Journal of General Microbiology (1983), **129**, 1815–1830. *Printed in Great Britain*

A Probability Matrix for Identification of some Streptomycetes

By S. T. WILLIAMS,[1]* M. GOODFELLOW,[2] E. M. H. WELLINGTON,[3] J. C. VICKERS,[1] G. ALDERSON,[4] P. H. A. SNEATH,[5] M. J. SACKIN[5] AND A. M. MORTIMER[1]

[1] *Department of Botany, University of Liverpool, Liverpool L69 3BX, U.K.*
[2] *Department of Microbiology, University of Newcastle, Newcastle upon Tyne NE1 7RU, U.K.*
[3] *Department of Biology, Liverpool Polytechnic, Liverpool L3 3AF, U.K.*
[4] *School of Studies in Medical Sciences, University of Bradford, Bradford BD7 1DP, U.K.*
[5] *Department of Microbiology, University of Leicester, Leicester LE1 7RH, U.K.*

(*Received 15 October 1982*)

The character state data obtained for clusters defined at the 77·5% S_{SM} similarity level in the phenetic numerical classification described by Williams *et al.* (1983) were used to construct a probabilistic identification matrix. The 23 phena included were the major clusters (19 *Streptomyces*, 2 *Streptoverticillium* and '*Nocardia*' *mediterranea*) and one minor cluster (*Streptomyces fradiae*). The characters most diagnostic for these clusters were selected using Sneath's CHARSEP and DIACHAR programs. The resulting matrix consisted of 41 characters × 23 phena.

Identification scores, determined by Sneath's MATIDEN program were used to evaluate the matrix. Theoretical assessment was achieved by determination of the cluster overlap (OVERMAT), the identification scores for the Hypothetical Medium Organism of each cluster (MOSTTYP), and the scores for randomly selected cluster representatives using the classification data of Williams *et al.* (1983). The matrix was evaluated practically by the independent re-determination of the characters for the same cluster representatives, which also provided a measure of test error. Finally it was used to identify unknown isolates from a range of habitats.

The results showed that the matrix was theoretically sound. Test error was within acceptable limits and did not distort identifications. Of the unknown isolates, 80% were clearly identified with a cluster. It is suggested that the matrix could form the basis for a more objective identification and grouping of the large number of *Streptomyces* species which have been described.

INTRODUCTION

An identification matrix for bacteria consists of the selected taxa together with their most diagnostic characters, which are ideally expressed as the percentage positive states for each character within each taxon. When this information is stored in a computer, it can be used for the probabilistic identification of unknown strains, which provides a measure of the likelihood that the identification is correct. In such a system, no one property is essential for taxon membership and hence the identification is polythetic in the sense used in numerical taxonomy (Sneath & Sokal, 1973). This is a most appropriate way to deal with the identification of bacteria where results of any one test can vary and where there are not always clear-cut distinctions between taxa of lower rank. Probabilistic schemes contrast with the widely used monothetic sequential keys which are especially susceptible to test error (Sneath, 1974). Reviews of the theory and practice of probabilistic identification of bacteria have been given by Sneath & Sokal (1973), Lapage (1974), Hill (1974), Sneath (1974, 1978) and Willcox *et al.* (1980).

Abbreviations: HMO, Hypothetical median organism; OTU, Operational taxonomic unit.

0022-1287/83/0001-0839 $02.00

Numerical classification not only defines clusters (phena) at selected levels of overall similarity, it also provides quantitative data on the test reactions of strains within each of the defined clusters. This is expressed as the percentage of strains within each cluster which show a positive state for each character used to construct the classification. Such data provide an ideal basis for the construction of an identification matrix (Sneath & Sokal, 1973; Hill, 1974; Sneath, 1978), which contains the minimum number of characters needed to discriminate between the clusters. Hill (1974) suggested that workers engaged in numerical taxonomy should add numerical identification sections to their classifications. However, surprisingly few numerical classifications of bacteria have been supported by probabilistic identification systems, possibly due to the problem of carrying out reproducibility studies on tests of presumptive diagnostic value. One of the few exceptions was the matrix produced by Wayne *et al.* (1980) for identification of slowly growing mycobacteria. On the other hand, most probabilistic identification systems have been constructed using data less comprehensive than those provided by a thorough numerical classification. Such studies include those on Gram-negative aerobic rods (Bascomb *et al.*, 1973; Lapage *et al.*, 1973; Willcox *et al.*, 1980), anaerobic bacteria (Kelley & Kellogg, 1978), Gram-negative, fermentative rods (Schindler *et al.*, 1979), *Bacillus* species (Willemse-Collinet *et al.*, 1980), nitrogen-fixing soil bacteria (Rennie, 1980), Gram-positive, aerobic cocci (Feltham & Sneath, 1982), and aerobic, Gram-negative, fermentative rods (Lefebvre & Gavini, 1982; Gavini *et al.*, 1982).

Most attempts to group the several hundred *Streptomyces* species described have been made using a few subjectively chosen criteria, usually morphological and pigmentation characters, to construct species, species-groups or 'series' (see Williams *et al.*, 1983). Such schemes are artificial and provide identifications of little general predictive value. It has also been shown by factor analysis that the taxa defined in this way were diffuse and often overlapped considerably due to the use of a few characters which were variable and prone to test error (Gyllenberg *et al.*, 1967; Gyllenberg, 1970). However, several attempts have been made to construct a numerical identification scheme for streptomycetes. The data from the first numerical classification of streptomycetes (Silvestri *et al.*, 1962) were used to construct a probabilistic identification key (Hill & Silvestri, 1962; Möller, 1962). Tests included in the key were selected by a method which ranked them hierarchically on the basis of their information content. The data included in the re-descriptions of type cultures of *Streptomyces* species in the International *Streptomyces* Project (ISP) (Shirling & Gottlieb, 1968*a*, *b*; 1969, 1972) were used to produce a matrix for computerized identification of streptomycetes (Gyllenberg, 1970, 1974, 1976; Gyllenberg *et al.*, 1975). Fifteen species groups were selected *a priori* and the ISP data provided 24 character states. The theoretical assessment of this matrix suggested that it provided a more objective means of analysing the ISP data than the construction of dichotomous keys.

A workable, polythetic identification system for *Streptomyces* species is, therefore, still required. The numerical classification of streptomycetes (Williams *et al.*, 1983) provided a suitable data base for the construction of an identification matrix and this has been attempted for the major clusters defined in that study.

METHODS

Details of the origins, cultivation and preservation of strains, together with the definition and composition of clusters, were given by Williams *et al.* (1983). Binomials in inverted commas are not cited in the *Approved Lists of Bacterial Names* (Skerman *et al.*, 1980). Genus names in inverted commas indicate that the generic status of the strain is still uncertain.

Selection of clusters. The clusters used for the construction of the identification matrix were the 22 major ones defined at the 77·5% S_{SM} level (Williams *et al.*, 1983), which consisted of 19 *Streptomyces*, 2 *Streptoverticillium* clusters and the '*Nocardia*' *mediterranea* cluster. The latter three clusters were included because they consisted of strains which can be easily confused with streptomycetes on superficial examination. In addition, although it contained only two strains, the *Streptomyces fradiae* minor cluster was included, as strains of this species are important antibiotic producers. A total of 309 strains was contained in these 23 clusters.

Selection of characters the most diagnostic of the clusters. A selection was made from the 139 unit characters used to construct the classification matrix, to determine the minimum number of characters necessary for discrimination between the 23 clusters.

As a first step, the number of clusters in which each character was $\geq 85\%$ positive or $\leq 15\%$ negative was determined, the product of these scores providing the separation index (S_i) of Gyllenberg (1963). Tests with high scores were of most general diagnostic value. A further selection of characters was achieved using the CHARSEP program (Sneath, 1979*b*), which calculates the usefulness of different characters for separating groups in which percentage positive values for the character states are known. Several separation indices are calculated for each character and those of poor diagnostic value (low scores) are readily determined. The separation indices included are Gyllenberg's (1963) sum of $C_{(i)}$ and rank measure $R_{(i)}$, the index of Niemalä *et al.* (1968), the VSP developed from Sneath & Johnson (1972) and an analagous index, CSP. The character selection was next checked by applying the DIACHAR program (Sneath, 1980*a*), which ranks the diagnostic scores of each character for each group in an identification matrix and also provides the sum of scores of all characters for each group. High scores are thus desirable. Finally, a few tests were included which, though of low overall separation value, were shown by DIACHAR (Sneath, 1980*a*) to be diagnostic for a particular cluster.

