
**Some Thoughts on Bacterial Classification**

BY P. H. A. SNEATH

**National Institute for Medical Research, London, N.W. 7**

**SUMMARY:** Scientific classification is virtually a branch of mathematics which describes the overall similarities of organisms. Catalogues do not do this. Many schemes of bacterial taxonomy are not classifications but catalogues. Similarity is best measured by the number of features in common between two strains, while division into taxa is based on correlated features. Other criteria for these two basic ideas are unsatisfactory whether any nonsense can make it much worse than it is. Hierarchical systems are a practical necessity, and simple mathematical methods are useful in bacterial classification. It is not necessary to know the evolutionary history of organisms in order to classify them in a scientific manner.

It might perhaps be best if this paper were given the subheading of ‘Issued but not published’ following the distinguished precedent of *An Unofficial Interpretation of the International Rules of Zoological Nomenclature* (1955). But whether published or not, bacterial classification is in such confusion that it seems doubtful whether any nonsense can make it much worse than it is. These reflections are an attempt to re-think the logic of bacterial classification. They present one point of view, perhaps carried to an extreme. None of the ideas is new; they are implicit in the work of many bacteriologists (e.g. Winslow & Rogers, 1906; Levine, 1918; Topley & Wilson, 1929; White, 1937) and particularly in the important and prophetic paper of Rogers, Clark & Davis (1914) which has been generally overlooked. These ideas underlie much of the matter read at the Discussion Meeting of the Society for General Microbiology (1954). They have long been discussed in the taxonomy of higher organisms (see Conclusions). Although in devising a particular taxonomic scheme some balance must be struck between the ideal and the practical, it is ideal scientific classifications which are considered in this paper, with an attempt to define the lines along which ideal systems might be sought. Scientific classification is used here in the sense of the ‘natural’ classifications of Bather (1927) or the ‘general’ classifications of Turrill (1952), indicating orthodox taxonomy; while the term catalogue approximates to Bather’s ‘artificial’ classifications and to Turrill’s ‘special’ classifications.

**A PHILOSOPHY OF CLASSIFICATION**

A classification is greatly influenced by the purpose for which it is devised. It may sometimes be useful to classify wild flowers by their colour—if one is teaching school children—but no one would submit this as a scientific classification. What are the principles which make us say that such a scheme is not scientific? We may first consider two purposes of classification in general:
Bacterial classification

(i) to give names or numbers to things, which is better called enumeration or cataloguing; (ii) to indicate similarity, and thus to increase our ability to think about and to use our observational material.

The giving of numbers to strains or to antigenic types is enumeration. It is questionable whether this is classification, or if so whether it is usually scientific (Cowan, 1955). It has, of course, great utility. Thus, if scientific classification were abolished, workers on food poisoning would immediately give letters or numbers to the big groups of bacteria which concerned them; group A for Clostridium botulinum and so on. They would classify strains by placing them into these groups, but the groups themselves would be enumerated—a catalogue. The groups would only be classified when someone pointed out that group A strains were closely related to group F (Clostridium welchii) but not to group B (Salmonella).

The underlying assumption of scientific classification is that there is a natural order, a system of similarities, which can be discovered by investigation. In its simplest form it says that A is very similar to B but is less similar to C. The distinctive concept of scientific classification is that it attempts to measure overall similarity; its distinctive value is in prediction. For example, we classify a mouse as a mammal because we believe it has sufficient overall similarity to other creatures which we already call mammals. The value of this course is that we can predict many features of any creature which is called a mammal. Other classifications do not contain such additional implied information; a classification of vertebrates into male and female, long and short, black and white, tells us nothing more, and it would never be proposed as a scientific classification. The overall similarity is measured by the number of similar features.

This assumption may be taken as axiomatic but it has a logical basis. Theoretically every feature would seem to be of equal weight for estimating overall similarity and for the following reasons. (1) Unless we prove that any particular feature should carry greater weight than another we must either give equal weight to every feature or else give weight on unproven or irrational grounds such as giving weight to features whose initial letter is A. (2) When we do consider that a feature has greater weight than another it is because it is correlated with other features, or because we believe that it ‘sums up’ a number of unknown features, which also implies correlation. (3) The greater weight is thus derived from correlation, and if we find an exceptional individual the feature is here no longer correlated with its usual congeners and, for classifying that individual, it loses its heavier value. Weighting is thus inseparable from correlation (and this also implies division). It is not the concept of number which is difficult but how one should estimate the number (see Conclusions).

The operative word here is number. Overall similarity is the concept of the sharing of many features, not of any one particular feature. Thus in Leviticus (Lev. xi. 19) the bat is catalogued with the birds because it can fly; the modern zoologist classifies it as a mammal, and his reasons are convincing, since the bat resembles other mammals in many features and resembles the
birds in but few. This concept does not consider evolution but it is not invalid on that account (see Conclusions).

