



Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines

We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content

The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute

The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

CONFERENCE ON MICROBIAL CLASSIFICATION
AUGUST 30TH - SEPTEMBER 5TH, 1964
LAVAL UNIVERSITY, QUEBEC CITY, CANADA

CONFERENCE PROGRAM

NATIONAL RESEARCH COUNCIL OF CANADA
DIVISION OF BIOSCIENCES
OTTAWA 2

SESSION No. 1

INTRODUCTION TO THE CONFERENCE

P. H. A. Sneath (National
Institute for Medical Research, London, N.W.7, England),
Chairman.

These notes are intended for general guidance, and though some of them have appeared earlier in the circulars, they are repeated here for convenience. We hope that you will all take a share in the program of the conference, and let us know of any improvements you would like to suggest.

The conference will consist of discussion sessions, each centred about a focal topic. The individual sessions may contain some teaching (in the interests of comprehension of the subject), but this will not be primarily a teaching conference. Strong emphasis will be placed on numerical taxonomy, particularly of bacteria. We hope you will all join freely in the discussions, which should be the most valuable part of the conference, and for this reason we would like the meetings to have an informal atmosphere.

The topics have been chosen to cover the ground fairly broadly, but we will not necessarily give equal time to each. The sessions have been planned on the following broad outline, so that they fall into a logical sequence as far as possible.

- Classification and its theory.
- Statistical ideas and numerical taxonomy.
- Data-processing methods.
- Obtaining the primary data:- techniques and cultures.
- Nomenclature.
- Identification.
- Relation to genetics and other subjects.

You will see from this outline that microbial classification has been treated as an interdisciplinary subject, and the sessions emphasize the links with other fields like genetics and statistics. This should encourage cross-fertilization between ideas in these fields. We hope there will be ample time for discussion. The evenings have been kept free as far as possible, to afford opportunity for you to discuss the day's meetings among yourselves informally.

We do not propose to publish the proceedings of this Conference, though a brief report may appear in Nature. Any communications should be regarded as unpublished personal communications, and the author's permission should be obtained if they are used elsewhere. I hope myself, however, that original work mentioned here will be published in due course, and would like to urge contributors to do this.

Whether the Conference would like to make any formal recommendations is a point we can discuss in the closing sessions. Please let us know your views on this fairly soon. It would seem in order for us to make recommendations to IAMS or its subcommittees, for example.

Lastly, may I express the hope that you will all find this Conference one of the most enjoyable and profitable you have attended, and thank everyone who has helped toward making it the success it is sure to be.

Introduction to the General Theory of
Classification

P. H. A. Sneath (National Institute
for Medical Research, London, N.W.7, England,
and after 1st October 1964, M.R.C. Microbial
Systematics Research Unit, Leicester University,
Leicester, England).

The first session is to "set the stage", so to speak, and it will deal with very general ideas and methods, emphasizing the logical order of the steps, and trying to make clear the meanings of technical terms. We will be discussing bacteria as a rule, but also other microorganisms where the taxonomic problems are similar. However, anything can be classified by these general methods (soils, words, pottery, etc.).

These notes are divided into two parts, descriptions of basic concepts, and of the logical steps in classification.

Descriptions and Definitions

These definitions are not to be taken too rigidly, as usages differ a lot.

Classification: used both for a process and a result. The result of the process of classification is a classification. Classification in the wide sense is the process of arranging entities according to their relationships (unspecified). In the narrow sense it implies placing entities into separate and distinct classes: however it is seldom restricted in this way (categorization can be used for this narrower usage). Here it is used in the wider sense, so that an array of bacteria without sharp divisions will be called loosely "a classification".

Taxonomy. The theoretical study of classification of living creatures. It is usually thus restricted to biology, but sometimes it is used for any subject, and then it becomes in effect metaclassification.

Systematics. The study of the diversity of organisms and of all of their relationships. This therefore embraces taxonomy, biological classification, nomenclature, and parts of many other subjects like genetics,

morphology, evolution, ecology, etc.

Relationship. This, like the word affinity, has so many usages that it is best to qualify it with an adjective. In particular, it does not necessarily mean relationship by ancestry. Phenetic relationship means overall similarity or resemblance, i.e. it is judged from the phenotype, and is not phyletic relationship. Whether certain sorts of genetic relationship are phenetic is a matter for debate.

Phyletic or phylogenetic relationship. It is essential to distinguish two sorts of phyletic relationship:-

- (a) cladistic, indicating descent alone, i.e. where or when the lineages branched, and saying nothing whatever about phenotypic resemblance. The relationship is measured in years, or number of generations, to the common ancestor.
- (b) Patristic, indicating phenetic resemblance, but only taking into account the characters that are not convergent (i.e. acquired independently in the two branches of the forked lineage). It is measured in the same units as phenetic resemblance, so that:-
patristic + convergent = total phenetic.

Resemblance. This is used as a neutral word to indicate overall similarity, dissimilarity, or affinity, and the arithmetic measures of it are called coefficients of resemblance.

Purpose in classification. When we pick (as we do) a certain classification as the best one, this implies we have a purpose in mind. This purpose may be a special one - such as the cataloguing of organisms by size or colour, and special classifications are then appropriate. An alphabetic catalogue is an example. But for any other purpose the classification is of little value. We can, however, make general-purpose classifications, like orthodox taxonomies, where no single purpose is overriding. These are often called "natural" classifications, or "polythetic" ones (see below). The purpose we have will always determine the classification methods we employ.

Character weighting. Intimately bound up with purpose is the problem of the relative weights to be given to characters, and there is much misunderstanding here.

