

SIMCA Pattern Recognition in the Analysis of Streptomyce Fatty Acids

By G. S. SADDLER,¹ A. G. O'DONNELL,^{2*} M. GOODFELLOW¹ AND D. E. MINNIKIN³

Departments of ¹Microbiology, ²Soil Science and ³Organic Chemistry, The University, Newcastle upon Tyne NE1 7RU, UK

(Received 13 November 1986; revised 28 December 1986)

Biomass from representative strains of *Streptomyces cyaneus* and related taxa was degraded using an alkaline methanolysis procedure. The resultant fatty acid methyl esters were analysed by gas chromatography and the quantitative data examined using the SIMCA statistical package. The *S. cyaneus* strains were recovered in two major clusters. The larger of these contained sixteen strains which had been previously classified as *S. cyaneus* in a numerical phenetic survey. The remaining *S. cyaneus* strains formed a heterogeneous aggregate cluster together with representatives of other streptomycete species. A third cluster encompassed a number of blue-spored streptomycetes from soil. Disjoint principal components analysis in conjunction with cross-validation analyses demonstrated that the *S. cyaneus* group and the blue-spored soil isolates could be represented by statistically stable principal component models. The fact that the initial, numerically circumscribed, *S. cyaneus* taxon was only partially recovered by the numerical analysis of fatty acid data supports the view that this taxon contains strains which are not entirely typical of the cluster, despite their high degree of similarity with other members. These preliminary findings suggest that models based on fatty acid data could be used for identifying unknown streptomycetes to established taxa and for the definition of novel streptomycetes.

INTRODUCTION

Industrial screening programmes designed to find and evaluate new microbial metabolites or biocatalytic processes rely on the selective isolation and characterization of novel organisms (Goodfellow & Williams, 1986). It is important that the circumscription of novel taxa is based on sound taxonomic principles and that data on new and established taxa are objectively analysed.

The importance of chemical data in actinomycete systematics is well established (Minnikin & O'Donnell, 1984). In particular numerical phenetic classifications can be evaluated in light of whole-organism fatty acid data (O'Donnell *et al.*, 1984; Embley *et al.*, 1984). Often, such comparisons have been somewhat subjective with fatty acid profile similarities determined qualitatively. Using standardized growth and analytical conditions, however, reproducible fatty acid profiles, suitable for multivariate statistical analysis can be obtained (Saddler *et al.*, 1986). This approach not only provides an objective alternative to visual comparisons of fatty acid profiles, but can also give valuable taxonomic information for identification and classification below the genus level (O'Donnell, 1985).

Streptomycetes which form a blue aerial spore mass have frequently been grouped together (Hütter, 1962; Pridham & Tresner, 1974; Korn *et al.*, 1978), and in an extensive numerical phenetic survey of the genus *Streptomyces* (Williams *et al.*, 1983) thirteen of the nineteen representative blue-spored taxa were recovered in a major phenon, the *S. cyaneus* cluster. This

Abbreviations: SIMCA, soft independent modelling of class analogy; PCA, principal components analysis; KNN, K nearest neighbours analysis; FAME, fatty acid methyl ester.

taxon also contained strains producing grey or red spores and was heterogeneous with respect to spore chain ornamentation as some strains formed smooth spores, but the remainder formed spiny spores. The major streptomycete clusters recovered by Williams *et al.* (1983) varied in size (6 to 71 strains) and in their homogeneity. It was therefore suggested that they be regarded as species-groups until further information was available.

In the present study the fatty acid profiles of representative *S. cyaneus* strains, additional blue-spored streptomycetes from soil and selected streptomycete reference strains from other numerically defined taxa (see Table 1) were examined together with recently isolated actinomycetes which contained LL-diaminopimelic acid and glycine in the wall peptidoglycan but lacked characteristic sugars (wall chemotype I, Lechevalier & Lechevalier, 1970). The fatty acid data were examined to determine whether the numerical analysis of fatty acid profiles using the SIMCA statistical package (Soft Independent Modelling of Class Analogy; Wold, 1976; Wold & Sjøström, 1977) could clarify the taxonomy of *S. cyaneus* strains and help to identify unknown streptomycetes.

METHODS

Organisms. The test strains (Table 1) were maintained as frozen glycerol suspensions at -25°C (Wellington & Williams, 1978). Binomials in inverted commas are not on the *Approved List of Bacterial Names* (Skerman *et al.*, 1980) and have not been published since the 1st of January 1980. ND, Not determined.

Table 1. *Cluster allocation, morphological properties and pigmentation properties of the test strains*

Cluster and ISP strain histories are after Williams *et al.* (1983). The acidophilic, blue-spored and facultatively autotrophic isolates were supplied by K. E. Simpson, T. Whitham and C. Falconer, respectively (Department of Microbiology, The University of Newcastle upon Tyne, UK). Data on the remaining strains is from Williams *et al.* (1983).