_catalogues._ The trouble with catalogues is that they are endless, arguable and of poor predictive value. They appear precise, but are vague, while paradoxically good classifications may appear vague but are precise. No one can prove that one catalogue is any 'truer to nature' than another. The enumerator is soon driven to defining 'groups' and to predicting other features of new strains because they fall into one of his groups. Before long he is back at devising schemes of similarity. For practical purposes, therefore, since we cannot easily think about catalogues or remember them, we are driven back to classification.

_on similarity and division_

There is a distinction to be made at this point between expressing similarity and dividing groups of organisms. A series of strains may vary in the amount of an enzyme which they produce, and this can be expressed in mathematical terms if one wishes, giving a set of figures which are a measure of similarity in this respect. When we start to split up this series of strains our aim is to group the infinite array of possible figures so that we can conveniently think of them. Now a classification can in theory be made by listing similarities without any attempt to separate the strains into taxonomic groups but in practice we need some device for compressing the information, which is why we use hierarchical systems of classification, with different ranks of taxa to indicate roughly the degree of similarity. God can consider every creature as an individual—we have to group them into such things as genera and species. These are artificial concepts; they have no actual existence. Only creatures exist (Adanson, 1763, Prêf. p. clxiii); Pirie (1955) has well summed this up.

_separation._ We are faced with the practical necessity of separating groups of organisms. How shall we do so? We may first consider the position of higher organisms. The history of classification abundantly shows the uncertainty of intuition in deciding what features are 'important' in classification (witness bats and whales) yet we all know of features that are of little importance. Let us consider how one separates two groups, for instance cats from dogs. There are features which we know are very variable and unimportant in such a separation, such as colour. We call them 'unessential' features. But how do we tell 'essential' features? No single feature will do. If one found a dog which had a cat's incisor teeth, it would indeed be a surprise! But this would not make the creature a cat, for we would thereafter relegate it to the 'unessential' features. The reason for regarding a feature as 'essential' is simply that it always goes with some other feature. Consider the position if this were not so, and if there were all possible combinations of the various features of cats and dogs: we should be unable to separate the two kinds of animal because there would in fact be no dividing line. This is associated in higher organisms with genetic exchange, such as the interbreeding of the breeds of dog. In bacteria there is probably a great deal more genetic exchange than we are aware of at present, which may be one reason why we often find intermediate forms between bacterial 'species'.
In higher organisms we also find intermediate forms, but this does not shake our faith in classification (that is to say degrees of similarity). It is simply an expression of the fact that relationship is an infinitely variable quantity. Our divisions are those of convenience. It is obvious that we place the lines of division where the resultant groups will hold the greatest content of information. There are 'points of rarity' in distributions of features where individuals with certain combinations of features are rare (for instance the platypus is close to the point of rarity between mammals, birds and reptiles) and since we are studying the world as it is we will place the divisions at the points of rarity. No one wishes to classify non-existent organisms, that is to say, those which are infinitely rare. This course is also the convenient one, for in order to think about groups of bacteria, to make mental concepts, to arrange groups, taxa, species or what not, we are forced for reasons of mental economy to consider chiefly the commoner forms. If there is no point of rarity, it is useless to make a division, since we say 'these cannot be separated'. Indeed they can be separated, and sometimes are, in a purely arbitrary fashion, as when we divide cats into white cats, brown cats and so on, but this is not a measure of the overall similarity between various cats, and may, by its very arbitrariness, obscure real relationships of a subspecific kind.

This leads to the concept of a taxon as a group of strains with many features in common but not necessarily always possessing any particular feature—what may well be called a 'species-group' or a biotype. Unfortunately these terms have other meanings; e.g. a biotype is defined, in the International Bacteriological Code of Nomenclature (1948) Recommendation 8a (2), as a physiological type for which physiotype might be a better word but for which no other suitable term seems available; and it seems to be necessary to create new names for the kind of groupings found when considering bacteria. Existing terms based on the degree of interbreeding (see Turrill, 1952) are not suitable. It is desirable to distinguish between the reactions of strains and of groups of strains. The ending-form is suggested for the former, and the ending-on (as an abbreviation of taxon) for the latter. The following situation, frequent in bacteria, may be envisaged. Two groups of strains occur (the strains within each group being very similar to one another), and each group might be termed a pleiston (from the Greek πλείστος, very many). The most common type within a pleiston—perhaps only a hypothetical strain—would be the modoform of the pleiston, and the other strains composing the pleiston might be called sychnoforms (from the Greek συχνός, frequent). The groups are connected by uncommon intermediate forms (interforms), and this connecting group if it needed a separate name might be termed a gephyran (from the Greek γέφυρα, a bridge). There may also be less common groups of strains clustered around the pleistons and connected to them by interforms: such groups might be called oligons (from the Greek δλιγός, scanty). The oligons are each composed of a modoform and of sychnoforms. A good example is the Bacillus cereus–Bacillus megaterium group (Smith, Gordon & Clark, 1946). The typical strains of B. cereus would form one pleiston and the typical strains of B. megaterium would form another pleiston, the B. cereus–B. megaterium
intermediates would be interforms, while anthrax bacilli would form an oligon attached to the \textit{B. cereus} pleiston. Although further work may lead to better separation of the organisms given in this example, the scheme is useful at a certain stage of knowledge. As Smith, Gordon \& Clark (1952) remark, the Rules of Nomenclature do not adequately cover situations such as this. For instance, if the anthrax bacillus is regarded as a variety of \textit{B. cereus}, according to the rule of priority the former should be named \textit{B. anthracis} var. \textit{anthracis} and the latter \textit{B. anthracis} var. \textit{cereus}. The two criteria used for defining these new terms are the degree of similarity and the commonness of the strains. These terms are not intended to have any taxonomic rank, since this is the reason for coining them instead of using terms such as species and variety. We may conjecture that this situation is found at a low level in classification, at species and genus level, and may be likened to the occasional leaves on the evolutionary tree which are partly fused to an adjoining leaf.