We are handling information, and each unit of information equates with a unit character. Numerical taxonomy has generally followed the Adansonian practice of avoiding a priori weighting, i.e. giving unit characters unit weight. It need not do so, however, as there are alternatives. But if we give unit information extra weight it implies we consider it more important for the purpose of the classification. So we have to be clear of our purpose every time. It is true that characters omitted have zero weight, but this need not greatly bias the classifications based on large samples, if statistical sampling theory holds, as it generally does. However, excessive weighting can introduce bias, so that equal-weighting is usually the safest procedure. There may be some doubt about the best unit of information.

Monothetic and polythetic classifications. The traditional Aristotelian logic based classification on the well-known A, not -A, principle. Each section was then subdivided further (e.g. B, not -B). In this system the divisions are ones such that their members possess a unique set of attributes that are both sufficient and necessary for membership. This defining set of attributes is unique, and the group is called monothetic (monos one, thetos, arrangement). They allow no "exceptions", and are hence "unnatural" or "artificial", but are clearly very suitable for diagnostic keys.

If, however, we allow the percentage of shared attributes to determine membership, we get another type of group, called polythetic. Here there need not be any single character that is both sufficient and necessary. Thus the polythetic group of organisms 1, 2, 3, 4 might be as follows:-

Organism	Characters
1	A, B, C
2	A, B, D
3	A, C, D
4	B, C, D.

Here there is no character, of the four, A, B, C, D, common to all the organisms, but each shares three with every other organism. The group is called fully polythetic. If there are some characters common throughout, the group is partly polythetic.

Polythetic groups are of the kind we call "natural" in the sense of Gilmour (1), i.e. they have the greatest "content of information", so that one can make the

maximum number of deductive propositions about a member. They are, of course, not so convenient for diagnostic keys as are monothetic groups. Polythetic classifications are general ones, and this is what biological taxonomies are (as opposed to special classifications, based for example on size or colour, which are usually monothetic). Polythetic methods require many characters, since they rely on many attributes, rather than on a few key attributes like monothetic methods.

Hierarchy. The nested hierarchy is very convenient as a summary of the relations between OTU's, since it greatly assists our memories. It may not, however, always be the best system, and will distort the relations in the interests of simplicity.

OTU's, or operational taxonomic units, are the entities to be classified. They can be of any kind (species, individuals, and so on). Any given one is symbolized by the letters j or k, and the total number is called t.

Character. Any attribute that differs in the collection of entities that are being studied. All-or-none characters are those that can be expressed in two states, or by + or - signs, and have been called features. Feature is too vague to be a good technical term, however. The states of a character are the alternatives in which it differs in the collection. Multistate characters have more than two states. Quantitative multistate characters have the states in a linear array (e.g. 1, 2, 4.5, 8.9) while qualitative multistate characters cannot be so arranged linearly (e.g. a, α , *, @). Any given character is symbolized by the letter i, and the total number is called n.

Homology. Like "species" this term has contradictory usages. It has now come to mean a character state derived from a common ancestor. Unfortunately, however, we usually don't know about this, and the word simply means in effect "the same character". This is operational homology. There are difficulties in defining it, but in microorganisms it is usually possible to do so in operational terms (e.g. "lactose plus" is such a reaction under such conditions).

Congruence. The agreement of two classifications of a given collection of entities, each based on a different set of characters (e.g. adult characters vs. larval characters, morphological vs. physiological).

Matrix. Just a fancy word for a table. The "data matrix" used to classify the OTU's is called the $n \times t$ matrix (or table of characters versus organisms). The "similarity matrix" records each OTU compared with every other, and is of size $t \times t$ (also called a matrix of resemblance, affinity, distance, or dissimilarity, according to the circumstances).

Cluster analysis. Now used as a general term for methods for creating clusters or groups of OTU's.

Dendrogram. A tree-like diagram of relationship (of any kind). A taxonomic hierarchy can be shown thus.

Taxon (plural taxa). This is just an abbreviation for taxonomic group (of any kind, nature, or rank).

Phenon nomenclature. A phenon is any group formed in numerical taxonomy. It approximates to a natural taxonomic group (for reasons that will be given, at least if the numerical method is polythetic), but it is not synonymous with the term taxon. A numerical prefix is often given to show the resemblance level at which a phenon is formed (e.g. 70-phenon means a group whose members are affiliated at 70% similarity or over).

Rank, taxonomic rank. A somewhat vague term, referring to the variation within a taxon, or the phenetic gap between adjacent taxa. It can be expressed precisely by the phenon nomenclature, but we should not think at present of an absolute criterion of rank applicable to all organisms. The named ranks are best called rank categories. Thus "tribe" is a category, the tribe Serratiae is a taxon, and tribal rank is the phenetic variability associated with the category of tribe. At the "species level" rank is used with several conflicting meanings.

Q-R relations. We have been thinking of classifying organisms into groups. This is called Q-study. We could also group characters into clusters, and this process, R-study, is clearly the complementary one to Q-study. R-studies give groups of characters that are similar in their distributions. We are now starting to explore the interesting relations between Q and R groups. In R-studies the characters are the OTU's.

Species. I hope that we will not get bogged down in the "species-concept". This is a confusion over words.

There are many concepts of species. The word just means kind. What kind depends on your own usage. Geneticists have lately pushed the concept of the genetic species as being the only "proper" one, even where it can seldom be applied (as in bacteria). If you want it that way, alright, but then let us only use the term when we mean it. However, we are forced in practice to use it, at other times, for instance when we talk about the species Pseudomonas aeruginosa (what other term could we use here?). Ravin (2) has shown up the whole thing for the hot air it is, by distinguishing the following usages:-

nomen species, the unit given a binomial name (e.g.