Determination of characters included in the identification matrix. As it is hoped that other workers may wish to use the identification system, full practical details for the determination of the 41 characters included in the matrix are given, although general details were covered by Williams *et al.* (1983). The character numbers quoted are those used in the identification matrix (Table 2). Unless stated, the inoculum consisted of spore or mycelial suspensions stored in 10% (w/v) glycerol at − 20 °C (Wellington & Williams, 1978). However, other forms of inoculum can be used. All incubations were at 25 °C.

(*i*) *Spore chain morphology* (*character numbers 1–4*). This was determined on a starch/mineral salts medium containing (l^{-1}): 10 g soluble starch, 2 g $(NH_4)_2SO_4$, 2 g $CaCO_3$, 1 g K_2HPO_4, 1 g NaCl and 20 g agar (Küster, 1959), adjusted to pH 7·0. The starch was prepared as a paste with a little distilled water, made up to 500 ml and added to the other constituents dissolved in 500 ml distilled water, before autoclaving. The medium was poured into Petri dishes and two sterile coverslips were inserted at an angle of about 45° into each plate while the medium was semi-molten. Spore or mycelial suspensions were inoculated on to the junction between the upper surface of the coverslips and the medium. After 14 d incubation, one of the coverslips was carefully removed and placed on to a microscope slide for examination of spore chains by light microscopy at × 400 magnification. Spore chain shape was described, according to the categories of Pridham *et al.* (1958), as rectiflexibiles (RF), retinaculiaperti (RA), spirales (S) or verticillati (V).

The same preparations were used to detect fragmentation of the non-sporing substrate mycelium (character 5) growing over the surface of the coverslip.

(*ii*) *Spore surface ornamentation* (*6–7*). This was determined using the second coverslip from the cultures used in (i). The coverslip was cut with a glass file and a suitable fragment with growth on it was mounted on a specimen stub, coated with gold-palladium under vacuum, and examined at magnifications of 10000 and 20000 with a scanning electron microscope. Spore surface ornamentation was categorized, according to Tresner *et al.* (1961) and Dietz & Mathews (1971), as smooth, warty, spiny, hairy or rugose.

(*iii*) *Pigmentation* (*8–13*). The colour of the spore mass (8–10) was determined using cultures on the starch/mineral salts medium incubated for 14 d. Plates were inoculated in a cross-hatched pattern to promote development of spores. After incubation, a segment of the agar supporting good, confluent sporing growth was removed and placed near to the end of a microscope slide. The spore colour was then determined by matching the segment against the tabs in the colour wheels of Tresner & Backus (1963). The colour wheel containing the tab giving the closest match was recorded. These comparisons should be made under conditions of standard illumination. This can be best achieved by using a white-painted box with an open front, illuminated from above by a source of illuminant C (fluorescent batten, Thorn Electricals, Thorn House, London, U.K.) which approximates to average daylight. If colour wheels are not available, spore mass colour should be categorized by eye, in standard conditions of illumination, as grey, red, green, or other (white, yellow, blue or violet); these colours were not sufficiently diagnostic to be included in the matrix.

Colour of substrate mycelium (11–12) was determined using cultures grown at 25 °C for 14 d on a medium containing (l^{-1}): 10 g glycerol, 1 g L-asparagine, 1 g K_2HPO_4 and 20 g agar, adjusted to pH 7·0. Segments bearing confluent growth were removed from the plates, inverted and excess medium removed from them to expose the substrate growth. A inverted segment was then placed at the end of a slide and the reverse colour matched to the tabs in the colour scheme of Prauser (1964), under conditions of standard illumination. Colours were categorized as yellow-brown (no distinctive pigment), red-orange or other. Soluble pigments of the same colour as the substrate mycelium are produced by some cultures. If the colour scheme is not available, these determinations may be made by eye. Workers lacking experience in examination of streptomycetes should seek expert advice before attempting to categorize the colour of spores and substrate mycelium by eye.

Production of a melanin pigment (13) was determined by multiple-loop inoculation of strains into 25-compartment Repli dishes (Sterilin, Teddington, Middlesex) containing peptone iron agar (Difco) adjusted to pH 7·0. The presence or absence of a melanin pigment was recorded after 4 d. Typically this is a brown to black pigment, but the basic colour may be slightly modified if other soluble pigments are produced.

(*iv*) *Utilization of carbon sources* (*14–22*). Appropriate weights of each carbon source were sterilized in sterile 50 ml flasks fitted with loose fitting bungs by covering the solid with diethyl ether which was allowed to evaporate at room temperature in a fume cupboard for 18 h. Final traces of ether were removed by placing the flasks in a laminar flow cabinet for 10 min. Alternatively, carbon source solutions may be sterilized by filtration. Each of the nine sterilized carbon sources was added, aseptically, to give a concentration of 1% (w/v), to a basal medium containing (l^{-1}): 2·64 g $(NH_4)_2SO_4$, 2·38 g KH_2PO_4, 5·65 g $K_2HPO_4 . 3H_2O$, 1·0 g $MgSO_4 . 7H_2O$, and 15 g agar, adjusted to pH 7·0 (modified from Pridham & Gottlieb, 1948). Repli dishes were inoculated with dry spores or mycelium from cultures grown on starch/mineral salts medium (any other medium which promotes sporing may be used); glycerol suspensions are clearly not suitable as inoculum for these tests. Cultures were incubated for 21 d. The ability of a strain to use a carbon source was determined by comparison of its growth with that on the positive control (1%, w/v, glucose) and the negative control (no carbon source), the latter being the more useful aid in making the decision.

(*v*) *Utilization of nitrogen sources* (*23–25*). Nitrogen sources were ether-sterilized (or filter-sterilized) and added aseptically to give a concentration of 1% (w/v) in a basal medium containing (l^{-1}): 10 g glucose, 1 g K_2HPO_4, 0·5 g $MgSO_4 . 7H_2O$, 0·5 g NaCl, and 15 g agar, adjusted to pH 7·0. Repli dishes were inoculated and incubated for 15 d. Utilization of the nitrogen source was determined by comparison of the growth with that in the positive control (1%, w/v, L-asparagine) and the negative control (no nitrogen source), the latter being the more useful.

(*vi*) *Degradation* (*26–32*). Allantoin degradation (26) was studied in tubes containing (l^{-1}) 0·1 g yeast extract (Oxoid), 9·1 g KH_2PO_4, 9·5 g $NaHPO_4$, 3·3 g allantoin, 0·01 g phenol red, and 7·5 g agar, pH 6·8 (Gordon, 1968). Control tubes without allantoin were also inoculated. After 28 d, a positive result was shown by a change of the indicator from orange-yellow to pink, red or purple.

Arbutin degradation (27) was studied in tubes containing (l^{-1}): 3 g yeast extract (Oxoid), 1 g arbutin, 0·5 g ferric ammonium citrate, and 7·5 g agar, pH 7·2 (Kutzner, 1976). Controls without arbutin were also inoculated. After 21 d, a positive result was indicated by a brown-black pigment. Comparison with controls is essential to avoid confusion with melanin production.

Xanthine degradation (28) was determined by inoculating pairs of organisms as single streaks on to plates containing (l^{-1}): 1 g yeast extract (Oxoid), 0·8 g Lab-lemco (Oxoid), 10 g glycerol, 2 g NZ amine A (or a similar casein digest), 4 g xanthine and 15 g agar, pH 7·0. After 21 d, a clear zone around the growth indicated a positive result.

Pectin degradation (29) was studied by spot-inoculation of each plate with nine strains using a simple multi-point inoculator. The medium used was that of Hankin *et al.* (1971) and consisted of three parts which were prepared and autoclaved separately. Part A contained 6 g Na_2HPO_4 and 4 g KH_2PO_4 in 200 ml distilled water; part B consisted of 5 g pectin in 200 ml distilled water, and was steamed well before autoclaving; part C contained 2 g $(NH_4)_2SO_4$, 0·001 g $FeSO_4 . 7H_2O$, 1 g yeast extract (Oxoid), 0·2 g $MgSO_4 . 7H_2O$, and 10 g agar in 600 ml distilled water. The final pH of the medium was 7·4. Plates were incubated for 6 d and then flooded with a warmed solution of 1% (w/v) hexadecyltrimethylammonium bromide and left for 1 h. Excess reagent was decanted and hydrolysis zones appeared as clear areas around colonies.