The alternative is to define a taxon as a group whose members all possess a specified key feature. Such features should be definable, recognizable and rigidly employed, yet they can neither be defined nor recognized before creating the taxon, and they are not in any event rigidly employed. If, for instance, the mammary gland is taken as the key feature of the mammals, how does one define it or recognize it before first making the taxon Mammalia, and is a mouse with congenital absence of mammary glands to be excluded from the Mammalia? A key feature is therefore not a feature for creating a classification, but is one which is useful for easy identification afterwards. There would seem, therefore, to be only one way in which we can decide if a feature is 'important' and that is by seeing whether it goes closely with other features.

Let us examine in bacteria some consequences of the principle of Division by Correlated Features. I do not know whether it has ever been explicitly formulated in bacteriology, though Rogers \textit{et al.} (1914) came close to doing so. The trouble with any other method is how to decide on the relative 'importance' of features. The importance is composed of two interdependent things—the usefulness for making a division and the content of information in the divisions thus made. We may consider a hypothetical genus of bacteria which is studied by four successive authors. If the first author divides a genus arbitrarily on nitrate reduction and on lactose fermentation then every strain will go into one of the four 'species' so formed. Of course, all sorts of bacteria can be fitted in, and not infrequently all sorts of bacteria are. The second author divides the genus on reactions with maltose and with citrate. Who is to decide which is the right (or, not to beg the question, the better) classification? Is the first author to be followed because he is first?—or the author who uses the commoner tests?—or the one who uses tests in use in a nearby genus?—or the one who follows the text-book which we prefer?—or the one who uses a feature of economic as opposed to academic importance? The third author proposes eight tests and makes 256 'species'; we feel that this is too much. Which tests shall we discard and on what grounds? Now if all these tests are 'unimportant' tests (on a par with coat colour in cats and dogs) then these questions are insoluble because these tests are in fact useless for dividing the
Bacterial classification

The disagreements spring from the assumption that the tests must be important, but no proof is offered for this assumption. This is but fighting with shadows. Such a system—really a catalogue—not only fails to show up the relationships clearly, but may conceal them. It is to react like the Eskimo who comes to Britain and decides that white cats and sheep are polar bears.

To follow our hypothetical genus, the fourth author devises some new tests J, K, L, M, N and finds that all strains are J+ K+ L+ M- N- or are J- K- L- M+ N+, that all pathogenic strains show the first set of reactions and are natural inhabitants of the bowel, and he achieves two things. He shows that the genus is separable into species, and he compresses into his classification much information of a useful kind; he has separated cats from dogs. Division by correlated features is also the best solution to the problem of the classification of intermediate forms, since if this method fails, then any arbitrary method will also fail.

The Enterobacteriaceae

A notable example of the difficulty in classifying a group of bacteria is the Enterobacteriaceae. These organisms show a bewildering array of features in almost all combinations, so that they cannot be easily and neatly divided. Kauffmann, Edwards & Ewing (1956) somewhat pessimistically reject the Linnaean hierarchy, apparently because it is Linnaean, not because it is hierarchical. But does anyone now accept Linnaean concepts of species, and is this not a different thing from Linnaean binomial monomenclature? If a scheme be rejected because it is hierarchical, then one has tacitly admitted that strains cannot be grouped into taxa of larger size. The above authors, however, admit that they have to use various groupings for practical reasons. They abandon the ‘artificial and arbitrary classification into “genera” and “species”’ and instead use biochemical groups, subgroups and types. In what way are the latter any less artificial or arbitrary? Or are the other difficulties they mention the reason for this change? Their main point seems to be summed up in the sentence “under these circumstances...it is impossible to decide whether a given classification is scientifically correct, but only whether it is more or less practical and expedient”. Although it is not clear to me just what they mean by ‘scientifically correct’, this is at any rate a candid admission of our poor knowledge, and is far better than creating multitudes of bad taxonomic groups. The Enterobacteriaceae can only be better classified when more work is done on correlated features and the frequency of intermediate forms. It seems probable (as has happened in other groups which are largely parasitic, e.g. some fungi—see Ainsworth, 1955) that in the past we have in the Enterobacteriaceae made our species too small (Cowan, 1956a); we may be trying to separate the inseparable.