Pseudomonas aeruginosa), the binom of Camp (3).

genospecies, the unit capable of gene-exchange between its members (the biological species, or genetic species).

taxospecies, the unit with high mutual phenetic resemblance, a phenon in fact.

If we use these terms whenever we are tempted to say "species" by itself, we will seldom have any trouble with the "species-concept". Note that these terms do not specify the detailed criteria, only the basic sort of criterion. It is up to the user to say what degree of gene-exchange he means when defining a genospecies, or what phenon level equals a taxospecies.

The Logical Steps in Classification

Against the background of these definitions we can now see how classifications must have a logical sequence of procedures, though when we make classifications intuitively this may not be very clear.

- 1) Data collecting. First, we must gather data that is pertinent to our purpose. This purpose overrides all other considerations, for it determines what we wish to classify, on what characters, and the methods we will use. Numerical taxonomy has up till now been generally polythetic (though it need not be) and aims at a selection of characters representative of the phenotype, and sufficiently numerous to give reliable results. We have t OTU's and n characters.

- 2) Arraying. Second, we must choose the sort of relationship we are going to employ (monothetic, polythetic, phenetic, etc.) and compare the OTU's with each other, thus getting the relationships. These may give arrays, but so far we have no distinct separate groups.
- 3) Clustering. Thirdly, we must cluster the OTU's into groups according to the relationships. This usually means losing some information due to simplification.
- 4) Prediction. Fourthly, we can ask questions of our groups, such as what characters are best for diagnosis. These questions imply we can make predictions of various sorts. Clearly we can give different character weights at this stage (a posteriori).

It is clear that these four steps (though we may jump back and forth in our minds) are logically consecutive. We cannot ask what are the relations before we choose the data. We cannot ask what groups are present before we have the relations to base them on. We cannot describe properties of the groups before we have the groups themselves.

Further Reading

A few general references are given for further reading (4 - 7).

References

- (1) Gilmour, J. S. L. A taxonomic problem. Nature, 139 : 1040-1042 (1937).
- (2) Ravin, A. W. Experimental approaches to the study of bacterial phylogeny. Am. Naturalist, 97 : 307-318 (1963).
- (3) Camp, W. H. Biosystematy. Brittonia, 7 : 113-127 (1951).

- (4) Ainsworth, G. C. and Sneath, P. H. A. (Editors), Microbial Classification, XIIth Symposium of the Society for General Microbiology. (Cambridge : Cambridge University Press. ix + 483 pp., illust., 1962.)
- (5) Davis, P. H. and Heywood, V. H. Principles of Angiosperm Taxonomy. (Edinburgh and London : Oliver & Boyd. xx + 556 pp., illust., 1963.)
- (6) Simpson, G. G. Principles of Animal Taxonomy. (New York : Columbia University Press. xii + 247 pp., illust., 1961.)
- (7) Sokal, R. R. and Sneath, P. H. A. Principles of Numerical Taxonomy. (San Francisco and London : W. H. Freeman. xvi + 359 pp., illust., 1963.)

ANNOTATED BIBLIOGRAPHY ON MICROBIAL CLASSIFICATION

This bibliography is intended as an introduction to the main literature on microbial classification, with emphasis on bacteria. It is not intended to be exhaustive, and up-to-date reviews have usually been preferred over the original literature. Nevertheless, it is hoped that it gives a well-balanced list of the key literature, or else of the sources where that literature is listed more fully.

The references cover methods and concepts rather than actual taxonomies: for these see in particular the lists in Ainsworth and Sneath (1962) Appendix III. Numerical taxonomies of bacteria are reviewed in Sneath (1962) and (1964). This bibliography is intended to cover all the sessions, but only very lightly. This seems preferable to a fuller bibliography which would take too long to read through when help was needed on some point. For this reason many papers of current interest have been left for mention in the individual sessions.

P. H. A. Sneath,
August 1964.

Adams, M. H. Bacteriophages. Interscience : New York
(xviii + 592 pp., illust., 1959).

Taxonomy of bacteriophages is discussed on pp.421-438.

Adanson, M., Familles des plantes, Vol.1. Vincent : Paris.
(cccxxv + 190 pp., 1763).

The celebrated historical work on the methods and philosophy of biological classification.

Ainsworth, G. C. Pathogenicity and the taxonomy of fungi. Pp.249-269 in Ainsworth and Sneath, 1962
(q.v.).

Contains a discussion of speciation and suggestions on infraspecific taxa.