Lecithin degradation (30) was determined on plates spot-inoculated with nine strains. The basal medium contained (l^{-1}): 10 g peptone (Oxoid), 5 g yeast extract (Oxoid), 1 g glucose, 10 g NaCl and 12 g agar, pH 7·0. Sterile egg yolk emulsion (Oxoid) was added to the sterile molten medium at 45 °C to give a concentration of 5% (v/v). After incubation for 6 d, the plates were viewed under strong light and a positive reaction was indicated by an opaque, creamy-yellow precipitate around colonies. Care must be taken to distinguish between this and lipolysis which gives a less opaque precipitate in a zone with a more irregular margin.

Nitrate reduction (31) was studied in tubes containing nutrient broth (Oxoid) supplemented with (l^{-1}): 2 g KNO_3 and 6 g agar, pH 7·0. After 14 d, nitrite production was detected by addition of 0·2 ml each of Griess–Ilosvay reagents I and II. Tubes were shaken well and observed for development of a pink-red colour over a period of 30 min. A trace of zinc dust was added to tubes with a negative reaction; development of a red colour after shaking confirmed the presence of nitrate.

Hydrogen sulphide production (32) was determined in the cultures used for the nitrate reduction test by insertion of a strip of sterile lead acetate paper into the mouth of each tube (Küster & Williams, 1964). After 14 d, any browning or blackening of the paper indicated a positive reaction.

(*vii*) *Growth inhibition* (*33–38*). Tests 33–36 were carried out in Repli dishes containing (l^{-1}): 1 g yeast extract (Oxoid), 0·8 g Lab-lemco (Oxoid), 10 g glycerol, 2 g NZ amine A or a similar casein digest, and 15 g agar, pH 7·3. Each potential inhibitor was added to this basal medium before autoclaving to give the appropriate final concentration; inhibitors were sodium azide (0·01%, w/v), NaCl (7%, w/v) and phenol (0·1%, w/v). After 14 d, growth was compared with that on the medium without the inhibitor; absence of growth or very weak growth was recorded as negative.

Growth at 45 °C (36) was determined in Repli dishes using the same basal medium. Plates were incubated for 14 d, enclosed in a moistened polythene bag to prevent desiccation, and presence or absence of growth was noted.

Resistance to neomycin (37) and rifampicin (38) (50 μg ml^{-1}) was also determined on this basal medium. Discs

(5 mm diameter) of filter paper (Whatman no. 1) were soaked in a filter-sterilized antibiotic solution at the selected concentration for 2 min; they were then freeze-dried for 90 min and stored in sealed vials at 4 °C. Test strain inoculum was spread over the surface of plates of basal medium which were then dried to remove excess moisture. Antibiotic discs (four per plate) were placed on to the plates using sterile forceps. Any definite inhibition zone was noted after 24–30 h; resistance was recorded as positive.

(*viii*) *Antibiosis* (*39–41*). Plates of nutrient agar (Oxoid) in glass Petri dishes were spot-inoculated (nine strains per plate) and incubated for 24–28 h to produce growth which was, as far as possible, non-sporing. Plates were then inverted, 1 ml chloroform was added to the lids and left to evaporate in a fume cupboard for 40 min to kill the colonies; plates were left open for 15 min in a laminar flow cabinet to remove remaining vapour. They were then overlaid with a molten 'sloppy' agar (Oxoid nutrient broth with 0·6%, w/v agar) at 45 °C, previously seeded with the test organism. The test organisms used were *Bacillus subtilis* (National Collection of Industrial Bacteria, Aberdeen, U.K.; NICB 3610), *Aspergillus niger* (Department of Botany, Liverpool University, Liverpool, U.K.; LIV 131) and *Streptomyces murinus* (International *Streptomyces* Project, ISP 5091). Overlaid cultures were examined for zones of inhibition after 24 h, which were recorded as a positive reaction.

Theoretical evaluation of the identification matrix. (*i*) *Assessment of cluster overlap.* This was achieved using the OVERMAT program (Sneath, 1980*c*), designed to determine overlap between groups in a matrix constructed using percentage positive values for character states. For each pair of groups, a disjunction index (W) and a corresponding nominal overlap (V_G) is calculated, the latter ranging from 1·0 for complete overlap to 0 for complete disjunction. The significance of the determined overlap is assessed, using a non-central t-statistic, against a selected critical overlap value (V_O). In this case the chosen critical value was 5%.

(*ii*) *Determination of identification scores.* These were obtained, in both the theoretical and practical assessment of the matrix, using the MATIDEN program (Sneath, 1979*a*), which provides the best identification scores for known or unknown strains against a matrix consisting of q taxa and m characters. Percentages in the matrix, with 0 changed to 1 and 100 to 99% (Lapage *et al.*, 1970), are converted by the program to proportions, P_{iJ} for the ith character of taxon J. The character state values of an unknown (u) are input and compared with each taxon in turn, identification coefficients are calculated and the best identifications printed. Only the m characters scored as + or − are considered, not those for which the value of u is unrecorded. Three of the identification coefficients provided by the program were selected.

1. Willcox probability (L_{uJ}) (Willcox *et al.*, 1973). This is the likelihood of u against taxon J divided by the sum of the likelihoods of u against all q taxa, i.e. $L_{uJ}/\Sigma^q L_{uJ}$. The nearer the score approaches 1·0, the better is the fit of an unknown with a group in the matrix.

2. Taxonomic distance (d) which is given by $\sqrt{[\Sigma(u_i - P_{iJ})^2/m^1]}$. This expresses the distance of an unknown from the centroid of the group with which it is being compared; hence low scores indicate relatedness.

3. The standard error of the taxonomic distance, which equals the constant (c) in the equation $d = \bar{d}_J + cs_{dJ}$, where $\bar{d}_J$ is the mean distance of OTUs of taxon J from the centroid and s_{dJ} is the standard deviation of those distances. An acceptable score is less than about 2·0–3·0 and approximately half the members of a taxon will have negative scores, indicating that they are closer to the centroid than average.

The program provides identification scores to the best group and the two next best alternatives. In addition, it gives properties of the unknown which are atypical of the best group ('characters against') and other properties which may help to distinguish the unknown from the two next best groups (Table 3).

(*iii*) *Determination of identification scores for the Hypothetical Median Organism* (*HMO*) *of each cluster.* This was achieved using the MOSTTYP program (Sneath, 1980*b*) which calculates the best identification scores achievable by an entirely typical example of each group in a matrix.

(*iv*) *Determination of identification scores for cluster representatives using data from the numerical phenetic classification.* The next stage in the evaluation of the matrix was to input the results of the appropriate tests obtained during the construction of the numerical classification (Williams *et al.*, 1983). Identification scores were obtained for a randomly selected member of each of the 23 clusters included in the matrix.

Practical evaluation of the identification matrix. (*i*) *Re-determination of character states of selected cluster representatives.* The same randomly selected representatives of each group were used. The character states of cultures, coded to avoid revealing their names, were re-determined by a worker not previously experienced in streptomycete taxonomy. The results were used to obtain identification scores for each strain. This also provided an opportunity to assess the degree of intra-laboratory error for the tests included in the matrix, by comparison of the original and re-determined results. Numbers of discrepancies were noted and the test variances (S_i^2) calculated from $S_i^2 = d/2t$, where d is the number of strains giving a discrepant result and t is the number of strains tested (Sneath & Johnson, 1972). The mean of test variances was used to calculate the average probability of an erroneous test result (Sneath & Johnson, 1972).

(*ii*) *Identification of unknown isolates.* A total of 64 isolates from various soil and water samples, which had been sent in for identification by various workers, was used for the final testing of the matrix. Character states were determined and identification scores obtained. Cluster identifications were made for those isolates with scores considered to be acceptable.