PRACTICAL ASPECTS OF CLASSIFICATION

Avoidance of preconceived notions. When collecting strains for study we should not be too selective, or too obsessed with the best tests to use. With all of us there is a tendency to exclude a strain because it differs in a ‘cardinal’
feature. There is little interest and no significance in a statement such as ‘All streptococci are Gram-positive’ if every Gram-negative organism is resolutely discarded. Some limit must be set but it is better to cast the net widely, and to exclude on several features, rather than risk discarding an aberrant streptococcus which is Gram-negative. Some taxonomists appear to exhibit strong intuitions about the value of certain tests, but it is difficult to think of any test which is invariably reliable in the separation of bacteria. The Gram reaction, flagellation, carbohydrate metabolism, pathogenicity, aerobiosis, anaerobiosis, antigenic structure—all these have proved to be unreliable in some genus or another, either because genuine exceptions occur, or because the conditions of testing are so critical that in practice one gets false results. It is well to be sceptical about well-established tests even in groups where they have proved their value; the proof may not be very sound.

**Definition of taxa and dichotomous keys**

If we accept the importance of divisions based on correlated features, what happens when we summarize our classification in order to define our taxa? We will use in our generic definition those features which are common to all strains. Of the other features we will obviously choose those with high correlation for dividing the genus into species, thus:

\[
\text{Genus is } A + B + C - D -
\]

\[
\text{Species 1 is } E + F + G - H -
\]

\[
\text{Species 2 is } E - F - G + H +
\]

To indicate the range of variability, we will add ‘tests I, J, K, L variable in both species’. A common complaint is that generic definitions contain features qualified by the word ‘usually’. There is some substance in such complaints, since this implies that the features listed are not well correlated, but from the theoretical point of view, every feature must be so qualified. If enough well-correlated features are used, the description will be adequate. One improvement on the usual way of listing features may be suggested: features can be given in two sections, (i) those which are almost always found and (ii) those which are less constant. The problem becomes most acute when we try to make the dichotomous keys so beloved of text-books. Shall we use more than one feature? If we use one feature, then the exceptional strain will be misclassified; if we use several features the exceptional strain will not fit in and the key may become unwieldy.

Single-feature keys are pernicious since they suggest that the group can be sharply divided. If one must have a key it is better to use several features, since the user can then see that his strain is aberrant in a single test, and it is possible that he will make the logical deduction that he has an aberrant strain. Single feature keys look neat, but the user thinks he has made a mistake (which he may have done) or that he has a new species (he is probably wrong), or else (and this time he is right) that the key is nonsense. The formal logical way is to use several features for the key and to add ‘intermediate forms may
Bacterial classification

occur'. Keys which are qualified by exceptions to cover likely difficulties presuppose a thorough knowledge of the likely aberrancies. The most successful general key of this kind is that of Skerman (1949).

Conditions of test. One practical difficulty is always arising, and that is the standardization of tests. This may be illustrated by the example of growth temperatures. If we have two species with different optimum temperatures should we do all our tests at one temperature, or test each species at its optimum? Our chief difficulty is that we do not know how growth at non-optimal temperatures will affect the results and we have not the time to investigate this. Individual strains will also have different optima. I feel that our practice should be devised with an eye for the future worker. He cannot repeat the tests unless the instructions are precise, and he may be too lazy to find out the optimum temperature for each strain and to marshal an array of incubators. It is best to use one temperature for both species, for the risk of all our tests being vitiated seems less than the risk of making our work unrepeatable. Cowan (1956a) recently made a forceful plea for standardization of descriptions and reform of nomenclature.

The use of mathematics in devising a plan of work

When one studies a group of bacteria one naturally begins by collecting strains and listing the tests to be used. Commonly one is faced with the dilemma—shall I do a lot of tests with a few strains, or a few tests with many strains? Some help may be obtained from a very elementary consideration of probability theory. We wish to find how to separate a group into two or more species, and make these assumptions:

(1) The majority of our strains belong to one species (Species 1) but we have in the batch of strains which we pick for testing at least two strains which belong to a different species (Species 2). At this stage we cannot tell which strains are Species 2, and the probability that a strain picked at random from those available will belong to Species 2 may be called \( p_s \).

(2) We must discover at least two tests which show absolute correlation in dividing the two species, for example Species 1 is \( A + B + \) and Species 2 is \( A - B - \), but no strain is \( A + B - \) or \( A - B + \). The probability that any test picked at random from available tests will prove a ‘good’ test in this sense, may be called \( p_t \).