- Ainsworth, G. C. and Sneath, P. H. A. Microbial Classification : 12th Symposium of The Society for General Microbiology. Cambridge University Press : Cambridge. (ix + 483 pp., illust., 1962)
This symposium covers most angles of the subject. Besides individual contributions (listed separately under the authors' names) there are Appendices on pp.456-476 on nomenclature, culture collections and key works in microbial systematics, including most of the useful monographs of bacterial taxa.
- Ainsworth, G. C. Ainsworth & Bisby's Dictionary of the Fungi. 5th edition. Commonwealth Mycological Institute : Kew, Surrey. (viii + 547 pp., illust., 1961.)
A standard work in systematic mycology. Type nomenclature is well treated (pp.273, 418).
- Beers, R. J., Fisher, J., Megraw, S. and Lockhart, W. R. A comparison of methods for computer taxonomy. J. Gen. Microbiol. 28 : 641-652 (1962).
Discusses especially different scoring systems for numerical taxonomy (see also the preceding paper, Vol.28, pp.633-640).
- Bergey's Manual of Determinative Bacteriology, 1st Ed., by D. H. Bergey, F. C. Harrison, R. S. Breed, B. W. Hammer and F. M. Huntoon. Williams and Wilkins : Baltimore. (xii + 461 pp., 1923.)
Contains a valuable historical review.
- Bergey's Manual of Determinative Bacteriology, 6th Ed., by R. S. Breed, E. G. D. Murray and A. P. Hitchens. Williams and Wilkins : Baltimore. (xvi + 1059 pp., 1948.)
This edition contains the most detailed coverage of the literature.
- Bergey's Manual of Determinative Bacteriology, 7th Ed., by R. S. Breed, E. G. D. Murray and N. R. Smith. Williams and Wilkins : Baltimore. (xviii + 1094 pp., 1957.)
The current standard treatment of bacterial systematics. The older and more obscure literature will be found in the 6th Ed.

- BIMD Computer Programs Manual, Division of Biostatistics, School of Medicine, University of California, Los Angeles, (1961).
An extensive series of statistical computer programs, mostly in Fortran, are described. The actual programs are supplied separately.
- Bisset, K.A. The phylogenetic concept in bacterial taxonomy. Pp. 361-373 in Ainsworth and Sneath (1962) (q.v.).
An eloquent exposition of the phyletic point of view.
- Brisbane, P.G. and Rovira, A.D. A comparison of methods for classifying rhizosphere bacteria. J. Gen. Microbiol. 26 : 379-392 (1961).
An interesting paper on wide-range problems, using varied methods.
- Cain, A.J. Deductive and inductive methods in post-Linnaean taxonomy. Proc. Linn. Soc. Lond. 170th session : 185-217 (1959a).
This and the following article are admirable expositions of the development of systematic theory.
- Cain, A.J. The post-Linnaean development of taxonomy. Proc. Linn. Soc. Lond. 170th session : 234-244 (1959b).
- Cain, A.J. The evolution of taxonomic principles. Pp. 1-13 in Ainsworth and Sneath (1962) (q.v.).
A clear summary of the history of taxonomy.
- Cain, A.J. and Harrison, G.A. An analysis of the taxonomist's judgment of affinity. Proc. Zool. Soc. Lond. 131 : 85-98 (1958).
A modern zoological point of view.
- Cain, A.J. and Harrison, G.A. Phyletic weighting. Proc. Zool. Soc. Lond. 135 : 1-31 (1960).
Distinguishes the patristic and cladistic components of phyletic relationship.
- Corliss, J.O. Taxonomic procedures in classification of protozoa. Pp. 37-67 in Ainsworth and Sneath (1962) (q.v.).
A discussion on how protozoa are classified.
- Cowan, S.T. The microbial species - a macromyth? Pp. 433-455 in Ainsworth and Sneath (1962) (q.v.).
A shrewd discussion on this topic.
- Cowan, S.T. and Steel, K.J. A device for the identification of micro organisms. Lancet, (i), 1172-1173 (1960).
A useful identification method.

- Cummins, G.S. Immunochemical specificity and the location of antigens in the bacterial cell. Pp. 212-241 in Ainsworth and Sneath, 1962 (q.v.).
- Davis, P.H. and Heywood, V.H. Principles of Angiosperm Taxonomy. Oliver and Boyd : Edinburgh (xx + 556 pp., illust., 1963).
The most comprehensive and up-to-date text on botanical taxonomy, with a tremendous coverage of the literature. It should be known to all taxonomists.
- Defayolle, M. and Colobert, L. L'espece Streptococcus faecalis. II. Etude de l'homogeneite par l'analyse factorielle. Ann. Inst. Pasteur, 103 : 505-522 (1962).
- de Ley, J. Comparative biochemistry and enzymology in bacterial classification. Pp. 164-195 in Ainsworth and Sneath, 1962 (q.v.).
An excellent review.
- Dijkstra, E.W. A Primer of Algol 60 Programming. Academic Press : New York and London (xi + 114 pp., 1962).
A useful introduction to Algol.
- Doty, P., Marmur, J., Eigner, J. and Schildkraut, C. Strand separation and specific recombination in deoxyribonucleic acids : physical chemical studies. Proc. Nat. Acad. Sci., U.S.A., 46 : 461-476 (1960).
A classic paper on base-pairing in soluble DNA.
- Dybowski, W., Franklin, D.A. and Payne, L.C. Computer for bacteriological diagnosis. Lancet, (ii), 866 (1963).
A brief description of this pioneering work. It is to be hoped that fuller details will soon be published.
- Ehrlich, P.R. Systematics in 1970 : Some unpopular predictions. Systematic Zool. 10 : 157-158 (1961).
Discussion of the future of systematics.
- Fisher, R.A. The uses of multiple measurements in taxonomic problems. Ann. Eugen. 7 : 179-188 (1936).
The paper describing the Discriminant Function. For a worked example, see Whitehead (1954).
- Floodgate, G.D. Some remarks on the theoretical aspects of bacterial taxonomy. Bacteriol. Rev. 26 : 277-291 (1962).
A discussion.
- Gilmour, J.S.L. A taxonomic problem. Nature, 139 : 1040-1042 (1937).

Gould, S.W. Family names of the plant kingdom. Vol. I. of the International Plant Index. IPIx : New Haven and New York (111 pp., 1962).

An example of modern data-processing in systematics.

Grainger, T.H., Jr. A Guide to the History of Bacteriology. Ronald Press : New York, (210 pp., 1958).