Cluster	Strain no.	Received as:	SIMCA no.	Spore characteristics		
				Mass colour	Chain morphology	Ornamentation
Cluster 18 (<i>S. cyaneus</i>)	5106	<i>S. azureus</i>	47	Blue	Spirales	Smooth
	5140	' <i>S. bicolor</i> '	43	Blue	Spirales	Spiny
	5085	<i>S. chartreusis</i>	41	Blue	Spirales	Spiny
	5144	<i>S. coeruleofuscus</i>	26	Blue	Spirales	Spiny
	5145	<i>S. coeruleorubidus</i>	25	Blue	Spirales	Spiny
	5146	<i>S. coerulescens</i>	23	Blue	Spirales	Spiny
	5107	<i>S. curacoi</i>	35	Blue	Spirales	Spiny
	5108	<i>S. cyaneus</i>	20	Blue	Spirales	Spiny
	5592	' <i>S. peruviensis</i> '	8	Blue	Spirales	Spiny
	5084	<i>S. caelestis</i>	37	Blue	Retinaculi- aperti	Smooth
	5185	<i>S. bellus</i>	32	Blue	Retinaculi- aperti	Spiny
	5483	<i>S. luteogriseus</i>	15	Grey	Spirales	Smooth
	5133	<i>S. resistomycificus</i>	31	Grey	Spirales	Smooth
	5293	<i>S. arenae</i>	45	Grey	Spirales	Spiny
	5013	<i>S. echinatus</i>	40	Grey	Spirales	Spiny
	5482	<i>S. iakyrus</i>	27	Grey	Spirales	Spiny
	5588	<i>S. neyagawaensis</i>	34	Grey	Spirales	Spiny
	5129	<i>S. collinus</i>	24,39	Grey	Retinaculi- aperti	Smooth
	MAV	<i>S. paradoxus</i>	16	Grey	Retinaculi- aperti	Smooth
	5212	<i>S. pseudovenezuelae</i>	28	Grey	Rectiflexi- bilis	Smooth
5467	<i>S. cinnabarinus</i>	19	Red	Spirales	Smooth	
5154	<i>S. fumanus</i>	29	Red	Spirales	Smooth	
5531	' <i>S. pallidus</i> '	17	Red	Spirales	Smooth	
5207	' <i>S. violochromogenes</i> '	44	Red	Spirales	Smooth	
5206	<i>S. janthinus</i>	42	Red	Spirales	Spiny	
5166	<i>S. longisporus</i>	22	Red	Spirales	Spiny	
5277	<i>S. roseoviolaceus</i>	18	Red	Spirales	Spiny	

Table 1 (continued)

Cluster	Strain no.	Received as:	SIMCA no.	Spore characteristics		
				Mass colour	Chain morphology	Ornamentation
Cluster 1A (<i>S. albidoflavus</i>)	5205	<i>S. violarus</i>	46	Red	Spirales	Spiny
	5209	<i>S. violatus</i>	36	Red	Spirales	Spiny
	5227	' <i>S. thermotolerans</i> '	10	Red	Spirales	Spiny
	5228	<i>S. afghaniensis</i>	49	Red	Retinaculi-aperti	Spiny
	5469	<i>S. griseorubiginosus</i>	50	Red	Rectiflexibilis	Smooth
	5042	<i>S. hawaiiensis</i>	48	Yellow	Retinaculi-aperti	Smooth
	5001	<i>S. canescens</i>	51	Yellow	Retinaculi-aperti	Smooth
	5324	<i>S. gougerottii</i>	56	Yellow	Retinaculi-aperti	Smooth
	5077	<i>S. rutgersensis</i>	52	Yellow	Retinaculi-aperti	Smooth
	Cluster 17 (<i>S. griseoviridis</i>)	5420	<i>S. chryseus</i>	54	Red	Spirales
5240		' <i>S. roseo-luteus</i> '	53	Red	Spirales	Smooth
Cluster 19 (<i>S. diastaticus</i>)	5429	<i>S. glomeraurantiacus</i>	55	Grey	Spirales	Smooth
	5256	' <i>S. coralus</i> '	38	White	Rectiflexibilis	Smooth
	5285	<i>S. phaeoviridis</i>	30	White	Rectiflexibilis	Smooth
Cluster 27 (<i>S. viridochromogenes</i>)	5381	<i>S. viridis</i>	21	Blue	Spirales	Spiny
Cluster 38 (<i>S. prasinosporus</i>)	5547	' <i>S. steffisburgensis</i> '	33	Blue	Spirales	Warty
Cluster 61 (<i>S. lavendulae</i>)	5550	<i>S. katrae</i>	58	Red	Rectiflexibilis	Smooth
	5069	<i>S. lavendulae</i>	57	Red	Rectiflexibilis	Smooth
Cluster 71 (<i>S. clavuligerus</i>)	G4	' <i>S. jumonijiensis</i> '	12	Blue	Spirales	Smooth
Single-member cluster	5219	<i>S. amakusaensis</i>	9	Blue	Spirales	Smooth
Blue-spored soil isolates	GSP1		11	Blue	Spirales	ND
	GSP2		5	Blue	Spirales	ND
	GSP3		2	Blue	Spirales	ND
	GSP4		3	Blue	Spirales	ND
	GSP5		7	Blue	Spirales	ND
	GSP6		13	Blue	Spirales	ND
	GSP7		4	Blue	Spirales	ND
	GSP8		14	Blue	Spirales	ND
	GSP11		6	Blue	Spirales	ND
	GSP13		1	Blue	Spirales	ND
Acidophilic soil isolates	AC269		59	White	Rectiflexibilis	ND
	AC272		60	White	Rectiflexibilis	ND
	AC648		61	Grey	Spirales	ND
	AC654		62	Grey	Spirales	ND
	AU10		63	Grey	Rectiflexibilis	Spiny
Facultatively autotrophic soil isolates	AU11		64	Grey	Rectiflexibilis	Spiny
	AU18		65	Grey	Rectiflexibilis	Spiny
	AU19		66	Grey	Rectiflexibilis	Spiny

Preparation of biomass. A thawed glycerol suspension of each strain was used to inoculate 100 ml of modified Sauton's medium (Mordarska *et al.*, 1972) in a 250 ml flask. The acidophilic actinomycetes were cultivated at pH 4.5 and the remaining strains at pH 7.2. The inoculated flasks were shaken (200 r.p.m.) for 3 d at 25 °C. Samples (10 ml) of the starter cultures were used to inoculate 300 ml of the same medium in 1 l flasks which were incubated as before, but for 7 to 10 d as recommended by Saddler *et al.* (1986). Cultures were checked for purity, killed by addition of 3 ml formalin (1%, v/v), harvested by centrifugation at 3000 g for 20 min, washed three times with distilled water, recentrifuged and freeze-dried.

