similarities between taxa in numerical taxonomy. This may seem like a very large number to be able to find in botanical materical. In two published numerical taxonomic studies on higher plants (Morishima and Oka, 1960, and Soria and Heiser, 1961), 42 and 26 characters, respectively, were employed by the authors. Both of these studies involved closely related species and races. The authors of one of these studies stated that "26 characters was the maximum number which we could obtain for this study." It is, of course, quite possible that fewer characters are found to vary among close relatives; also that in some groups of organisms there are fewer visible characters than in others. This is certainly true for animals, where animals in shells such as mollusks are likely to have fewer easily distinguishable characters than flies, for example. Based on the amount of genetic differentiation which we can safely assume exists among species and higher categories in any group, an abundance of characters must exist but may be difficult to perceive. Thus studies of internal anatomy, or histology, as well as physiology, biochemistry and ecology may have to be undertaken in order to procure a sufficient number of characters for constructing a stable classification in groups with few visible characters. This clearly cannot be achieved in the immediate future, but in any case conventional taxonomy has been moving in the same direction of utilizing such characters. Numerical taxonomy does not present a radical departure in this respect, but only underlines the need for many observations to be obtained in order to construct adequately based taxonomies.

Another point in this connection is that the scarcity of characters may in some instances be more apparent than real and may be based on the traditional conceptions of what we mean by characters. In the conventional taxonomic sense, characters are properties of the taxa being classified which distinguish them one from the other. Characters which do not "make sense" are generally ignored by taxonomists after a few so-called good characters have established a rough classification. Characters which do not agree with this general outline are frequently rejected. In numerical taxonomy all characters would be used, and if in searching for characters taxonomists free themselves from these preconceptions just mentioned, many more characters can usually be found. In a study of a very well known genus of mosquitoes, one of my students (Rohlf, 1962) was able to find more than two dozen *new* characters on the legs alone, when looking at character variation from this empirical point of view. The more pronounced lack of a fossil record in botany as compared with certain zoological groups, coupled with the somewhat weaker penetration of phylogenetic thinking into botany as compared with zoology, should make plant taxonomy particularly amenable to numerical taxonomy. Modern botanists have in any case rejected some of the more fantastic homologies and phylogenies erected by phylogenetic evolutionists at the turn of the century. There are therefore fewer "well established" relationships and ancestries to be upset or corroborated by numerical taxonomy.

Two further problems in plant taxonomy require mention. These are the problems raised by cytogenetic work and the problem of hybridization. Plant cytology more than any other field can lay claim to giving insights into the true relationships among organisms. How will cytogenetic results agree with numerical taxonomy, and are the two methods to be jointly employed in evaluating relationships or used alternatively? Cytologists themselves would be the first to admit that their interpretations have in different instances varying degrees of probability. Thus, while in some case the ancestry of some species of plants can be established with near certainty on the basis of the karyotype, in other cases the cytological interpretation is very problematical. Numerical taxonomy on the other hand is an objective measurement of the similarities based on as many characters as the investigator wishes to use. The system as such can be in error only because of poor or inadequate choice of chaıacters. No comparative studies have been made between the results of these two approaches. At the moment it seems to me that they should be worked on separately, because in botany in those cases where ancestry is pretty well worked out cytogenetically we have the opportunity for measuring the amount of phenetic divergence which certain definite evolutionary paths have brought about. However. I would include characters of the karotype, as such. in the measurement of phenetic similarity whenever they are known. This would include the number of chromosomes. the morphology and any other characteristics of the karotype that differ among the taxonomic units being compared.

Hybridization is known in other groups of organisms, but its most frequent occurrence is in plants. Numerical taxonomy by being able to put quantitative values on similarities between various forms should be able to afford a measure of the degree of resemblance which hybrids have to their presumed ancestral forms (and any other forms, for that matter). When working with putatively hybrid forms the dendritic type of representation of Fig. 2 may not be the best one, and the notion of points plotted in a hyperspace which we used for distance may be the best way of looking at these.

SUMMARY

The tremendous advances in electronic data processing are likely to result in revolutionary changes in the theories and practices of taxonomy. The process of classification is being removed from speculations regarding the origin of the taxa being classified. A natural classification is one whose taxa share the largest number of properties and which is most useful for a wide range of purposes. The principles of numerical taxonomy are stated briefly and illustrated by means of diagrammatic examples. The relative roles of the taxonomist and computer are discussed and estimates given of computer time and costs involved in numerical taxonomic work. The numerical taxonomic work done in botany so far is discussed and the paper concludes with a brief mention of several problems of numerical taxonomy with regard to botanical work. These are: scarcity of characters, correlations between cytogenetic work and phenetic similarities, and problems raised by hybridization.

References

CAIN, A. J. and G. A. HARRISON 1958 — An analysis of the taxonomist's judgement of affinity. Proc. Zool. Soc. Lond. 131: 85-98.

GILMOUR, J. S. L. 1961 — Taxonomy, p. 27-45. *In* A. M. MacLeod and L. S. Cobley, [ed.], Contemporary botanical thought. Oliver and Boyd, Edinburgh and Quadrangle Books, Chicago. 197 p.

HAMANN, U. 1961 — Merkmalsbestand und Verwandtschaftsbeziehungen der Farinosae. Ein Beitrag zum System der Monokotyledonen. Willdenowia 2: 639-768.

MICHENER, C. D. and R. R. SOKAL 1957 — A quantitative approach to a problem in classification. Evolution 11: 130-162.

MORISHIMA, H. and H. OKA 1960 — The pattern of interspecific variation in the genus *Oryza:* its quantitative representation by statistical methods. Evolution 14: 153-165.

PEARSON, K. 1926 — On the coefficient of racial likeness. Biometrika 18: 105-117.

REMANE, A. 1956 — Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Theoretische Morphologie und Systematik. I. 2nd. ed. Akademische Verlagsges. Geest and Portig, Leipzig. 364 p.

ROGERS, D. J. and T. T. TANIMOTO 1960 — A computer program for classifying plants. Science 132: 1115-1118.

ROHLF, F. J. 1962 — A numerical taxonomic study of the genus *Aedes* (Diptera: Culicidae) with emphasis on the congruence of larval and adult classifications. Ph. D. Thesis. Univ. of Kansas. 98 p.

SIMPSON, G. G. 1961 — Principles of animal taxonomy. Columbia Univ. Press, New York. 247 p.

SNEATH, P. H. A. 1957*a* — Some thoughts on bacterial classification. J. Gen. Microbiol. 17: 184-200.

SNEATH, P. H. A. 1957*b* — The application of computers to taxonomy. J. Gen. Microbiol. 17: 201-226.

SNEATH, P. H. A. and R. R. SOKAL 1962 — Numerical taxonomy. Nature. 193: 855-860.

SOKAL, R. R. 1961 — Distance as a measure of taxonomic similarity. Systematic Zool. 10: 70-79.

SOKAL, R. R. and C. D. MICHENER 1958 — A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38: 1409-1438.

SOKAL, R. R. and P. H. A. SNEATH 1963 — The principles of numerical taxonomy. W. H. Freeman, San Francisco and London. (in press).