An extensive annotated bibliography to a wide range of subjects, including systematics.

Greig-Smith, P. Quantitative Plant Ecology. Butterworth : London, and Academic Press : New York (198 pp., 1957).

A standard text in a field closely related to systematics.

A new edition is just appearing.

Cyllenberg, H. A general method for deriving determination schemes for random collections of microbial isolates. Ann. Acad. Sci. Fennicae, Ser. A. IV. Biol., No. 69 (23 pp., 1963).

A novel clustering method with a simple way of distinguishing the discriminative tests.

Index Bergeyana (in press).

This will give a complete coverage of the nomenclature of bacteria, and will be indispensable.

International Code of Botanical Nomenclature, adopted by the Ninth International Botanical Congress, Montreal, August 1959.

International Bureau of Plant Taxonomy and Nomenclature : Utrecht (372 pp., 1961).

International Code of Nomenclature of Bacteria and Viruses, 1958.

Iowa State Univ. Press : Ames, Iowa; reprinted with corrections, 1959 (156 pp.).

The current Bacteriological Code, with very valuable anonymous annotations, due almost entirely, so it is believed, to Professor R.E. Buchanan. The status of viruses under this code is uncertain, pending an expected new Code for viruses.

International Code of Nomenclature for Cultivated Plants.

International Bureau for Plant Taxonomy and Nomenclature, Utrecht. (Regnum Vegetabile No. 22, 30 pp., 1961).

Many ideas on nomenclature of cultivated varieties could be adapted to micro-organisms.

International Code of Zoological Nomenclature, adopted by the XV International Congress for Zoology, London, July 1958, London :

International Trust for Zoological Nomenclature, 19 Belgrave Square, London, S.W.1 (176 pp., 1961).

Protozoa, considered as animals, are treated under this Code.

- Jahn, T.L. Man versus machine: a future problem in protozoan taxonomy. Systematic Zool. 10 : 179-192 (1961).
The future application of data-processing machinery.
- Kendrick, W.B. and Proctor, J.R. Computer taxonomy in the Fungi Imperfecti. Can. J. Bot. 42 : 65-88 (1964).
A full discussion of scoring and coding in mycology.
- Liston, J., Weibe, W. and Colwell, R.R. Quantitative approach to the study of bacterial species. J. Bacteriol. 85 : 1061-1070 (1963).
An important paper, discussing the definition of a species in taxometric terms and the idea of the typical strain.
- Lockhart, W. R. and Hartman, P.A. Formation of monothetic groups in quantitative bacterial taxonomy. J. Bacteriol. 85 : 68-77 (1963).
A monothetic method that approaches the polythetic solution under many circumstances, and capable of handling very many OTUs. Also notes on scoring methods.
- Luria, S.E. Molecular and genetic criteria in bacterial classification. Pp. 604-616 in Recent Progress in Microbiology: Symposia held at the VIII International Congress for Microbiology, Montreal, 1962, ed. N.E. Gibbons, Univ. of Toronto Press : Toronto (xiv + 721 pp., illust., 1963).
A review of present knowledge: that of Marmur et al. (1963) is yet more recent.
- Maccacaro, G.A. La misura delle informazione contenuta nei criteri di classificazione. Ann. Microbiol. Enzimol. 8 : 231-239 (1958).
An early paper on "repartition" and key-making: see also Rescigno and Maccacaro (1960).
- McCarthy, B.J. and Bolton, E.T. An approach to the measurement of genetic relatedness among organisms. Proc. Nat. Acad. Sci. U.S.A. 50 : 156-164 (1963).
The fullest study to date on nucleic acid pairing.
- McCracken, D.D. A guide to Fortran programming. John Wiley : New York. (viii + 88 pp., 1961).
A useful Fortran manual.
- Marmur, J., Falkow, S., and Mandel, M. New approaches to bacterial taxonomy. Ann. Rev. Microbiol., 17 : 329-372 (1963).
The most thorough review to date of genetic and molecular biological aspects.

- Mattick, A.T.R., Cheeseman, G.C., Berridge, N.J. and Bottazzi, V.
The differentiation of species of lactobacilli and streptococci
by means of chromatography. J. Appl. Bacteriol., 19 : 310-321
(1956).
Description of this useful technique.
- Metcalf, Z.P. The construction of keys. Systematic Zool. 3,
38-45 (1954).
One of the few papers on how to make diagnostic keys.
- Michener, C.D. Some future developments in taxonomy. Systematic
Zoology, 12 : 151-172 (1963).
Some very interesting suggestions are made.
- Michener, C.D. and Sokal, R.R. A quantitative approach to a
problem in classification. Evolution, 11; 130-162 (1957).
An early paper on numerical taxonomy.
- Microbial Systematics Newsletter. Issued by the Microbial Systematics
Group of the Society for General Microbiology, Institute of Biology,
41, Queen's Gate, London, S.W.7, England.
A new venture. For those interested in the classification of
bacteria and actinomycetes.
- Müller, F. Quantitative methods in the systematics of Actinomycetales.
IV. The theory and application of a probabilistic identification
key. Giorn. Microbiol., 10 : 29-47 (1962).
Describes a method of identification which also gives the chance
of misidentification.
- Moroney, M.J. Facts from Figures, 3rd ed. Penguin books :
Harmondsworth, Middlesex. (viii + 472 pp., illust., 1956).
One of the clearest expositions of elementary statistics.
- Murray, R.G.E. Fine structure and taxonomy of bacteria.
Pp. 119-144 in Ainsworth and Sneath, 1962 (q.v.).
A useful summary of current knowledge.
- Norris, K.P. Infra-red spectroscopy and its application to
microbiology. J. Hyg., 57 : 326-345 (1959).
A review.
- Orla-Jensen, S. Die Hauptlinien des natürlichen Bakterien-
systems. Zentr. Bacteriol. 2 Abt. 22 : 305-346 (1909).
The most extreme form of chemical taxonomy of bacteria.
- Uchterlony, O. Diffusion-in-gel methods for immunological analysis.
Progress in Allergy, 5, 1-78 and 6 : 30-154. Karger: Basle
and New York (1958 and 1962).
Extensive reviews.