Extraction and analysis of fatty acids. Dried biomass (50 mg) was degraded at 75 °C for 3 h with 5% (w/v) aqueous tetrabutylammonium hydroxide (2 ml) followed by reaction with 2.5% (v/v) iodomethane in dichloromethane (1 ml) at room temperature for 1 h. Evaporation of the lower dichloromethane layer gave crude methyl esters (D. E. Minnikin, A. G. O'Donnell, J. H. Parlett and M. Goodfellow, unpublished data). Analytical and preparative thin-layer chromatography (TLC) of methyl esters were done as described by O'Donnell *et al.* (1982).

Gas chromatography (GC) of light petroleum (b.p. 60–80 °C) solutions of purified fatty acid methyl esters (FAMES) was done using a Shimadzu GC Mini-2 chromatograph equipped for simultaneous analysis on both 50 m bonded OV-1 and 25 m non-bonded Silar 10C (Alltech) fused silica capillary columns. Retention times and the relative proportions of each FAME, expressed as a proportion of the total area, were measured using a Shimadzu CE1B computing integrator. The identity of individual esters was established by comparison with the retention times of standard FAMES mixtures and by their chromatographic behaviour both on the polar and apolar columns. Straight chain standards (C₁₂–C₁₈) were obtained from Sigma. Branched chain standards (C₁₄–C₁₈) were extracted from *Bacillus subtilis* NCIB 3610. The percentage fatty acid composition, expressed relative to the total peak area, for each test strain is presented in Table 2.

Statistical methods. Principal components analysis (PCA; Gower, 1966) and K nearest neighbours analysis (KNN; Fix & Hodges, 1951) of the quantitative fatty acid data were done using the SIMCA package. SIMCA is a commercially available program (Sepanova AB, Enskede, Sweden) which can run on a variety of microcomputers and mainframes. Nearest neighbours analysis, by providing information on the relationships between individual strains, made it possible to assess the relevance of the clusters recovered using principal components analysis (see Discussion for further details). When used for identification, SIMCA describes each taxon or group of strains as a separate principal components model or class. This procedure of disjoint principal components analysis (Wold, 1976; Wold & Sjostrom, 1977) means that the final data matrix can include information on a variety of taxa in which the within group variation is modelled independently and therefore not assumed to be equal.

Separate principal components models (classes) were constructed for both blue-spored soil isolates and *S. cyaneus* strains. The classes were not separately scaled (Blomquist *et al.*, 1979b, c). The number of statistically significant components necessary to describe each class model was determined by cross-validation (Wold, 1978). This latter procedure provided a useful estimate of how much of the variance within a class data matrix accurately described the class (systematic variation) and how much was random noise (Blomquist *et al.*, 1979b, c). In this way, the statistical stability and therefore the predictive value of each class model was determined (Blomquist *et al.*, 1979b, c). Geometrically, a principal component or class model can be seen to represent a region in multidimensional space in which the standard deviation can be used to define the confidence regions of the group. In the present study, linear multiple regression was used to fit every fatty acid profile in the data matrix (Table 1) to each class model (Blomquist *et al.*, 1979b, c). Strains were considered to be members of a class when they fell within the class confidence interval. Detailed mathematical descriptions of the SIMCA method are available (Wold & Sjostrom, 1977; Blomquist *et al.*, 1979a).

RESULTS

Whole-organism methanolysates of the test strains showed the presence of predominant spots (R_F 0.8) corresponding to non-hydroxylated FAMES (Minnikin *et al.*, 1980); hydroxylated fatty acids (R_F 0.2) were not detected. Results of the GC analysis of the non-hydroxylated long-chain components are shown in Table 2. All of the strains produced qualitatively similar profiles with 12-methyltetradecanoic (*anteiso*-15) and 14-methylpentadecanoic acid (*iso*-16) as the major components. However, marked quantitative variations with respect both to the major components and to 12-methyltridecanoic (*iso*-14), 13-methyltetradecanoic (*iso*-15) and hexadecanoic acids (16:0) were observed between strains.

The results obtained by applying PCA to the data in Table 2 are shown in Fig. 1. Results of KNN are imposed on top of the scatter diagram, with each strain linked to its nearest neighbour. This latter analysis highlights relationships which are not evident by PCA alone. Thus, the relationship of *S. cinnabarinus* ISP 5467 (strain 19) with the facultatively autotrophic

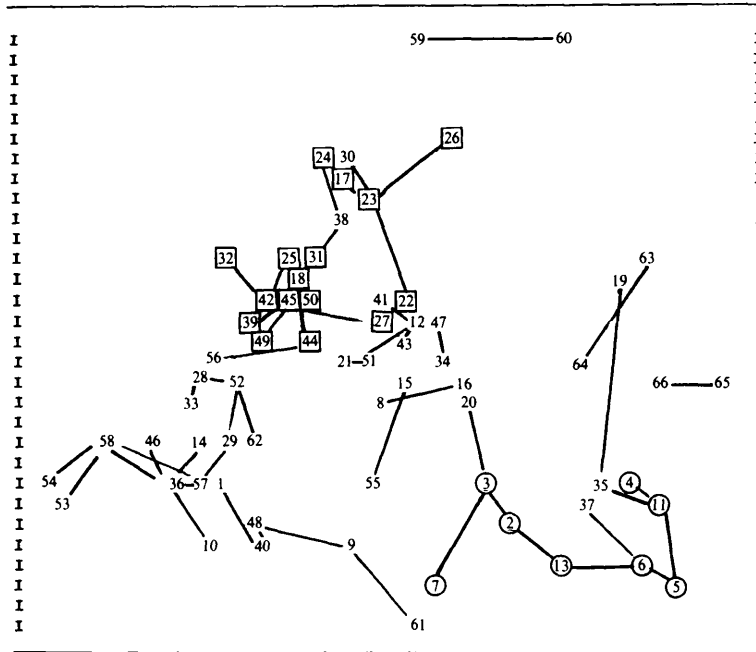


Fig. 1. PCA of FAMES of the test strains (Table 2). The strains have been linked together using KNN. Strains used to define the models (□, *S. cyaneus* strains; ○, blue-spored soil isolates) have been identified.

actinomycetes was more apparent than real as this organism is more closely related to *S. curacoi* ISP 5107 (strain 35).