Table 1. *Examples of results obtained by CHARSEP to determine the separation values of characters*

Characters	No. of clusters in which character is predominantly:		Separation indices				
					Gyllenberg (1963)		Niemela *et al.* (1968)
	+ve	−ve	VSP (%)	CSP	$\Sigma C_{(i)}$	$\Sigma R_{(i)}$	index
Examples of characters selected for the matrix:							
Spore chain spirales	7	6	54·85	0·55	19·01	801·36	7·45
Spore mass grey	12	3	40·47	0·52	19·63	706·82	6·12
Production of melanin	11	3	38·59	0·52	19·80	653·43	5·90
Use of DL-α-aminobutyric acid	8	3	33·38	0·40	18·23	437·59	5·12
Degradation of lecithin	13	4	48·57	0·58	19·87	1003·24	7·77
Degradation of xanthine	5	9	53·76	0·59	20·02	900·81	7·60
Antibiosis to *Bacillus subtilis*	4	5	44·43	0·44	18·29	365·84	4·84
Resistance to neomycin (50 μg ml^{-1})	14	4	41·99	0·58	20·70	1159·14	8·03
Resistance to phenol (0·1%, w/v)	6	8	55·83	0·57	19·48	935·14	8·01
Use of D-mannitol	4	17	42·05	0·61	21·68	1474·14	8·70
Examples of characters rejected for the matrix:							
Spore surface hairy	21	0	0·32	0·13	22·12	0	0
Spore mass blue	22	0	0·20	0·10	22·37	0	0
Production of blue soluble pigment	22	0	0·06	0·07	22·56	0	0
Use of L-arginine	0	19	1·32	0·22	21·25	0	0
Proteolysis	0	14	3·02	0·27	20·45	0	0
Degradation of testosterone	1	14	8·18	0·33	20·62	288·72	2·71
Resistance to streptomycin (50 μg ml^{-1})	16	1	11·81	0·37	20·95	335·23	2·83
Growth at 37 °C	1	13	11·04	0·35	19·48	253·24	2·64
Growth with NaCl (13%, w/v)	21	1	5·04	0·25	22·26	467·50	3·09
Use of L-arabinose	1	13	8·42	0·40	19·52	253·79	2·64

RESULTS

Selection of characters the most diagnostic of the clusters

Calculation of S_i separation indices (Gyllenberg, 1963) provided an initial indication that some of the 139 characters used to define the 23 clusters (Williams *et al.*, 1983) were of limited diagnostic value. The CHARSEP program (Sneath, 1979*b*), incorporating five other separation indices, provided further evidence of diagnostic values (Table 1), high values indicating those tests of most potential value. The examples presented give some idea of the range of values obtained. A few tests with high values were excluded due to practical difficulties in their determination. On the other hand, a few characters with low values (e.g. VSP $< 25\%$) were selected because it could be seen by eye that they were useful for recognition of particular clusters; these included biverticillate spore chains (for clusters 55 and 58) and green spores (for cluster 37).

The DIACHAR program (Sneath, 1980*a*) provided a final assessment of the 41 tests selected. The sums of values for characters in each cluster, which should be high if the cluster is well defined by those characters, ranged from 9·363 to 18·563. This was regarded as acceptable, but the lower values indicated that a few clusters were less clearly defined than the others. Therefore, the final identification matrix consisted of 23 clusters × 41 characters (Table 2).

Theoretical evaluation of the identification matrix

(*i*) *Assessment of cluster overlap.* It is clearly desirable that overlap between groups in an identification matrix should be minimal. The OVERMAT program (Sneath, 1980*c*) provided an assessment of overlap in the constructed matrix. At the selected critical overlap value (V_O) of 5%, there was no significant overlap between any of the pairs of clusters.

(*ii*) *Determination of identification scores.* An example of the print-out provided by the MATIDEN program (Sneath, 1979*a*), which was used to determine identification scores, is given in Table 3. This is included to emphasize the diversity of information which should be considered before making a positive identification.

(*iii*) *Identification scores for the Hypothetical Median Organism of each cluster* (*Table 4*). The identification scores obtained by the MOSTTYP program (Sneath, 1980*b*) can be regarded as the best possible scores attainable within each cluster. A good matrix should therefore give excellent scores for all coefficients of the HMO against its own taxon. The results were satisfactory, with Willcox probabilities of 0·999 or 1·000 (except for *S. diastaticus*), low taxonomic distances (0·137–0·252, the latter for *S. diastaticus*) and values for the standard errors of taxonomic distance all negative.

(*iv*) *Identification scores for cluster representatives using data from the classification study* (*Table 5*). As expected, scores were somewhat less impressive than those for the HMOs. Nevertheless, they were generally satisfactory; all strains identified to their clusters, the majority doing so with good identification scores. The representative of cluster 20 ('*Elytrosporangium*' *brasiliense*) had a noticeably low Willcox probability with high taxonomic distance and standard error. It was presumably a somewhat atypical member of this cluster.

Practical evaluation of the identification matrix

(*i*) *Re-determination of character states of selected cluster representatives* (*Table 6*). Generally there was little deterioration of the identification scores compared with those obtained using the original classification data (Table 5). Exceptions were *Streptomyces griseus* (cluster 1), *S. umbrinus* (cluster 5) and '*Elytrosporangium*' *brasiliense* (cluster 20); these still identified to their clusters but with reduced scores.

The assessment of intra-laboratory test error, provided by comparison of these independently determined results with those from the classification study, is summarized in Table 7. Overall agreement was high, with 55 discrepancies in 943 determinations (5·8%), a mean variance of 0·029 and a mean probability of error of 3·0%, which is below the acceptable limit of 5·0% for test error within a laboratory (Sneath & Johnson, 1972). The maximum number of discrepancies was four, with 26 characters having either one or none.

Table 2. *A percentage positive probability matrix for streptomycete clusters*

Cluster	Spore chains V	Spore chains RA	Spore chains RF	Spore chains S	Fragmentation of mycelium	Spores smooth	Spores rugose	Spores grey	Spores red	Spores green	Reverse yellow-brown	Reverse red-orange	Melanin	Adonitol	Cellobiose	D-Fructose	*meso*-Inositol	Inulin
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. *Streptomyces albidoflavus*	1	13	83	9	1	99	1	20	3	1	96	1	7	51	99	93	37	10
3. *S. atroolivaceus*	1	11	99	1	1	99	1	67	1	1	56	1	1	11	89	78	1	1
5. *S. exfoliatus*	1	1	99	1	1	99	1	41	41	1	89	11	61	1	99	56	6	33
6. *S. violaceus*	1	13	38	50	1	63	1	13	13	1	99	1	88	13	99	99	63	1
10. *S. fulvissimus*	1	22	67	22	1	99	1	13	75	1	11	89	67	89	99	99	99	11
12. *S. rochei*	1	40	4	64	1	48	1	83	8	4	89	8	4	35	99	89	96	27
15. *S. chromofuscus*	1	1	22	78	1	67	1	43	1	1	99	1	33	22	99	99	89	1
16. *S. albus*	1	1	1	99	1	99	1	1	1	1	86	1	1	99	99	67	33	1
17. *S. griseoviridis*	1	1	1	99	1	99	1	1	99	1	99	1	1	67	99	99	67	17
18. *S. cyaneus*	1	13	5	82	1	41	1	32	32	1	69	21	97	82	95	90	92	80
19. *S. diastaticus*	1	11	42	58	5	99	1	53	18	1	79	16	47	16	99	99	84	63
20. *S. olivaceoviridis*	1	14	1	86	1	99	1	71	14	1	71	1	14	14	99	86	57	43
21. *S. griseoruber*	1	1	1	99	1	89	1	88	13	1	1	67	22	22	99	99	99	11
29. *S. lydicus*	1	1	1	99	1	64	1	91	9	1	99	1	1	82	73	91	91	18
32. *S. violaceoniger*	1	1	1	99	1	1	99	99	1	1	99	1	1	67	99	83	67	33
37. *S. griseoflavus*	1	99	1	1	1	1	1	1	17	83	99	1	17	1	99	83	83	17
40. *S. phaeochromogenes*	1	33	1	67	1	99	1	1	1	1	99	1	17	99	99	99	99	99
42. *S. rimosus*	1	1	14	57	1	86	1	1	1	1	99	1	1	99	99	86	99	1
53. *'Nocardia' mediterranea*	1	1	38	13	99	38	13	1	1	1	25	75	1	13	99	86	99	17
55. *Streptoverticillium griseocarneum*	99	1	1	1	1	99	1	11	67	1	99	1	78	11	1	99	78	13
58. *Stv. blastmyceticum*	99	1	1	1	1	99	1	1	40	20	99	1	40	1	20	60	40	11
61. *S. lavendulae*	1	1	92	17	1	99	1	8	83	1	99	1	99	8	83	75	25	1
68. *S. fradiae*	1	99	1	50	1	99	1	1	99	1	99	1	1	1	1	1	1	1