- Felczar, H.J. (and many others). Manual of Microbial Methods. McGraw Hill : New York (x + 315 pp., illust., 1957). This, with Skerman's Guide, is a main source-book for techniques and tests.
- Rescigno, A. and Maccacaro, G.A. The information content of biological classifications. Pp. 437-446 in C. Cherry (ed.), Information Theory - A Symposium Held at the Royal Institution, London, August 29 - September 2nd, 1960. Butterworth : London (476 pp., 1960).
An extension of Maccacaro's earlier work.
- Rogers, D.J. and Tanimoto, T.T. A computer program for classifying plants. Science, 132 : 1115-1118 (1960).
Several important taxometric ideas are introduced in this paper.
- Savory, T. Naming the living world : an introduction to the principles of biological nomenclature. English Universities Press : London (xiii + 128 pp., 1962).
A readable introduction to the history and purposes of the Rules of Nomenclature.
- Silvestri, L., Turri, M., Hill, L.R. and Silardi, E. A quantitative approach to the systematics of actinomycetes based on overall similarity. Pp. 333-360 in Ainsworth and Sneath (1962) (q.v.).
A thorough numerical study, with details of cluster analysis.
- Simpson, G.G., Roe, A., and Lewontin, R.C. Quantitative Zoology. Harcourt, Brace and Co. : New York. (440 pp., illust., 1960).
A useful text on statistics.
- Simpson, G.G. Principles of animal taxonomy. Columbia Univ. Press : New York. (xii + 247 pp., illust., 1961).
A phylogenetic viewpoint of taxonomy but with many good chapters, and lucidly written.
- Skerman, V.B.D. A Guide to the Identification of the Genera of Bacteria. Williams and Wilkins : Baltimore. (ix + 217 pp., illust., 1959).
An indispensable book, both for its keys and its collection of little-known techniques.
- Skerman, V.B.D. Species concept in bacteria. Pp. 213-221 in The Evolution of Living Organisms. Melbourne University Press : Melbourne (459 pp., 1962).
A discussion on central data-storage and processing in bacterial systematics.

- Sneath, P.H.A. Some thoughts on bacterial classification. J. Gen. Microbiol., 17 : 184-200 (1957a).
This and the next are two early papers on numerical taxonomy.
- Sneath, P.H.A. The application of computers to taxonomy. J. Gen. Microbiol., 17 : 201-226 (1957b).
- Sneath, P.H.A. The construction of taxonomic groups. Pp. 289-332 in Ainsworth and Sneath, (1962) (q.v.).
A summary of numerical taxonomy in bacteriology, with a review of its results (brought up to date in Sneath, 1964).
Includes a historical section.
- Sneath, P.H.A. New approaches to bacterial taxonomy: the use of computers. Ann. Rev. Microbiol., 18 (in press, 1964).
A review bringing the subject up-to-date and giving with the earlier reviews (in Sokal and Sneath, 1963 or in Sneath, 1962) a fairly complete coverage of methods and findings.
- Sneath, P.H.A. Comparative biochemical genetics in bacterial taxonomy. In Leone, C.A. (ed.), Taxonomic Biochemistry, Physiology, and Serology. Ronald Press : New York (in press).
Contains summary of inter-species bacteriophage cross-reactions up to 1962, though probably by no means complete.
- Sneath, P.H.A. and Cowan, S.T. An electro-taxonomic survey of bacteria. J. Gen. Microbiol., 12, 551-565 (1958).
A study covering a wide range of bacteria of different Orders.
- Sneath, P.H.A. and Skerman, V.B.D. A first list of type and reference strains of bacteria. (to be published in Internatl. Bull. Bact. Nomencl. Taxon.)
This will provide a comprehensive list of strains whose reliability for taxonomic work is fairly high. At present the most useful source is the latest American Type Culture Collection Catalogue (7th ed. 1964).
- Sneath, P.H.A. and Sokal, R.R. Numerical taxonomy. Nature, 193, 855-860 (1962).
A brief outline.
- Snedecor, G.W. Statistical Methods. 5th ed. Iowa State University Press : Ames, Iowa. (xiii + 535 pp., illust., 1956).
A useful textbook.
- Sokal, R.R. and Rohlf, F.J. The comparison of dendrograms by objective methods. Taxon, 11 : 33-40 (1962).
Describes the cophenetic technique.