With the exception of acidophilic cultures AC648 (strain 61) and AC654 (strain 62) and two blue-spored soil isolates GSP13 (strain 1) and GSP8 (strain 14) which appeared to cluster with some of the *S. cyaneus* type strains, the recent soil isolates (Table 1) were recovered as distinct groups. On the basis of the KNN analysis, *S. curacoi* ISP 5107 (strain 35), *S. caelestis* ISP 5084 (strain 37), *S. cyaneus* strains, *S. cinnabarinus* ISP 5467 (strain 19) and *S. cyaneus* ISP 5108 (strain 20) were also associated with the blue-spored soil isolate group.

Three major groups were revealed by KNN. The 19 strains assigned to the largest group included 16 organisms classified in the *S. cyaneus* cluster (Williams *et al.*, 1983). These 16 strains, which were heterogeneous with respect to classical morphological criteria such as spore arrangement and colour of aerial spore mass, were then used to form a SIMCA class model using disjoint principal components analysis (Wold, 1976; Wold & Sjøström, 1977). The second group accommodated 18 strains comprising seven representatives of the *S. cyaneus* cluster, seven streptomyces reference strains, two acidophilic isolates and two blue-spored soil isolates. This apparently heterogeneous group was not studied further.

The final group contained eight of the ten blue-spored soil isolates and four *S. cyaneus* strains (35, 37, 19 and 20). The former were subject to further study using disjoint PCA, whilst the latter were treated as unknowns. Both the *S. cyaneus* and blue-spored soil isolate classes were represented by statistically stable principal components models, significant to three principal components, as determined by cross-validation (Wold, 1978). The 17 strains excluded from the three major groups were recovered as minor groups which, because of their limited size, were unsuitable for further statistical analysis.

The result of plotting the distance (unweighted object standard deviation), calculated using linear multiple regression, of each of the test strains from the class model derived from the blue-spored environmental isolates (x-axis) against that for the class model of the *S. cyaneus* strains (y-axis) is shown in Fig. 2. Imposing 95% confidence intervals for each class divides the plot into

Table 2. Percentage fatty acid composition of the test strains including those grouped into the blue-spored soil isolate and *S. cyaneus* classes
 Fatty acid abbreviations exemplified by: *i*-14, 12-methyltridecanoic acid; 14:0, tetradecanoic acid; *ai*-15, 12-methyltetradecanoic acid; 16:1, hexadecenoic acid; A, unknown ester having relative retention time to 16:0 of 1.05.