We may now ask what are the chances that we shall ‘hit the jackpot’—i.e. discover that there are two or more strains of the second species which are separable on two or more ‘good’ tests. This problem is a variation of the problem of finding two or more successes in a random sample of \( n \) trials when the possibility of success per trial is \( p \). The required probabilities can be obtained from the binomial expansion, and some of them are tabulated in Table 1. In Table 1 \( p \) can stand for \( p_s \) and \( n \) for the number of strains chosen (giving the chance of including two or more Species 2 strains), or else \( p \) can stand for \( p_t \) and \( n \) for the number of tests employed (giving the chance of finding two or more ‘good’ tests). To solve our problem we must succeed in both of these respects, and the overall chance will be the product of the chances of each individual eventuality. An example may be given: we are sent a
P. H. A. Sneath

thousand strains of streptococci, which are mostly enterococci, but twenty
are group A streptococci. We pick one hundred strains at random, and there-
fore \( n = 100 \) and \( p_s = 0.02 \), and the chance of having two or more group A
streptococci in this selection of a hundred strains is 0.60 (ignoring the 'two-
tailed' distribution of \( P_s \)). We have a hundred tests available, and of these
five are 'good' tests, so that \( p_t = 0.05 \). We choose fifty tests at random, and
the chance of picking two or more good tests is 0.72. We test the hundred
strains in the fifty tests: the chance of showing that there are two kinds of
streptococcus on these results will be \( 0.60 \times 0.72 = 0.43 \), about an even chance.

Table 1. The probability of finding two or more successes in \( n \) independent
trials when the probability of success in a single trial is \( p \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( p_s )</th>
<th>( 0.5 )</th>
<th>( 0.2 )</th>
<th>( 0.1 )</th>
<th>( 0.05 )</th>
<th>( 0.02 )</th>
<th>( 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>100</td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>0.25</td>
<td>0.81</td>
<td>0.99</td>
<td>1.00</td>
<td>0.97</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>0.26</td>
<td>0.26</td>
<td>0.61</td>
<td>0.96</td>
<td>0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.01</td>
<td>0.08</td>
<td>0.26</td>
<td>0.26</td>
<td>0.06</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.06</td>
<td>0.26</td>
<td>0.99</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2. The probabilities of 'hitting the jackpot' are given under the headings
(a), (b), and (c), where (a) is 10 strains in 100 tests, (b) is 32 strains in
32 tests, (c) is 100 strains in 10 tests for various values of \( p_s \) and \( p_t \)

<table>
<thead>
<tr>
<th>( p_s )</th>
<th>( p_t )</th>
<th>( (a) )</th>
<th>( (b) )</th>
<th>( (c) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.99</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.99</td>
<td>0.84</td>
<td>0.28</td>
</tr>
<tr>
<td>0.5</td>
<td>0.05</td>
<td>0.95</td>
<td>0.48</td>
<td>0.09</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5</td>
<td>0.26</td>
<td>0.84</td>
<td>0.99</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.26</td>
<td>0.71</td>
<td>0.26</td>
</tr>
<tr>
<td>0.1</td>
<td>0.05</td>
<td>0.25</td>
<td>0.40</td>
<td>0.09</td>
</tr>
<tr>
<td>0.05</td>
<td>0.5</td>
<td>0.09</td>
<td>0.48</td>
<td>0.95</td>
</tr>
<tr>
<td>0.05</td>
<td>0.1</td>
<td>0.09</td>
<td>0.40</td>
<td>0.25</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.09</td>
<td>0.28</td>
<td>0.09</td>
</tr>
</tbody>
</table>

If the values of \( p_s \) and \( p_t \) are both 0.5 we shall quickly get the answer, while
if either is less than 0.01 we are unlikely to do so without hundreds of trials.
The practical range will therefore be 0.5 to 0.01. Here is an example. We
decide to use only 1000 culture tubes, and can do (a) 10 strains in 100 tests;
(b) about 32 strains in 32 tests; (c) 100 strains in 10 tests. The probabilities
that we will 'hit the jackpot', though they vary with \( p_s \) and \( p_t \), are usually in
favour of the second course (b). In Table 2 are given the probabilities from this
eample for some values of \( p_s \) and \( p_t \). It is seen that course (a) or course (c) is
seldom very much better than course (b), while the superiority of course (b) is
very evident when both \( p_s \) and \( p_t \) are small.

If we knew the value for \( p_s \) or \( p_t \) we should know what to do without the
aid of mathematics, but if we do not, then it seems wisest to choose the middle
Bacterial classification

course (b). The example given is greatly over-simplified, but the general statistical conclusions would seem to hold for more complex instances, particularly if we wish to find three or more species (or 'good' tests), when the individual probabilities become smaller for given values of $p_1$ or $p_2$.

We may therefore conclude that the answer to our dilemma is to choose a middle course. This is at variance with the teachings of those who emphasize the importance of using many strains. The importance of using enough tests is emphasized here, if need be by decreasing the number of strains. There is no shortage of possible tests; Clarke (1955) has suggested ways of devising a great many new tests. Another advantage of numerous tests may be mentioned. If we use many tests we will find more 'good' tests, and our division will be more firmly established—it is better to have two firmly established species than to find (by using many strains) four or five poorly established species. There are enough of the latter already cluttering the literature.