- Sokal, R.R. and Sneath, P.H.A. Principles of Numerical Taxonomy. W.H. Freeman : San Francisco and London. (xvi + 359 pp., illust., 1963).
A comprehensive text on this subject, with sections also on philosophy, comparative serology, phylogeny and applications to other fields.
- Sporne, K.R. The phylogenetic classification of the angiosperms. Biol. Rev., 31 : 1-29 (1956).
A review of phylogenetic trees, dendrograms and ways of representing them.
- Steel, R.G.D. and Torrie, J.H. Principles and procedures of statistics. McGraw-Hill : New York (481 pp., illust., 1960).
A useful textbook.
- Steel, K.J. The practice of bacterial identification. Pp.405-432 in Ainsworth and Sneath (1962) (q.v.).
A wide review of great practical use.
- Taxometrics (Newsletter). Issued by the Progetto Sistematica Actinomiceti, Istituto P. Stazzi, Via Celoria, 10, Milan, Italy.
This is now the usual medium for exchange of information on computer programs etc. in numerical taxonomy, and also carries bibliographies, news items, etc. No 2 pages 3-1 to 3-5 has recommendations on program conventions.
- van Niel, C.E. The classification and natural relationship of bacteria. Spring Harbor Symposia Quant. Biol., 11 : 285-301 (1946).
A fine historical review.
- Whitehead, F.H. An example of taxonomic discrimination by biometric methods. New Phytologist, 53 : 496-510 (1954).
A worked example of Fisher's Discriminant Function.
- Wildy, P. Classifying viruses at higher levels : symmetry and structure of virus particles as criteria. Pp. 145-163 in Ainsworth and Sneath (1962) (q.v.).
A stimulating paper on viruses; N.W. Pirie's article in the same volume (Pp. 374-393) should also be read.
- Wilson, A.C. and Kaplan, N.O. Enzyme structure and its relation to taxonomy. In O.A. Leone (ed.) Taxonomic Biochemistry, Physiology, and Serology. Ronald Press : New York (in press).
A readable, stimulating, article on molecular biology and systematics.
- Woodger, J.H. On biological transformations. Pp. 94-120 in W.E. LeGros Clark and P.E. Medawar, (eds.), Essays on Growth and Form presented to D'Arcy Wentworth Thompson. Clarendon Press : Oxford. (408 pp., illust., 1945).
An important paper on the nature of operational homology.

①

Mention - AAAS symp 1960 - Usos computer Bot.
1-statement - sensitivity unit in taxonomy.

AAAS - Taxometrics 12/29/64

1. Definitions

1. objects - indiv, group, or taxon (syn. = O.T.U.)
2. character (or -istic) - a quant. or qual. descriptive term (mnemonic device) selected to define an object.
3. attribute - a state within a character - usually mutually exclusive. (syn. = taxonomic data.)
4. similarity ratio - (coefficients) - a comparison of 2 objects, in which the attributes of the objects are compared, the number of matches (identical attributes) are counted, and a figure representing the closeness of the 2 objects is derived.
5. Cluster - a general term for groups of similar objects.
6. Cluster analysis - methods to determine which objects may be placed in a cluster, or to determine the boundaries (parameters) of clusters.

2. Place of computers - use slide. -

1. Retrieval systems of \pm complex card sorting - Gould's method does no more than rapid card sorting.

3. Models + programs reflecting taxonomic classification procedures -

1. What is a model
2. " " " Program?

(2)

4. Historical

A. Pre-computer models - mostly statistical - developed as much by psychologists as by biologists.

B. Difficult to attract math. minds to the complexity of biolog. problems -

1. No clean-cut mechanisms -
2. No measurement system of precise operation.

3. Inter-related cells, organs, organisms,

4. Living, or alive - concepts difficult of precise definition.

C. Computer advent -

1. at first, mostly physical sciences.

2. Built in logic - decisions of yes-no, maybe.

3. New types math - rediscovery of Boolean algebras, non-statistical, capable of dealing with non-numerical values in a numerical sense - this ultimately most important for systematics.

D. Developers of models for taxonomy -

1. Basically, only three workers, whose first attempts appeared from 1957-1960 in order of appearance

a. Sneath. Eng. Doct. D. Social -

taxonomist & Botanist.

D 2 - Gregg - in 1949(?) publication attempted to explain taxonomy in terms of set theory - a useful, but not complete tool for systematics.

5. Processes: presently at work -

A. In spite of all the talk about Adanson vs non Adanson, Phylogenetic vs phenetic, etc.

only 3 processes have been looked at.

1. Data recording + coding
2. Measures of similarity
3. Clustering techniques.

B. Statistical vs non statistical models

1. Statistics ^{models} assumes an averaging, or "distance" from a stated norm - data are measured as deviants from a normal - hardly appropriate where no distributions, no norms are known.

2. Non statistical models - measures using the data provided directly -

3. Coding process may be dictated by the statistical model.

C. Coding process -

1. Depends upon subjective ability of taxonomist - selection of

④

5 C.2. No objective process to determine the value of a characteristic
- Mention importance of studies.

3. Binary vs multiple states

6. Similarity ratio - (coefficient)

A. Having coded data, some measure needed to show relationships.

1. All so far are pair-wise ratios.

may develop 3-4-way comparisons.

B. Several methods, giving 0-1 or fractions.

1. Sokal - Statistical = Pearson's product-moment coefficient.

$$r_{AB} = \frac{\text{Av. of Prod. of Deviations}}{(\text{Stand. Dev. A}) (\text{Stand. Dev. B})}$$

With this method, possible to have

2 objects that actually are related to show 0 relationship.

C. Sneath matching coefficient

Use with binary data - bacterial tests - all other info must be converted to binary.

D. Rogers + Tanimoto - no name given -
More counting of 1's + dividing
by no. of attributes

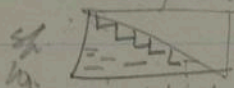
E. Several others - no ratio, but of similar -
- None so far are entirely satisfactory

(5)

7. Clustering - a separate set of methods from The Similarity procedure
- Search matrix of Similarity for close relations -
 - Put close (or highly similar) in same group, (or cluster)
 - Separate into hierarchical order - nested categories -
 - Use some procedure to exhibit the relations.