SIMCA no.	Strain	Percentage of total fatty acid												
		<i>i</i> -14	14:0	<i>i</i> -15	<i>ai</i> -15	15:0	<i>i</i> -16	16:1	16:0	A	<i>i</i> -17	<i>ai</i> -17	17:0	18:0
Blue-spored soil isolates class														
2	GSP3	10.4	0.0	8.5	18.2	5.8	26.9	1.3	7.5	3.4	1.3	3.6	3.3	1.0
3	GSP4	7.2	0.0	8.8	16.9	5.0	25.1	1.9	8.4	4.6	2.0	4.2	3.1	1.3
4	GSP7	7.8	0.0	7.1	14.8	4.6	32.3	0.0	10.6	1.5	3.0	6.1	2.5	1.3
5	GSP2	14.4	0.0	8.2	14.4	4.8	33.8	0.0	7.1	3.0	1.4	2.6	2.5	1.3
6	GSP11	15.6	0.0	8.1	13.6	5.6	30.6	0.0	7.4	4.7	1.0	1.9	2.6	1.0
7	GSP5	14.5	0.0	6.1	20.3	7.0	22.8	0.0	6.9	6.8	0.0	2.3	3.0	0.0
11	GSP1	7.0	0.0	6.9	15.4	4.0	35.5	2.4	8.3	4.2	1.5	4.2	4.0	1.5
13	GSP6	15.4	1.0	10.3	16.5	8.2	27.3	1.0	7.8	1.9	1.3	2.4	2.0	1.2
<i>S. cyaneus</i> class														
17	5531 ' <i>S. pallidus</i> '	1.2	1.0	5.3	11.7	4.3	13.3	7.1	18.9	2.2	2.0	8.7	1.1	6.3
18	5277 <i>S. roseoviolaceus</i>	1.6	2.3	4.6	13.1	7.3	11.4	5.3	12.1	2.0	1.2	3.8	1.4	1.6
22	5166 <i>S. longisporus</i>	3.6	2.2	5.3	20.0	6.4	21.4	3.8	22.9	1.4	0.0	5.9	1.2	1.0
23	5146 <i>S. coeruleus</i>	1.0	1.4	5.9	14.9	5.2	16.0	5.4	21.5	2.9	2.0	5.8	4.4	2.1
24	5129 <i>S. collinus</i>	0.0	0.0	2.2	10.3	1.2	12.5	2.5	15.4	3.9	0.0	12.5	8.9	2.2
25	5145 <i>S. coeruleorubidus</i>	1.6	1.0	9.8	18.0	6.0	14.6	6.8	17.8	2.4	3.7	8.5	1.3	2.0
26	5144 <i>S. coeruleofuscus</i>	2.4	1.7	6.7	11.4	2.5	19.3	11.0	24.7	2.1	2.2	7.5	1.0	0.0
27	5482 <i>S. iakyrus</i>	2.7	1.6	9.7	17.4	7.8	18.8	5.1	16.0	1.8	1.8	6.7	3.1	1.6
31	5133 <i>S. resistomycificus</i>	2.3	0.0	4.3	20.7	0.0	17.6	3.1	17.0	3.0	3.1	14.5	8.7	0.0
32	5185 <i>S. bellus</i>	2.3	2.9	6.8	18.0	5.7	8.9	5.9	16.7	2.2	1.5	6.2	3.2	0.0
39	5129 <i>S. collinus</i>	4.2	0.0	2.7	16.9	1.4	10.4	2.0	10.4	1.5	1.6	7.7	4.5	1.1
42	5206 <i>S. janthinus</i>	2.6	1.8	7.2	17.9	6.7	14.2	7.0	16.1	2.5	1.4	7.1	1.6	1.4
44	5207 ' <i>S. violochromogenes</i> '	5.1	1.2	7.3	16.3	7.2	12.8	3.6	13.8	1.3	1.8	4.0	1.9	2.2
45	5293 <i>S. arenae</i>	3.2	0.0	3.7	16.9	1.9	12.7	1.6	13.0	2.4	1.9	8.9	5.1	0.0
49	5228 <i>S. afghaniensis</i>	3.0	1.9	7.3	14.7	4.3	10.1	6.0	8.2	1.2	0.0	3.4	1.8	0.0
50	5469 <i>S. griseorubiginosus</i>	2.0	3.7	4.3	16.0	3.1	13.7	4.6	13.7	1.7	1.7	5.6	2.2	0.0
Strains not assigned to class models														
8	5592 ' <i>S. peruvianensis</i> '	3.9	0.0	15.4	15.7	7.3	17.8	4.9	8.3	3.0	3.1	7.1	1.0	1.2
10	5227 ' <i>S. thermotolerans</i> '	7.7	1.4	11.9	24.5	11.8	11.9	1.9	5.5	1.7	1.2	3.0	1.9	2.1
15	5483 <i>S. luteogriseus</i>	4.0	1.0	5.4	17.9	7.4	21.0	5.0	12.5	2.6	1.8	5.1	4.7	1.5
16	MAV <i>S. paradoxus</i>	5.2	2.2	17.5	13.4	5.7	20.0	4.3	12.3	1.7	2.6	5.3	1.2	0.0
19	5467 <i>S. cinnabarinus</i>	1.2	0.0	4.9	9.2	1.8	31.0	7.5	14.8	3.4	3.7	11.3	1.4	1.9
20	5108 <i>S. cyaneus</i>	4.1	0.0	3.6	17.2	2.1	24.8	1.3	10.0	5.1	1.6	6.4	6.8	1.4
28	5212 <i>S. pseudovenezuelae</i>	1.6	0.0	4.4	23.3	1.1	13.4	6.1	11.0	3.2	1.5	12.0	1.5	0.0
29	5154 <i>S. fumanus</i>	2.9	1.0	5.4	25.0	3.3	16.8	2.9	7.6	4.2	1.9	12.6	5.0	1.3
34	5588 <i>S. neyagawaensis</i>	3.6	1.5	11.7	16.6	3.5	22.2	9.7	12.8	2.3	4.2	6.1	2.3	1.0