Table 2. (*continued*)

Cluster	Mannitol	Raffinose	L-Rhamnose	D-Xylose	DL-α-Aminobutyric acid	L-Histidine	L-Hydroxyproline	Allantoin	Arbutin	Xanthine	Pectin	Lecithin	NO_3^- reduction	H_2S production	Sodium azide (0·01%, w/v)	NaCl (7·0%, w/v)	Phenol (0·1%, w/v)	Growth at 45 °C	Neomycin resistance	Rifampicin resistance	*Aspergillus niger* antibiosis	*Bacillus subtilis* antibiosis	*Streptomyces murinus* antibiosis
1.	94	17	62	93	54	65	23	83	99	97	45	6	55	92	56	83	90	7	1	54	32	28	39
3.	89	22	78	67	33	78	11	44	99	99	89	11	89	99	1	44	56	1	11	89	11	33	38
5.	1	33	61	89	61	78	89	83	99	94	61	50	83	89	23	22	72	17	11	11	6	56	39
6.	38	63	38	99	88	25	88	75	99	88	13	75	99	63	63	38	99	1	25	99	25	50	88
10.	99	89	22	99	89	99	78	78	68	99	1	44	1	67	1	22	44	22	89	99	22	78	99
12.	99	69	96	96	12	77	8	50	96	96	42	4	27	92	62	92	96	77	8	89	27	35	39
15.	99	22	67	99	67	78	1	33	99	22	22	11	22	89	56	44	22	67	1	33	1	11	22
16.	99	33	17	67	1	99	67	99	99	83	1	1	1	83	99	99	17	99	1	99	17	17	17
17.	99	51	83	83	1	83	17	17	99	83	67	50	33	99	99	83	83	67	1	83	33	99	80
18.	95	99	92	90	31	85	28	54	54	80	56	10	36	90	15	18	64	41	1	46	10	44	62
19.	90	84	95	90	32	68	21	32	53	53	68	1	47	79	5	32	95	16	1	68	11	21	5
20.	99	99	99	43	57	99	1	14	1	29	14	1	14	99	57	29	86	86	1	86	1	1	1
21.	99	99	99	89	33	99	22	78	99	78	99	1	89	99	11	78	89	1	1	78	1	1	22
29.	91	82	18	27	9	36	55	18	99	82	1	64	9	1	18	55	9	1	18	9	99	73	99
32.	99	83	83	67	99	99	11	50	99	1	50	1	83	99	50	1	1	50	1	83	33	67	50
37.	99	33	83	99	17	17	17	1	50	1	99	1	83	99	38	33	99	1	1	50	1	17	67
40.	99	86	99	99	67	83	67	99	67	99	50	1	17	83	67	83	99	67	1	50	1	17	1
42.	99	86	1	86	1	99	29	86	71	86	14	86	86	14	99	99	71	43	99	99	99	99	99
53.	99	63	88	86	1	86	38	63	13	1	13	1	38	99	1	1	1	1	88	99	1	38	13
55.	11	11	33	11	11	99	78	1	1	56	1	99	1	22	11	1	1	11	99	33	99	78	89
58.	20	20	1	1	20	40	1	1	1	1	1	60	1	48	1	20	1	99	80	1	80	99	80
61.	8	8	17	33	42	8	42	50	92	83	8	99	50	42	1	1	58	17	50	33	75	92	99
68.	1	50	99	99	1	1	1	1	99	1	50	99	50	1	1	1	1	99	50	1	50	99	50

Table 3. *Example of the output provided by the MATIDEN program to identify an unknown strain against the identification matrix*

INPUT THE CHARACTER VALUES OF THE UNKNOWN

REFERENCE NUMBER OF UNKNOWN IS P571

ISOLATE P571 BEST IDENTIFICATION IS, S. ALBIDOFLAVUS

SCORES FOR COEFFICIENTS: 1 (Willcox probability), 2 (Taxonomic distance), 3 (Standard error of taxonomic distance)

	1	2	3
S. ALBIDOFLAVUS	0·999	0·276	−1·653
S. CHROMOFUSCUS	$0{\cdot}671 \times 10^{-5}$	0·420	2·362
S. ATROOLIVACEUS	$0{\cdot}379 \times 10^{-5}$	0·420	3·099

CHARACTERS AGAINST S. ALBIDOFLAVUS - NONE

ADDITIONAL CHARACTERS THAT ASSIST IN SEPARATING S. ALBIDOFLAVUS FROM:
S. CHROMOFUSCUS - NONE
S. ATROOLIVACEUS - NONE

Table 4. *Identification scores for the Hypothetical Median Organism of each cluster provided by the MOSTTYP program*

Cluster*	Identification score: Willcox probability	Taxonomic distance	Standard error of taxonomic distance
1. *Streptomyces albidoflavus* (71)	0·999	0·237	−2·701
3. *S. atroolivaceus* (9)	0·999	0·202	−3·167
5. *S. exofoliatus* (18)	0·999	0·240	−2·744
6. *S. violaceus* (8)	1·000	0·234	−2·813
10. *S. fulvissimus* (9)	1·000	0·188	−3·615
12. *S. rochei* (26)	0·999	0·223	−3·037
15. *S. chromofuscus* (9)	0·999	0·215	−3·172
16. *S. albus* (7)	1·000	0·137	−3·967
17. *S. griseoviridis* (6)	0·999	0·190	−3·192
18. *S. cyaneus* (38)	0·999	0·251	−2·850
19. *S. diastaticus* (20)	0·974	0·252	−2·739
20. *S. olivaceoviridis* (7)	0·999	0·191	−3·251
21. *S. griseoruber* (9)	0·999	0·133	−4·449
29. *S. lydicus* (11)	1·000	0·183	−3·607
32. *S. violaceoniger* (6)	1·000	0·225	−2·162
37. *S. griseoflavus* (6)	1·000	0·177	−3·346
40. *S. phaeochromogenes* (7)	0·999	0·189	−2·926
42. *S. rimosus* (7)	1·000	0·143	−3·944
53. *'Nocardia' mediterranea* (7)	1·000	0·180	−3·501
55. *Streptoverticillium griseocarneum* (9)	1·000	0·149	−4·020
58. *Stv. blastmyceticum* (5)	0·999	0·191	−3·044
61. *S. lavendulae* (12)	1·000	0·225	−2·760
68. *S. fradiae* (2)	1·000	0·207	−0·699

* The figures in parentheses are the number of strains in the cluster.

(*ii*) *Identification of unknown isolates.* The criteria adopted for a successful identification were:
(*a*) a Willcox probability greater than 0·850 with low scores for taxonomic distance and its standard error;
(*b*) first group scores substantially better than those against the next best two alternatives;
(*c*) 'characters against' should be zero or few.
Examples to illustrate the range of scores obtained for identified and non-identified isolates are given in Table 8. The scores for the identification coefficients generally followed the expected pattern, with those for taxonomic distance, and its standard errors (and 'characters against') increasing as the Willcox probabilities decreased. Of the 64 isolates originating from a variety of

Table 5. *Identification scores for cluster representatives obtained using classification tests data*