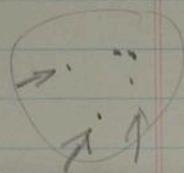
D. Methods

1. Sneath - no method, but a visual sorting of S_{ij} s along a diagonal



Useful for small no's of objects
not for large number.

2. Soxal - using highly correlated (with high coefficients) as nuclei; attempts to add other objects, on basis of uncomputed S_{ij} - to center of a cluster - demands a 3 dimensional geometry -



Criterion for clustering - if new obj. has coeff.

uses Pearson's coeff. for one added object;
Spearman's for more than

6

7D3 - Rogers-Tanimoto -

Model developed to eliminate dimensionality of cluster, to indicate degree of homogeneity within cluster -

Used arbitrary, though probabilistic constants - EM, delta, to set off clusters.

~~Ex.~~ New techniques

1. Graph clustering -

Does not have dimensionality
" " " arbitrary constants.

2. Requirements of methods of spherical nature not good for taxonomy, should be able to cluster odd shapes - cones and spheres.

8. Conclusions -

A. Confusion of mass of data w/ simplicity of the taxonomic procedures

B. All parts of process open - not finished

C. The simplest model will probably be best.

D. Whichever procedure will probably be more info-carrying than present procedures.

E. Naming of concepts - procedures, etc. very poor

Sokal's Method

1. Measurement of Data

i Continuously measured descriptive characters, divided into a finite number of states usually 2-9 in number.

ii Specimens are coded into descriptive character states.

iii The distribution of ^{objects assigned to} each state is transformed so that it is approximately a normal distribution.

iv The character states are relabeled in each character so that the mean value at a character state is 0, and the standard deviation of each distribution is 1.

v Specimens are recoded with the normalized character states.

2. Computation of Similarity

Pearson's Coefficient of Similarity is used.

$$r_{AB} = \frac{\text{Average of the Product of Deviations}}{(\text{Standard Deviation of A}) \times (\text{Standard Deviation of B})} \quad 15$$

computed as follows:

- i the average code value for A is found, similarly for B
 - ii A is recoded in terms of deviation from its average code value, similarly for B
 - iii the standard deviation for the coding of A is computed, similarly for B
 - iv the product of the deviations (in ii) at each character is taken and the average of those products is found
- A matrix of Pearson coefficients is made

A	B
2	3
2	3
2	4
2	2
1	4
4	8
5	13

0	+1
+1	+2
0	0
-1	+2
0	+5

3. Establishment of Clusters

Nowhere is a concise dictum for a good cluster stated.

"Good clusters" are what the following procedure says they are.

- i All two member Strong Circles are established as good clusters
- ii Membership for candidates for these "core clusters" is established as follows:

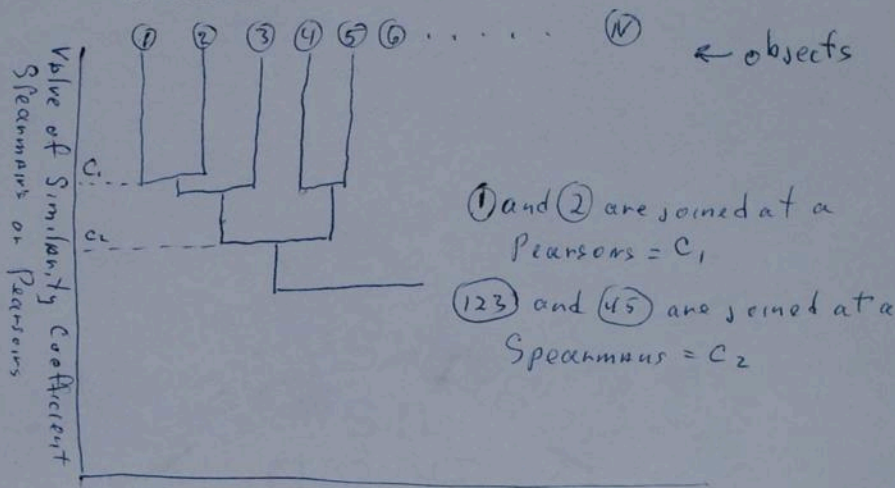
1. The average Pearson's correlation from the candidate to each of the core members is computed.
2. Spearman's coefficient between core clusters is computed as follows. A and B are clusters. A and B have N and M Objects.

$$S_{AB} = \frac{\sum_{a \in A, b \in B} r_{a,b}}{\left[\left(N + 2 \sum_{a \in A, c \in A} r_{a,c} \right) \left(M + 2 \sum_{b \in B, d \in B} r_{b,d} \right) \right]^{1/2}}$$

3. For each core cluster the largest relating coefficient as computed in 1, and 2, is noted, and if this coefficient is not more than .03 lower than the average* coefficient in the core cluster, then this candidate specimen or cluster is admitted to the core cluster. This is done until no more candidates can be admitted.

* There are alternate methods in use at this point. The one described, or the new Spearman's, can be compared to the specimens used at the previous level, for a test to see if the value dropped more than .03.

- iii A new matrix of similarities is now made between all clusters and points made above, using Spearman's where appropriate.
- iv All clusters are now considered as core clusters and, two member Strong Circles are isolated (members may now be clusters as well). A candidacy procedure is now done similar to that described in (ii, 3) above until no more clusters come together. Steps iii and iv are repeated at this point, until all objects have been grouped together.
- v The information is summarized in a pitchfork diagram as follows



4. Interpretation

Your guess is as good as mine

PROCESSING TAXONOMIC PROBLEMS