35	5107	<i>S. curacoi</i>	6-8	1-0	8-0	14-1	3-3	31-2	5-9	7-6	2-4	2-4	6-3	1-0	1-1
36	5209	<i>S. violatus</i>	5-9	3-2	8-8	28-0	8-3	12-1	5-1	9-0	2-1	1-1	5-0	1-0	1-0
37	5084	<i>S. caelestis</i>	20-7	1-0	4-7	13-4	3-0	25-7	3-5	11-7	2-8	1-1	2-3	3-3	1-0
40	5013	<i>S. caelinatus</i>	7-8	1-1	6-3	26-9	3-4	17-7	2-2	6-1	1-6	1-1	6-5	0-0	0-0
41	5085	<i>S. chatreusis</i>	1-9	0-0	6-6	10-1	3-0	15-8	5-9	8-4	3-9	2-3	8-7	2-0	0-0
43	5140	<i>'S. bicolor'</i>	2-7	0-0	5-3	14-5	2-6	20-7	6-7	9-9	2-6	2-3	12-3	1-9	1-1
46	5205	<i>S. azureus</i>	5-0	4-8	7-9	24-7	8-4	8-9	2-9	9-0	1-4	0-0	3-3	1-1	0-0
47*	5106	<i>S. azureus</i>	4-4	1-0	7-2	15-8	4-3	20-5	8-4	14-4	1-8	1-7	5-5	2-0	1-3
48	5042	<i>S. hawaiiensis</i>	5-2	0-0	3-0	28-5	1-9	18-8	0-0	5-9	4-1	1-4	8-3	3-7	0-0
51	5001	<i>S. canescens</i>	3-2	2-0	7-3	18-1	2-7	19-0	8-5	10-8	3-9	0-0	7-9	3-9	2-2
52	5077	<i>S. rugosens</i>	3-3	0-0	4-2	25-9	3-4	17-1	6-8	12-1	3-9	0-0	11-8	5-1	1-5
56	5324	<i>S. gougerotti</i>	5-8	1-7	6-3	21-1	10-4	10-4	5-8	12-4	2-1	1-1	4-4	6-7	2-9
53	5240	<i>'S. roseo-luteus'</i>	3-1	0-0	5-3	35-1	4-4	12-4	3-4	8-4	4-1	0-0	13-9	2-3	0-0
54	5420	<i>S. chryseus</i>	5-5	0-0	3-7	33-3	5-5	9-7	8-9	8-4	4-3	1-3	7-7	2-9	1-8
30*	5285	<i>S. phaeoviridis</i>	3-1	1-6	4-6	15-8	4-5	13-3	5-6	25-1	2-5	2-2	6-2	2-9	2-1
38*	5256	<i>'S. coralus'</i>	1-3	0-0	4-1	16-2	1-0	16-4	3-3	15-9	4-3	3-0	11-2	11-2	1-8
55	5429	<i>S. glomeraurantiacus</i>	7-0	0-0	8-2	23-6	10-4	20-9	1-8	11-2	2-6	1-4	4-2	3-3	1-9
21	5381	<i>'S. viridis'</i>	3-4	1-0	10-0	18-2	2-8	17-1	2-3	11-0	3-5	4-9	9-9	2-5	2-3
33	5547	<i>'S. steffsburgensis'</i>	2-6	1-0	9-6	22-3	8-7	12-2	7-7	7-6	4-0	1-4	14-4	0-0	2-0
57	5069	<i>S. lavendulae</i>	2-7	1-1	4-6	29-7	3-9	18-1	4-9	7-6	3-2	1-1	10-3	5-9	1-1
58	5550	<i>S. katrae</i>	1-6	1-0	5-6	32-9	1-0	13-3	5-7	9-8	2-7	1-7	9-2	5-5	1-1
12	G4	<i>'S. jumonijjensis'</i>	3-0	0-0	5-4	15-2	1-1	20-5	7-6	11-8	5-1	2-3	10-4	4-2	0-0
9	5219	<i>S. amakusaensis</i>	6-2	0-0	3-7	25-1	5-3	24-5	2-8	5-0	4-8	1-3	9-3	4-1	1-2
1	GSP13		6-7	1-2	4-6	26-5	6-6	14-8	0-0	10-1	0-0	2-5	10-6	2-4	1-9
14	GSP8		5-6	2-5	7-4	25-0	3-4	13-9	8-6	7-1	3-0	1-8	6-6	5-4	0-0
59	AC269		0-0	0-0	3-8	11-6	0-0	16-7	6-8	23-0	9-4	0-0	7-0	22-4	0-0
60	AC272		0-0	0-0	4-5	10-9	0-0	25-7	7-6	26-7	0-0	0-0	6-1	21-5	0-0
61	AC648		12-4	1-2	5-3	26-6	6-4	26-6	0-0	7-9	0-0	1-6	6-3	0-0	0-0
62	AC654		6-6	1-3	0-0	28-0	7-7	17-9	4-4	14-2	3-0	0-0	8-6	1-7	0-0
63	AU10		1-5	1-1	11-3	9-3	6-0	29-3	0-0	18-8	0-0	8-4	9-6	3-8	0-0
64	AU11		3-0	1-0	10-9	14-4	6-2	29-0	0-0	15-5	0-0	4-2	11-7	3-4	0-0
65	AU18		4-9	1-4	6-7	11-8	9-0	36-2	0-0	17-3	0-0	2-0	6-2	2-8	0-0
66	AU19		5-0	1-8	7-3	12-1	9-3	32-1	0-0	16-4	0-0	2-8	7-5	2-5	0-0

* Denotes strains recovered in the *S. cyanus* class that were not originally assigned to the group.

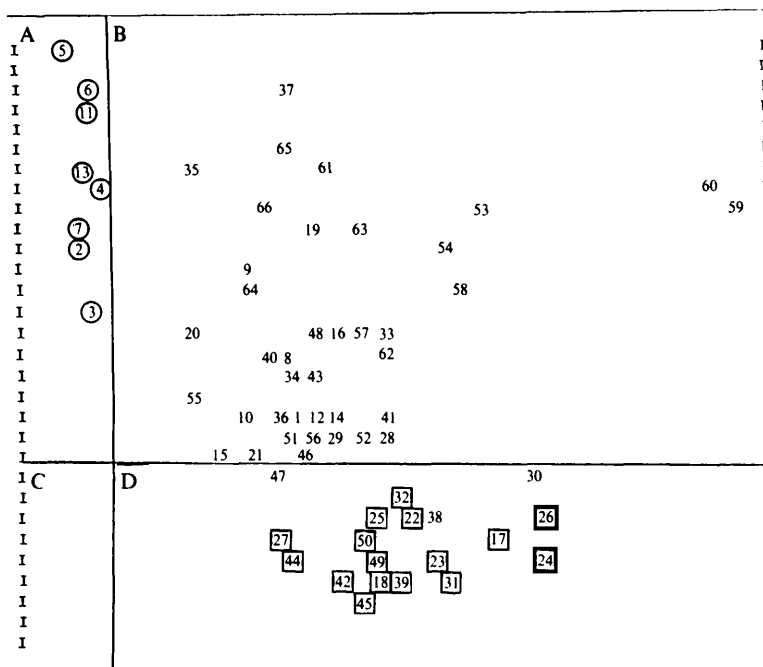


Fig. 2. Residual SD plot of the blue-spored environmental class (x-axis) against the *S. cyaneus* class (y-axis). Region A contains strains uniquely identified to the environmental isolates group (class); region D contains strains recovered in the *S. cyaneus* group; region C contains intermediates between the two taxa and region B consists of strains not identified to any group. All the strains assigned to the soil isolates class were identified to the group, as were all of the strains assigned to the *S. cyaneus* class; however, in addition strains 30, 38 and 47 were also identified to this group. Strains used to define the models (\square , *S. cyaneus* strains; \circ , blue-spored soil isolates) have been identified.

four distinct regions. Strains recovered in region A can be identified as belonging to the blue-spored environmental isolate class and those in region D can be assigned to the *S. cyaneus* class. Organisms allocated to region C in such plots represent intermediate strains which cannot be unambiguously assigned to either group. Region B, an area outside the class confidence intervals of both taxa, contains strains which are distinct from the two class models.