The use of mathematics in testing the significance of our findings

We start with the two assumptions made previously, namely, that we have at least two strains of Species 2 which is separated on two or more completely correlated tests. We make out a large table, with strains down the side and tests along the top and fill in the results. We see a pattern which suggests we have 'hit the jackpot'! Is this pattern due to chance? We might have chosen strains and tests which by chance gave the pattern we see. It is obvious that if we used four strains and two tests we could not place much reliance on the assertion that species 1 was $A+B+$ and species 2 was $A-B-$ . If, however, in 100 strains we had 50 $A+B+$ and 50 $A-B-$ strains it would be superfluous to do a test of significance. The best method of showing one's results is to publish the full table of all the results with every strain. Editors, wishing to save space, will object to this; they should be resisted if possible, for if we condense our table we may lose vital information—literally vital in the sense that our classification will live or die by the information we retain. It is for example very difficult to analyse the massive data on the Enterobacteriaceae, since the reactions of individual strains are usually not given. Most of the relevant information will be preserved if we use correlated features for our classification, but there may be imperfectly correlated data which are of great importance.

One way of condensing the table, which discards little information (and is a useful way of preparing the information for statistical analysis) is to take a pair of tests and count the numbers of strains positive in both, negative in both and positive with the first but not the second and vice versa, and arrange them in a square to give a $2 \times 2$ table—e.g. for 92 strains we may have:

\[
\begin{array}{c|c|c}
& \text{Test A} & \\
\hline
+ & 40 & 0 \\
\hline
- & 1 & 51 \\
\end{array}
\]
This is repeated for each way of pairing-off the tests, and arranged in a larger table as shown below. The last tests (A in the rows and E in the columns) are redundant and are omitted. The triangle gives all the possible ways of pairing the tests. It is a saving of space, and can accommodate the results of any number of strains. One can easily see the degree of correlation between tests, and calculate the probabilities where necessary. A further saving, though it causes loss of more information, is to omit the numbers of strains from such a table, and enter only the probabilities against the distribution in each square being due to chance. Alternatively, the information can be given in a form rather like this: 'Tables of contingency were made for all possible pairs of tests and the probabilities that the associations were due to chance were calculated. None showed values below 0.05 except for the pairs AB ($P = 0.001$), AC ($P = 0.01$) and BC ($P = 0.02$). The most logical method, to tabulate the number of similarities and differences between strains, is scarcely suited for publication, although it is the way in which the data would be analysed in a calculating machine, as is mentioned later.

The probabilities are not a measure of how close the correlation is, for an absolute correlation based on only a few strains may have the same probability as an imperfect correlation based on a large number of strains. We lose this information in any event. The probability is a measure of whether we are entitled on our results to say they are correlated at all.

Two tests will be found of general use—the $\chi^2$ test and Fisher's Exact Method for $2 \times 2$ tables. These tests are well known and simple to calculate and can be found in most elementary text-books of statistics. The best account of them for the non-mathematician is that of Race & Sanger (1954, pp. 59, 126, 272). The $\chi^2$ test is useful for tables with large figures, or with more than four values (such as $3 \times 8$ tables) provided no entry is less than 5. Fisher's method is useful where one entry is zero or a small number, and is easily calculated using a table of the logarithms of factorials. These tests are not readily misapplied, and are unlikely to give false results.

Having made a significant finding, one has to consider its cause. Statistics will not tell one this. The reason for the correlation may be that the two tests are really testing the same property—for instance caseinase and peptonization in litmus milk. The lack of one enzyme concerned with glucose metabolism might influence apparently unrelated things like the ability to utilize fructose and the ability to grow anaerobically. If two features are always associated throughout bacteria, this may be because they are two properties of the same thing. A similar argument applies to negative features; for instance, since no bacterium is known to require chloroform as a growth factor it would be superfluous to count chloroform-independence as a feature. The safeguard is to study enough features, which will make the uncertainty less important. For reasons of this sort, one should be reluctant to create a division on only two correlated features. Four or five such features (not all belonging
Bacterial classification

One more problem will arise, and that is how to express the probability that five or six tests, not just two, are correlated. This can be done if \( \chi^2 \) values are calculated, since it is permissible to add together the various values for \( \chi^2 \) and the various values for the degrees of freedom used in this test and then get the probability for \( \Sigma \chi^2 \) and \( \Sigma \) d.f. which expresses the overall probability. If need be, the \( \chi^2 \) values can be obtained by special methods with entries below 5.