Cluster no.	Cluster representative*	Identification score: Willcox probability	Identification score: Taxonomic distance	Identification score: Standard error of taxonomic distance
1.	*Streptomyces griseus* ISP 5236	0·988	0·304	−0·914
3.	*'S. scabies'* ISP 5078	0·999	0·286	−0·760
5.	*S. umbrinus* ISP 5278	0·924	0·345	−0·027
6.	*S. violaceus* ISP 5082	1·000	0·281	−1·557
10.	*S. spectabilis* ISP 5512	1·000	0·285	−0·820
12.	*S. griseorubens* ISP 5160	0·999	0·255	−2·172
15.	*S. argenteolus* ISP 5226	0·985	0·304	−0·765
16.	*S. albus* ISP 5313	0·999	0·243	−0·060
17.	*'S. rosa'* ISP 5533	0·999	0·277	−0·544
18.	*S. chartreusis* ISP 5085	0·999	0·376	0·219
19.	*S. diastatochromogenes* ISP 5449	0·993	0·385	0·563
20.	*'Elytrosporangium' brasiliense* MO86	0·879	0·341	1·273
21.	*S. violaceolatus* ISP 5438	0·999	0·230	−1·119
29.	*'S. fasiculatus'* ISP 5054	1·000	0·294	−0·332
32.	*S. violaceoniger* ISP 5563	1·000	0·274	−0·682
37.	*S. hirsutus* ISP 5095	1·000	0·283	0·039
40.	*'Chainia' nigra* ISP 5302	0·999	0·329	1·574
42.	*S. albofaciens* ISP 5268	1·000	0·262	0·256
53.	*'Nocardia' mediterranea* N730	0·999	0·358	1·957
55.	*Streptoverticillium cinnamoneum* ISP 5005	0·998	0·285	0·550
58.	*Stv. griseoverticillatum* ISP 5507	0·999	0·203	−2·659
61.	*S. lavendulae* ISP 5069	1·000	0·352	0·751
68.	*S. fradiae* ISP 5063	0·999	0·338	0·581

Table 6. *Identification scores for cluster representatives obtained by re-determination of classification test data*

Cluster no.	Cluster representative*	Identification score: Willcox probability	Identification score: Taxonomic distance	Identification score: Standard error of taxonomic distance
1.	*Streptomyces griseus* ISP 5236	0·856	0·313	−0·668
3.	*'S. scabies'* ISP 5078	0·997	0·266	−1·326
5.	*S. umbrinus* ISP 5278	0·641	0·360	0·355
6.	*S. violaceus* ISP 5082	0·999	0·311	−0·760
10.	*S. spectabilis* ISP 5512	1·000	0·266	−1·385
12.	*S. griseorubens* ISP 5160	0·999	0·289	−1·250
15.	*S. argenteolus* ISP 5226	0·984	0·304	−0·765
16.	*S. albus* ISP 5313	0·999	0·288	1·591
17.	*'S. rosa'* ISP 5533	0·999	0·262	−0·994
18.	*S. chartreusis* ISP 5085	0·982	0·292	−1·864
19.	*S. diastatochromogenes* ISP 5449	0·999	0·339	−0·582
20.	*'Elytrosporangium' brasiliense* MO86	0·749	0·346	1·425
21.	*S. violaceolatus* ISP 5438	0·997	0·291	0·987
29.	*'S. fasiculatus'* ISP 5054	0·999	0·326	0·614
32.	*S. violaceoniger* ISP 5563	1·000	0·243	−1·635
37.	*S. hirsutus* ISP 5095	0·999	0·310	0·915
40.	*'Chainia' nigra* ISP 5302	0·999	0·330	0·720
42.	*S. albofaciens* ISP 5268	1·000	0·262	0·256
53.	*'Nocardia' mediterranea* N730	0·999	0·294	−0·021
55.	*Streptoverticillium cinnamoneum* ISP 5005	0·999	0·294	0·836
58.	*Stv. griseoverticillatum* ISP 5507	0·999	0·237	−1·621
61.	*S. lavendulae* ISP 5069	1·000	0·368	1·201
68.	*S. fradiae* ISP 5063	0·999	0·301	−3·080

* Binomials in inverted commas are not cited in the *Approved Lists of Bacterial Names* (Skerman *et al.*, 1980). Genus names in inverted commas indicate that the generic status of the strain is still uncertain. For strain origins see Williams *et al.* (1983).

Table 7. *Assessment of test error by independent re-determination of character states for selected cluster representatives*

Character	No. of discrepancies	Variance
1. Spore chains RF 2. Spore chains RA 3. Spore chains S 4. Spore chains V 5. Fragmentation of mycelium 6. Spore surface smooth 7. Spore surface rugose 9. Spore colour red 10. Spore colour green 11. Reverse yellow-brown 12. Reverse red-orange 13. Melanin production 17. Use of *meso*-inositol 19. Use of mannitol 29. Degradation of pectin	0	0
8. Spore colour grey 15. Use of cellobiose 19. Use of inulin 20. Use of raffinose 21. Use of L-rhamnose 22. Use of D-xylose 24. Use of L-histidine 26. Degradation of allantoin 28. Degradation of xanthine 32. H_2S production 35. Tolerance of phenol (0·1%, w/v)	1	0·044
16. Use of D-fructose 27. Degradation of arbutin 30. Degradation of lecithin 31. Nitrate reduction 36. Growth at 45 °C	2	0·044
14. Use of adonitol 25. Use of L-hydroxyproline 33. Tolerance of sodium azide (0·01%, w/v) 37. Resistance to neomycin (50 μg ml^{-1}) 39. Antibiosis to *Aspergillus niger* 41. Antibiosis to *Streptomyces murinus*	3	0·065
23. Use of DL-α-aminobutyric acid 34. Tolerance of NaCl (7·0%, w/v) 38. Resistance to rifampicin (50 μg ml^{-1}) 40. Antibiosis to *Bacillus subtilis*	4	0·087

Total test discrepancies = 5·8%; mean test variance = 0·029; mean probability of error = 3·0%.

habitats, 52 (81·3%) were identified to one of six clusters (Table 9). Most were placed into *S. albidoflavus* (the '*griseus*' group), which was the largest cluster defined in the numerical phenetic classification and is known to be common in soil and other habitats.

DISCUSSION

The broader relevance of this study to bacterial taxonomy is that it has demonstrated the application of some recent theoretical concepts in the development of a probabilistic identification scheme from numerical classification data. There are several essential stages in this process and each was achieved by using a specifically designed computer program. Selection of diagnostic tests, which cannot be accurately achieved by eye with a large amount of data, was made by applying the programs of Sneath (1979*b*, 1980*a*). Evaluation of the quality of

Table 8. *Examples of identification scores for unknown isolates*

Origin of isolates	Cluster identification	Identification score: Willcox probability	Identification score: Taxonomic distance	Identification score: Standard error of taxonomic distance	Characters against
Salt marsh	*S. albidoflavus*	0·999	0·287	−1·361	0
Indian soil	*S. albidoflavus*	0·999	0·337	−0·044	1
Salt marsh	*S. albidoflavus*	0·998	0·337	−0·046	1
Indian soil	*S. rochei*	0·998	0·341	0·126	2
Marine sediment	*S. diastaticus*	0·998	0·352	−0·260	2
Marine sediment	*S. rochei*	0·992	0·347	0·270	2
Indian soil	*S. rochei*	0·980	0·366	0·798	2
Salt marsh	*S. albidoflavus*	0·974	0·365	0·697	3
Indian soil	*S. rochei*	0·956	0·373	0·967	3
Woodland soil	*S. diastaticus*	0·935	0·382	0·481	3
Marine sediment	*S. griseoruber*	0·913	0·318	1·903	4
Tobacco waste	*S. chromofuscus*	0·864	0·424	2·473	6
Indian soil	Not identified	0·838	0·410	3·348	6
Indian soil	Not identified	0·633	0·405	1·765	5
Indian soil	Not identified	0·482	0·426	3·543	6

Table 9. *Summary of identification of unknown isolates*

Cluster identification*	No. of isolates placed in cluster	Willcox probability: 0·995–0·999	0·990–0·994	0·850–0·989	<0·850
Streptomyces albidoflavus (1)	31	22	2	7	—
S. rochei (12)	8	2	1	5	—
S. diastaticus (19)	7	1	0	6	—
S. chromofuscus (15)	2	0	0	2	—
S. griseoruber (21)	2	0	1	1	—
S. exofoliatus (5)	2	2	0	0	—
Total	52 (81·3%)	27 (42·2%)	4 (6·3%)	21 (32·8%)	12 (18·8%)

* The figures in parentheses indicate the number of the cluster in the identification matrix (Table 2).

the matrix, which is essential before its application, was aided by two other programs (Sneath, 1980*b*, *c*). Identification of strains against the matrix was achieved using the range of assessments provided by the MATIDEN program (Sneath, 1979*a*). All of these programs are designed to operate with percentage positive test data, which are an inevitable product of a numerical phenetic classification, such as that previously constructed for streptomycetes (Williams *et al.*, 1983).