It is evident that the blue-spored soil isolates form a distinct group with no overlapping strains (region A; Fig. 2). Similarly all 16 members of the *S. cyaneus* class were recovered in region D, together with *S. azureus* ISP 5106 (strain 47; cluster 18), '*S. coralus*' ISP 5265 (strain 30; cluster 19) and *S. phaeoviridis* ISP 5285 (strain 38; cluster 19). Since none of the test strains were recovered in region C, the *S. cyaneus* and soil isolate SIMCA classes are distinct and show no overlap. The strains assigned to region B included 17 strains originally assigned to cluster 18 (*S. cyaneus*), three to sub-cluster 1A (*S. albidoflavus*), two to cluster 17 (*S. griseoviridis*), one to cluster 19 (*S. diastaticus*), one to cluster 27 (*S. viridochromogenes*), one to cluster 38 (*S. prasinosporus*), two to cluster 61 (*S. lavendulae*), one to cluster 71 (*S. clavuligerus*) and one strain to a single member cluster (*S. amakusaensis*). The acidophilic and the facultatively autotrophic strains along with two blue-spored soil isolates were also recovered in region B.

DISCUSSION

The detection of major amounts of saturated *iso*- and *anteiso*-branched fatty acids of between 14 and 18 carbons in all the test strains was in good agreement with the results of previous studies on streptomycetes (Lechevalier, 1977; Kroppenstedt & Kutzner, 1978; Kroppenstedt, 1985). It

is difficult to evaluate the significance of fatty acid profiles in streptomyces systematics generally, since much of the early work was carried out on poorly described strains examined using a variety of experimental methods. Further, the overall similarities in streptomyces fatty acid profiles make it necessary to use numerical analyses to exploit the data for taxonomic purposes.

Despite the potential of quantitative analyses of fatty acid profiles in bacterial systematics most studies still rely on a qualitative interpretation of data where test strains are grouped according to whether they possess 'large amounts' or 'small amounts' of particular components. With the exception of a few numerically based studies (O'Donnell, 1985) this approach has generally resulted in fatty acid data being used to define taxa at generic and suprageneric levels. The development of simpler extraction and analysis systems together with the ability to obtain reproducible profiles (Saddler *et al.*, 1986) provides a readily available but under-exploited source of taxonomic information.

The numerical approach adopted here involved two distinct processes. Initially, PCA was used to classify the test strains. The two-dimensional plot obtained gave a first approximation of the relationships between the strains, but the overall between-strain variation is rarely recovered exactly and inter-strain distances may be misrepresented (Sneath & Sokal, 1973). The latter recommended that ordination methods be used in conjunction with numerically defined phenograms since these are more reliable for describing relationships between individual strains. The alternative approach adopted in the present study involved superimposing the results of a nearest neighbours analysis on the principal components plots.

It was particularly interesting that both the facultatively autotrophic and acidophilic strains were distinguished from the remaining test strains in the numerical analysis of fatty acid data. However, comparative studies on acidotolerant actinomycetes, that is, strains which grow at pH 4.5 and 7.2, are needed to measure the effect of pH on fatty acid composition. The separation of the acidophilic from the neutrophilic actinomycetes and their recovery in more than one centre of variation was in good agreement with earlier findings (Williams *et al.*, 1983; Lonsdale, 1985). Indeed, acidophilic actinomycetes are now known to be taxonomically very diverse (Lonsdale, 1985). The assignment of *S. azureus* ISP 5106 (strain 47), '*S. coralus*' ISP 5256 (strain 30) and *S. phaeoviridis* ISP 5285 (strain 38) to the *S. cyaneus* class was also interesting as all three strains were classified in the *S. cyaneus* cluster in the S_J , UPGMA analysis (Williams *et al.*, 1983); in the corresponding S_{SM} analysis the latter two cultures were classified in the *S. diastaticus* cluster.

Two statistically significant principal component models were obtained when the fatty acid data were analysed by cross-validation (Wold, 1978). The latter operates by a process of removing and inserting elements from the data matrix and provides a useful method of separating the variable or object independent systematic variation from random noise. The concept of describing reproducibility in terms of systematic variation and random noise has no parallel in discriminant analysis (MacFie & Gutteridge, 1982) and is discussed in detail by Blomquist *et al.* (1979a). This approach provided a means of evaluating the statistical stability and therefore the predictive potential of the groups recovered using PCA and KNN. This practice is of primary importance in any bacterial screening or identification system and in the evaluation of media formulations devised for the isolation of defined groups of bacteria (Williams *et al.*, 1984; Goodfellow & Williams, 1986). A problem arises, however, over whether a cluster or class deemed to be statistically significant is in fact valid microbiologically as judged by existing taxonomic criteria.

The test strains used to define the SIMCA class models were selected from *a priori* information on the numerical phenetic and chemotaxonomic relationships of the strains. Consequently, strains which appeared to be closely associated with those used to define a given class model on the basis of their fatty acid data (Fig. 1) but which showed no relationship to them with respect to other criteria were not included. These strains were therefore treated as unknowns. Such an approach represents a working compromise since the question of strain selection for use in identification systems remains an as yet unresolved problem in microbial systematics and is in need of further investigation.