One of the pioneers in the application of statistics to bacterial classification was Levine (1918), while Dible (1921) applied his methods in a classic paper. It is possible to express, as these authors did, the correlation in terms of coefficients of association, but this leads into a tangle of statistics, since the significance of the figures thus obtained may be doubtful. Such coefficients are usually unnecessary. The instances considered here are tests which score as positive or negative. Tests which give quantitative results can be analysed by the well-known Coefficients of Correlation. For such tests, a scatter diagram is very useful.

Justification. A justification of introducing mathematics into bacteriological classification may be needed. Mathematics and statistics are already firmly entrenched in things like viable counting, but they may seem foreign to the taxonomist. However, in another field, that of blood groups (which at first must have seemed a very non-mathematical subject) the use of methods as simple as those suggested here has caused an enormous expansion of valuable conclusions and of profitable experimentation.

CONCLUSIONS

The development of ‘natural’ systems of classification is well summarized by Bather (1927), Gilmour (1940) and Turrill (1942). The address of Bather is very valuable and he traces the subject from its inception by Linnaeus (1748; perhaps foreshadowed by Caspar Bauhin), through the systems of Buffon, de Candolle, de Jussieu, Cuvier and Whewell. However, these systems weighted the features according to their relative importance and this produced a source of uncertainty. Importance was never satisfactorily defined, but it clearly cannot be defined as importance in classification, for this is what we wish to discover.

The first taxonomist who conceived of the use of every feature impartially and with equal weight was Michel Adanson (1768). His views were far ahead of his time. He attacked the arbitrary methods of the day and insisted that every part of a plant should be used in making a classification (Préf. p. clv). He constructed in turn all the classifications which he could, each one based on a single part of the plants, and on comparing them he found that some of the classifications divided up the plants in a manner which was almost identical with that of other classifications. These he chose as indicating the most natural groupings, which were therefore groupings based on correlated features
of virtually equal weight. He also realized the nature of the dividing lines which were thus made (Préf. p. clxiv). Adanson's views were not highly esteemed, since his contemporaries considered that some features were more important than others, although they could never agree which features these should be. Adanson seems to me to have been right, for the concept of 'importance' seems to play little part in the concept of 'greatest content of information.' The viewpoint of Gilmour (1940) is close to that of Adanson. Gilmour, on epistemological grounds, reaches the same conclusion as is reached here, that the ideal classification is the one with the greatest predictive value. Such a concept is based on giving every feature equal weight. Such classifications may conveniently be called 'Adansonian'. The basic difficulties which remain are to define a 'unit character' and to handle quantitative data, and an attempt has been made to do this in the succeeding paper. In quantitative tests the phrase 'number of features' is not directly applicable, and it would seem necessary to take as units of measurement units of such size that they are in reasonable relation to the range of variation found in the material. We have, in fact, assumed the truth of this thesis by usually adjusting our tests so that the division lies near the centre of variability of the quantitative feature concerned. The use of 'number of features' may also be justified on the grounds that the quantitative differences depend ultimately on qualitative differences in genes or in nucleotides, to which the concept of number is applicable.

Similarity is better defined in terms of the number of features which are not known to be different properties of the same thing. This is to avoid a logical dilemma, e.g. if a single gene mutation causes many diverse effects shall we score this as one difference or as many correlated differences, and the mutant forms as one species or as two? Since the raison d'etre of a scientific classification is its high content of information (its predictive value) the two forms will be classified as one species, for this procedure gives a greater content of information—viz. it adds that one form can turn into the other (in a sense also they are similar in their potential features). The paradox is that the value of a correlation in classification is lost when our knowledge is sufficient to explain why this correlation exists, but it does not matter that what used to be counted as several correlated features is now counted as a single feature since the function of scientific classification is to enable us to make predictions about unknown and unexplained properties. Similar considerations make one classify insect larvae with the adult forms. However, this point may seem only of academic interest in bacteria.

The most startling new concept in classification is the 'dynamic system' of Hayata (1921). This likens the relationships of plants to a net or a sponge, and springs from the observation that there are many cross-relationships between very diverse families. It is an expression of the fact that similarity is multidimensional; there is one dimension for each feature and to represent it one requires multidimensional space, so that the net or sponge is an oversimplification. Hayata fully recognized that in practice the relationships must be summarized into a hierarchical system, but suggested that instead of a
linear array of families it is better to flank each family name with the names
(in small print) of the families which are most closely related, i.e. the nearby
knots of the net. Hayata's system is well discussed by Du Rietz (1930) and
Airy Shaw (1950) who believed that it may have a large part to play in botany.
It would also seem particularly appropriate to bacteria and will repay closer
study. It has been further developed (Hayata, 1981) to bring it more into
line with orthodox genetics.