The classification results showed that there was a wide range of phenetic variation in the genus *Streptomyces*; there were 22 major clusters and the other 35% of the strains fell into 51 minor clusters or occurred as single member clusters at the 77·5% S_{SM} level. Therefore, as it was not practicable to construct an identification matrix for all the strains included in the classification study, the major taxa (plus one minor cluster, *Streptomyces fradiae*), were selected. Although these clusters showed no significant overlap using 5% as the critical value, this is not a stringent value compared, for example, to that of 1% used for Gram-positive cocci by Feltham & Sneath (1982). As concluded by Williams *et al.* (1983), these clusters may represent some overlapping variation, rather than entirely clearly defined, well separated species. It is therefore most appropriate to regard them as species-groups until their taxonomic status can be evaluated by other methods. Any identification system can only be as good as the classification on which it is based (Hill *et al.*, 1978).

The characters most diagnostic of the clusters which were included in the matrix were selected as objectively as possible. Some of these, such as morphology and pigmentation characteristics, are not wholly independent because they exist in more than two states, but these effects are minimized by the large number of other characters used. Although the total of 41 characters is quite large, in practice several of them are determined using the same culture. The number of characters is a reflection of the variation within the clusters. Also it is necessary in practice to have at least as many tests as taxa in a matrix (Sneath & Chater, 1978). There is, therefore, no simple, rapid procedure for the objective identification of streptomycetes.

A wide range of criteria for identification is provided by the MATIDEN program (Sneath, 1979*a*), the most frequently used previously being the Willcox coefficient (Willcox *et al.*, 1973). Workers applying this coefficient to other groups of bacteria have required a score of 0·999 for positive identification (Bascomb *et al.*, 1973; Lapage *et al.*, 1973; Willcox *et al.*, 1973; Hill *et al.*, 1978; Wayne *et al.*, 1980; Willcox *et al.*, 1980; Feltham & Sneath, 1982), although a score of 0·900 was accepted by Schindler *et al.* (1979) to identify Gram-negative, fermentative rods. In the present study, a Willcox probability of 0·850 was accepted as one indication of a positive identification. This can be justified by the fact that the identifications were being made to species-groups which were less narrowly defined than the bacterial species studied by previous workers. Also several other identification criteria provided by the MATIDEN program were taken into account. This is important because the Willcox probability alone can give an apparent positive identification if the group to which an unknown belongs is not included in the matrix, due to the normalization process in the calculation of this coefficient (Willcox *et al.*, 1980). However, values for taxonomic distance and its standard error are not deceptive in this way (Sneath, 1979*a*).

It is clearly an advantage to have the output of the MATIDEN program, providing several criteria for evaluating an identification. However, as pointed out by Sneath (1979*a*) for this program and by Lapage *et al.* (1973) for the Willcox probability, it is advisable that an experienced bacterial taxonomist should interpret the computer output. Interpretation of the output from the MATIDEN program (Table 3) requires several value judgements and it is possible that our evaluation of identification data for streptomycetes will improve with experience. With the exception of the Willcox probability, we lack data for the critical comparison of other identification scores.

The need to evaluate matrices by using strains not involved in their construction has been emphasized (Sneath & Sokal, 1973). Therefore, the matrix constructed here was finally assessed by using it to identify unknown isolates from a range of habitats. Using the criteria for identification outlined previously, 81·3% of these isolates were identified to a cluster (species-group). However, the success rate for identification of unknown streptomycetes at the Willcox probability levels applied by previous workers (0·995–0·999) was 42·2% (Table 9). This compares with those of 89·4% for field strains of Gram-negative, fementative rods; 70·8% for Gram-negative, non-fermentative rods (Lapage *et al.*, 1973); 50% for field strains of 'coryneform' bacteria (Hill *et al.*, 1978) and 47% for slow-growing mycobacteria (Wayne *et al.*, 1980). Our comparatively low success rate at the higher probability levels is not surprising. Many of the clusters were sufficiently heterogenous to be regarded as species-groups and not all *Streptomyces* species were included in the matrix.

The need to group the large number of *Streptomyces* 'species' described into larger, more manageable taxa has long been recognized (see Williams *et al.*, 1983). The majority of systems proposed for the construction and recognition of species or species-groups (e.g. Waksman, 1961; Hütter, 1967; Pridham & Tresner, 1974; Pridham, 1976) have used only a few, readily determined characters (such as morphology and pigmentation). Such systems have the attractions of speed and simplicity, but they provide identifications with little information content. The matrix presented here has been objectively constructed and thoroughly tested; we believe its limitations are largely a reflection of the difficulties of streptomycete taxonomy. It is hoped that it will provide a workable system to assist in the identification of many streptomycete strains and it is at present proving useful in several ecological studies in our laboratories. As it will permit many strains to be allocated to broad regions of *Streptomyces*, it should serve

(together with the numerical phenetic classification) as a basis for future improvements in the taxonomy of this important genus. Ultimately it is desirable to construct an identification matrix which includes all valid *Streptomyces* species without increasing the number of tests required beyond practically unacceptable limits.

This work was supported by research grants from the Science and Engineering Research Council.

REFERENCES

BASCOMB, S., LAPAGE, S. P., CURTIS, M. A. & WILLCOX, W. R. (1973). Identification of bacteria by computer: identification of reference strains. *Journal of General Microbiology* **77**, 291–315.

DIETZ, A. & MATHEWS, J. (1971). Classification of *Streptomyces* spore surfaces into five groups. *Applied Microbiology* **21**, 527–533.

FELTHAM, R. K. A. & SNEATH, P. H. A. (1982). Construction of matrices for computer-assisted identification of aerobic Gram-positive cocci. *Journal of General Microbiology* **128**, 713–720.

GAVINI, F., OGER, C., LEFEBVRE, B., IZARD, D. & LECLERC, H. (1982). Development of a computer identification system for coliform strains. *Journal of Applied Bacteriology* **52**, 329–332.

GORDON, R. E. (1968). The taxonomy of soil bacteria. In *The Ecology of Soil Bacteria*, pp. 293–321. Edited by T. R. G. Gray & D. Parkinson. Liverpool: Liverpool University Press.

GYLLENBERG, H. G. (1963). A general method for deriving determination schemes for random collections of microbial isolates. *Annales Academiae scientiarum fennicae, Series A (IV. Biologica)* **69**, 1–23.

GYLLENBERG, H. G. (1970). Factor analytical evaluation of patterns of correlated characteristics in streptomycetes. In *The Actinomycetales*, pp. 101–105. Edited by H. Prauser. Jena: Gustav Fischer Verlag.

GYLLENBERG, H. G. (1974). Identification models for streptomycetes. *Postepy higieny i medycyny doświadczalnej* **28**, 409–418.

GYLLENBERG, H. G. (1976). Application of automation to the identification of streptomycetes. In *Actinomycetes: The Boundary Microorganisms*, pp. 299–321. Edited by T. Arai. Tokyo: Toppan Co.

GYLLENBERG, H. G., WÓZNICKA, W. & KURYŁOWICZ, W. (1967). Application of factor analysis in microbiology. 3. A study of the "yellow series" of streptomycetes. *Annales Academiae scientiarium fennicae, Series A (IV. Biologica)* **114**, 3–15.

GYLLENBERG, H. G., NIEMELA, T. K. & NIEMI, J. S. (1975). A model for automatic identification of streptomycetes. *Postepy higieny i medycyny doświadczalnej* **29**, 357–383.

HANKIN, L., ZUCKER, M. & SANDS, D. C. (1971). Improved solid medium for the detection and enumeration of pectolytic bacteria. *Applied Microbiology* **22**, 205–209.

HILL, L. R. (1974). Theoretical aspects of numerical identification. *International Journal of Systematic Bacteriology* **24**, 494–499.

HILL, L. R. & SILVESTRI, L. G. (1962). Quantitative methods in the systematics of Actinomycetales. III. The taxonomic significance of physiological–biochemical characters and the construction of a diagnostic key. *Giornale di microbiologia* **10**, 1–28.

HILL, L. R., LAPAGE, S. P. & BOWIE, I. S. (1978). Computer assisted identification of coryneform bacteria. In *Coryneform Bacteria*, pp. 181–215. Edited by I. J. Bousfield & A. G. Callely. London: Academic Press.

HÜTTER, R. (1967). *Systematik der Streptomyceten*. Basel: Karger.

KELLEY, R. W. & KELLOGG, S. T. (1978). Computer-assisted identification of anaerobic bacteria. *Applied and Environmental Microbiology* **35**, 507–511.

KÜSTER, E. (1959). Outline of a comparative study of criteria used in the characterization of the actinomycetes. *International Bulletin of Bacteriological Nomenclature and Taxonomy* **9**, 97–104.

KÜSTER, E. & WILLIAMS, S. T. (1964). Production of hydrogen sulphide by streptomycetes and methods for its detection. *Applied Microbiology* **12**, 46–52.

KUTZNER, H. J. (1976). *Methoden zur Untersuchung von Streptomyceten und einigen anderen Actinomyceten*. Darmstadt: Teilsammlung Darmstadt am Institut für Mikrobiologie der Technischen Hochschule.

LAPAGE, S. P. (1974). Practical aspects of probabilistic identification of bacteria. *International Journal of Systematic Bacteriology* **24**, 500–507.

LAPAGE, S. P., BASCOMB, S., WILLCOX, W. R. & CURTIS, M. A. (1970). Computer identification of bacteria. In *Automation, Mechanization and Data Handling in Microbiology*, pp. 1–22. Edited by A. Baillie & R. J. Gilbert. London: Academic Press.

LAPAGE, S. P., BASCOMB, S., WILLCOX, W. R. & CURTIS, M. A. (1973). Identification of bacteria by computer: general aspects and perspectives. *Journal of General Microbiology* **77**, 273–290.

LEFEBVRE, B. & GAVINI, F. (1982). Theory and programming of a computer identification system for coliform strains. *Journal of Applied Bacteriology* **52**, 325–328.

MÖLLER, F. (1962). Quantitative methods in the systematics of Actinomycetales. IV. The theory and application of a probabilistic identification key. *Giornale di microbiologia* **10**, 29–47.

NIEMALÄ, S. I., HOPKINS, J. W. & QUADLING, C. (1068). Selecting an economical binary test battery for a set of microbial cultures. *Canadian Journal of Microbiology* **14**, 271–279.

PRAUSER, H. (1964). Aptness and application of colour codes for exact description of colours of streptomycetes. *Zeitschrift für allgemeine Mikrobiologie* **4**, 95–98.

PRIDHAM, T. G. (1976). Contemporary species concepts in Actinomycetales. In *Actinomycetes: The Boundary Microorganisms*, pp. 163–174. Edited by T. Arai. Tokyo: Toppan Co.

PRIDHAM, T. G. & GOTTLIEB, D. (1948). The utilization of carbon compounds by some Actinomycetales as

aid for species determination. *Journal of Bacteriology* **56**, 107–114.

PRIDHAM, T. G. & TRESNER, H. D. (1974). *Streptomyces* Waksman and Henrici 1943. In *Bergey's Manual of Determinative Bacteriology*, 8th edn, pp. 758–829. Edited by R. E. Buchanan & N. E. Gibbons. Baltimore: Williams & Wilkins.

PRIDHAM, T. G., HESSELTINE, C. W. & BENEDICT, R. G. (1958). A guide for the classification of streptomycetes according to selected groups. Placement of strains in morphological sections. *Applied Microbiology* **6**, 52–79.

RENNIE, R. J. (1980). Dinitrogen-fixing bacteria: computer assisted identification of soil isolates. *Canadian Journal of Microbiology* **26**, 1275–1283.

SCHINDLER, J., DUBEN, J. & LYSENKO, O. (1979). Computer-aided numerical identification of Gram-negative fermentative rods on a desk-top computer. *Journal of Applied Bacteriology* **47**, 45–51.

SHIRLING, E. B. & GOTTLIEB, D. (1968*a*). Co-operative description of type cultures of *Streptomyces*. II. Species descriptions from first study. *International Journal of Systematic Bacteriology* **18**, 69–189.

SHIRLING, E. B. & GOTTLIEB, D. (1968*b*). Co-operative description of type cultures of *Streptomyces*. III. Additional species descriptions from first and second studies. *International Journal of Systematic Bacteriology* **18**, 279–391.

SHIRLING, E. B. & GOTTLIEB, D. (1969). Co-operative descriptions of type cultures of *Streptomyces*. IV. Species descriptions from the second, third and fourth studies. *International Journal of Systematic Bacteriology* **19**, 391–512.

SHIRLING, E. B. & GOTTLIEB, D. (1972). Co-operative description of type strains of *Streptomyces*. V. Additional descriptions. *International Journal of Systematic Bacteriology* **22**, 265–394.

SILVESTRI, L. G., TURRI, M., HILL, L. R. & GILARDI, E. (1962). A quantitative approach to the systematics of actinomycetes based on overall similarity. *Symposia of the Society of General Microbiology* **12**, 333–360.

SKERMAN, V. B. D., MCGOWAN, V. & SNEATH, P. H. A. (1980). Approved lists of bacterial names. *International Journal of Systematic Bacteriology* **30**, 225–420.

SNEATH, P. H. A. (1974). Test reproducibility in relation to identification. *International Journal of Systematic Bacteriology* **24**, 508–523.

SNEATH, P. H. A. (1978). Identification of microorganisms. In *Essays in Microbiology*, pp. 10/1–10/32. Edited by J. R. Norris & M. H. Richmond. Chichester: Wiley.

SNEATH, P. H. A. (1979*a*). BASIC program for identification of an unknown with presence-absence data against an identification matrix of percent positive characters. *Computers and Geosciences* **5**, 195–213.

SNEATH, P. H. A. (1979*b*). BASIC program for character separation indices from an identification matrix of percent positive characters. *Computers and Geosciences* **5**, 349–357.

SNEATH, P. H. A. (1980*a*). BASIC program for the most diagnostic properties of groups from an identification matrix of percent positive characters. *Computers and Geosciences* **6**, 21–26.

SNEATH, P. H. A. (1980*b*). BASIC program for determining the best identification scores possible for the most typical example when compared with an identification matrix of percent positive characters. *Computers and Geosciences* **6**, 27–34.

SNEATH, P. H. A. (1980*c*). BASIC program for determining overlap between groups in an identification matrix of percent positive characters. *Computers and Geosciences* **6**, 267–278.

SNEATH, P. H. A. & CHATER, A. O. (1978). Information content of keys for identification. In *Essays in Plant Taxonomy*, pp. 79–95. Edited by H. E. Street. London: Academic Press.

SNEATH, P. H. A. & JOHNSON, R. (1972). The influence on numerical taxonomic similarities of errors in microbiological tests. *Journal of General Microbiology* **72**, 377–392.

SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman.

TRESNER, H. D. & BACKUS, E. J. (1963). System of color wheels for streptomycete taxonomy. *Applied Microbiology* **11**, 335–338.

TRESNER, H. D., DAVIES, M. C. & BACKUS, E. J. (1961). Electron microscopy of *Streptomyces* spore morphology and its role in species differentiation. *Journal of Bacteriology* **81**, 70–80.

WAKSMAN, S. A. (1961). *The Actinomycetes*, vol. II. Baltimore: Williams & Wilkins.

WAYNE, L. G., KRICHEVSKY, E. J., LOVE, L. L., JOHNSON, R. & KRICHEVSKY, M. I. (1980). Taxonomic probability matrix for use with slowly growing mycobacteria. *International Journal of Systematic Bacteriology* **30**, 528–538.

WELLINGTON, E. M. H. & WILLIAMS, S. T. (1978). Preservation of actinomycete inoculum in frozen glycerol. *Microbios Letters* **6**, 151–157.

WILLCOX, W. B., LAPAGE, S. P., BASCOMB, S. & CURTIS, M. A. (1973). Identification of bacteria by computer: theory and programming. *Journal of General Microbiology* **77**, 317–330.

WILLCOX, W. R., LAPAGE, S. P. & HOLMES, B. (1980). A review of numerical methods in bacterial identification. *Antonie van Leeuwenhoek* **46**, 233–299.

WILLEMSE-COLLINET, M.E., TROMP, T. F. J. & HUIZINGA, T. (1980). A simple and rapid computer-assisted technique for the identification of some selected *Bacillus* species using biochemical tests. *Journal of Applied Bacteriology* **49**, 385–394.

WILLIAMS, S. T., GOODFELLOW, M., ALDERSON, G., WELLINGTON, E. M. H., SNEATH, P. H. A. & SACKIN, M. J. (1983). Numerical classification of *Streptomyces* and related genera. *Journal of Microbiology* **129**, 1743–1813.