No mention has been made of evolution, which some believe to be essential
to a scientific classification. But, as the Irishman said, there is a red herring
in the air and it must be nipped in the bud. One can make a classification
without considering phylogeny (indeed the outlines of the classification of
higher organisms were laid down before Darwin—see Bather, 1927; Airy
Shaw, 1950; Pirie, 1955), although it will necessarily be incomplete. The
phylogenist may amplify it and modify it in details. His ability to do this is
limited: he cannot deny that the degree of similarity is a good measure of
ancestral relationship for he must use this very criterion to choose the mem-
bers of an ancestral tree. The bacteriologist has no choice in the matter; those
who hanker after phylogenetic schemes in bacteria should read the critical
review of Sporne (1956) on this problem in angiosperms, in which the fossil
record is scanty. He rejects as invalid most of the criteria for primitive
features which have been proposed and concludes that a fossil record is
needed for the application of the others. The bacterial fossil record is sum-
marized by Pia (1928); some substitute may be provided by bacterial genetics
—what may be called 'micro-evolution'.

We have concluded that the number of features in common is the basis for
deducing similarity but that correlated features should be the basis of division
into taxa. Does this mean that an indivisible group should always be of
lower taxonomic rank than an unrelated but divisible group? I do not think
that it should. One species may be divisible into distinct subspecies and races,
while another may not and the causes may be various (lack of genetic exchange,
restriction to certain microenvironments, etc.). Since the concept of species
is an artificial one, there is no need to make divisibility a bed of Procrustes.
At present, we have almost nothing to guide us on the rank which should
be given to a taxon of bacteria—it is a matter of personal preference. The
theoretical method is to lay down that strains of one species should have a
certain percentage of common features, but until more work is done upon this
it is desirable to observe two working rules. The first is that a mutant
should not be given specific rank. Bacteriologists do not do this, but they are
sometimes tempted to give specific rank to strains which might be naturally
occurring mutants differing from the common type in only a single feature.
The second is that one should be chary of separating into different genera
bacteria which undergo genetic exchange.

Cowan (1955) posed some interesting questions which it is fashionable to
attempt to answer. The last three were admirably answered by Howie in
the discussion on Cowan's paper. The first two questions are: (1) Does the
hierarchical structure of a Linnaean system satisfy the requirements of
microbiologists? (2) Can we accept the species concept, and all that this implies, or must we view our organisms as a huge spectrum composed of gradually merging forms?

The two questions are inter-related, for if bacteria form a continuous spectrum then, since we cannot think in terms of such spectra, we are forced to use a hierarchical system of some kind: we may as well use the Linnaean system as any other, and we may as well use the term species where we can. In fact I do not believe that such a spectrum is so continuous that we can never summarize it.

The spectrum can, however, be summarized in different ways, as Cowan (1956b) pointed out, to suit the needs of various workers. van Niel (1946) also suggested a similar approach, but wisely did not call it taxonomy. These alternatives to the ideal classification are often not scientific classifications but are catalogues; in so far as each alternative approaches the ideal classification they will approach one another. They may be necessary steps to that ideal, but they are obsolescent, and should not be indiscriminately called taxonomies, but schemes of division, or keys for identification. A classification based on the number of similar features and their correlation, in other words the classification which would emerge if one fed one’s data into a calculating machine, would in theory be the best classification which could be made with the available data. One can instruct the machine to compare each strain with every other and tabulate the number of differences, and then place together into groups those strains with no differences, add to these groups the strains with one difference and so on. Similarity can then be given a numerical value and this allows taxonomic rank to be defined in terms of similarity, as is only logical. An attempt has been made to do this in the next paper (Sneath, 1957). In this sense classification is a branch of mathematics.

The practical man may complain that he does not want a classification based on correlated features (all qualified by the word usually); he just wants a test to identify a pathogenic bacterium. But, if there is no correlation between pathogenicity and any other feature, it simply means that no such test is known. No re-arrangement of the taxonomy will help him, and if he groups all pathogenic strains into one species, this species means no more than ‘pathogenic strains of genus A’. Is it heretical to suggest that it is simpler to say this rather than to use a Latin name?

Classifications based on the ideas discussed here might with profit be tried out on viruses and bacteriophages before the taxonomy of these becomes overloaded with preconceived notions of how to do it. Antigenic analysis will be largely used and this will cause some difficulties; but the only logical way would be to assume only such number of distinct antigens as are essential to explain the serological results, that is to use Occam’s Razor, and to use these results as the basis for counting similarities and differences.

Will divisions based on correlated features stand the test of time? For the answer we may consider a well-known example such as mammals and birds. New techniques in embryology, biochemistry, cytology and so on reinforce
Bacterial classification

the separation of Aves from Mammalia and the separation becomes more reliable and the classification of the vertebrates becomes more closely knit, precisely because the new discoveries correlate with the old knowledge. Similarly, in the genus *Bacillus* the work of Knight & Proom (1950) on nutritional requirements has reinforced the taxonomy of Smith et al. (1946). A bad classification is pulled apart by new discoveries.

Any interest which this may possess is due to stimulating discussions—and arguments—with Drs J. Mandelstam, S. T. Cowan, M. R. Pollock and B. A. D. Stocker and many others, whose help is gratefully acknowledged.

REFERENCES


200

P. H. A. Sneath


(Received 26 February 1957)