The lack of congruence between classifications based on different analytical techniques can be expected to become increasingly common in bacterial systematics given the availability of an increasing number of techniques for generating taxonomic information on the same group of organisms (Sneath, 1985). When compared with the numerical phenetic classification (Williams *et al.*, 1983) the results presented here support the inclusion of blue-, grey- and red-spored strains within *S. cyaneus* and the separation of this taxon from other numerically circumscribed streptomycete taxa. Further, the fact that not all of the strains received as *S. cyaneus* had quantitatively similar fatty acid profiles lends weight to the view that the numerically defined major streptomycete clusters may encompass strains that are not entirely typical of the cluster despite their high overall similarity with other members (Williams *et al.*, 1983). The present data indicate that the *S. cyaneus* cluster is heterogeneous but further comparative studies using, for example, serological and nucleic acid pairing techniques need to be applied to determine the taxonomic status of the two statistically significant taxa obtained. Ridell *et al.* (1986) found that streptomycetes which shared a high number of precipitinogens belonged to the same numerically defined cluster, while strains less serologically related usually fell into different clusters. Similarly, Mordarski *et al.* (1986) obtained good congruence between numerical phenetic and DNA homology data with respect to strains assigned to the *S. albidoflavus* sub-cluster.

The increased use of analytical instrumentation in microbial systematics provides taxonomists with valuable new tools with which to study relationships between organisms. In many instances the data generated are multivariate and require the development and application of suitable data handling procedures. Since SIMCA pattern recognition techniques accommodate taxa with different within-group heterogeneities and uses them independently, they are likely to have an increasingly important role in microbial systematics. The data presented here demonstrate the potential of SIMCA pattern recognition in improving the interpretation of bacterial fatty acid data. As an identification or screening system the success of such a procedure depends on the construction and testing of statistically, as well as microbiologically, significant principal components models. The confidence limits set on these models can be tailored to suit particular problems although it is likely that the analysis of closely related taxa, in which there is a gradual shift from one species to another along a spectrum, will require the application of more stringent confidence intervals than the 95% interval used here. These problems will only be met by the application of such techniques to a large number of well defined taxa. It is possible that no one analytical or statistical technique will provide all the answers but the present data suggest that by analysing quantitative fatty acid data using SIMCA pattern recognition it should be possible to identify unknown actinomycetes to established taxa.

The authors are indebted to the Science and Engineering Research Council (grant GR/C/18830) for supporting part of this work and to Professor S. Wold for providing the SIMCA package. G. S. Saddler gratefully acknowledges receipt of an SERC/CASE award tenable at the University of Newcastle and ICI (Natural Products Group).

REFERENCES

- BLOMQUIST, G., JOHANSSON, E., SÖDERSTROM, B. & WOLD, S. (1979*a*). Data analysis of pyrolysis chromatograms by means of SIMCA pattern recognition. *Journal of Analytical and Applied Pyrolysis* **1**, 53–65.
- BLOMQUIST, G., JOHANSSON, E., SÖDERSTROM, B. & WOLD, S. (1979*b*). Reproducibility of pyrolysis–gas chromatographic analyses of the mould *Penicillium brevi-compactum*. *Journal of Chromatography* **173**, 7–17.
- BLOMQUIST, G., JOHANSSON, E., SÖDERSTROM, B. & WOLD, S. (1979*c*). Classification of fungi by means of pyrolysis–gas chromatography pattern recognition. *Journal of Chromatography* **173**, 19–32.
- EMBLEY, T. M., GOODFELLOW, M., MINNIKIN, D. E. & O'DONNELL, A. G. (1984). Lipid and wall amino acid composition in the classification of *Rothia dentocariosa*. *Zentralblatt für Bakteriologie, Mikrobiologie und Hygiene, series A* **257**, 285–294.
- FIX, E. & HODGES, J. L. (1951). Discriminatory analysis, non-parametric discrimination. *USAF School of Aviation Medicine, Randolph Field, Texas, Project 21-49-004, Report 4*.
- GOODFELLOW, M. & WILLIAMS, S. T. (1986). New strategies for the selective isolation of industrially important bacteria. *Biotechnology and Genetic Engineering Reviews* **4**, 213–262.
- GOWER, J. C. (1966). Some distance properties of

- latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.
- HÜTTER, R. (1962). Zur Systematik der Actinomyceten. 8. Quirlbildende Streptomyceten. *Archiv für Mikrobiologie* **43**, 356–391.
- KORN, F., WEINGARTNER, B. & KUTZNER, H. J. (1978). A study of twenty actinophages: morphology, serology, relationship and host range. In *Genetics of Actinomycetales*, pp. 251–270. Edited by E. Freerksen, I. Tárnok & J. H. Thumin. Stuttgart: Gustav Fischer Verlag.
- KROPFENSTEDT, R. M. (1985). Fatty acid and menaquinone analysis of actinomycetes and related organisms. In *Chemical Methods in Bacterial Systematics*, pp. 173–199. Edited by M. Goodfellow & D. E. Minnikin. London: Academic Press.
- KROPFENSTEDT, R. M. & KUTZNER, H. J. (1978). Biochemical taxonomy of some problem actinomycetes. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene, Abteilung I, Supplement* **6**, 125–133.
- LECHEVALIER, M. P. (1977). Lipids in bacterial taxonomy – a taxonomist's viewpoint. *Critical Reviews in Microbiology* **5**, 109–210.
- LECHEVALIER, M. P. & LECHEVALIER, H. A. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. *International Journal of Systematic Bacteriology* **20**, 435–443.
- LONSDALE, J. T. (1985). *Aspects of the biology of acidophilic actinomycetes*. PhD thesis, The University of Newcastle upon Tyne, UK.
- MACFIE, H. J. H. & GUTTERIDGE, C. S. (1982). Comparative studies on some methods for handling quantitative data generated by analytical pyrolysis. *Journal of Analytical and Applied Pyrolysis* **4**, 175–204.
- MINNIKIN, D. E. & O'DONNELL, A. G. (1984). Actinomycete envelope lipid and peptidoglycan composition. In *The Biology of the Actinomycetes*, pp. 337–388. Edited by M. Goodfellow, M. Mordarski & S. T. Williams. London: Academic Press.
- MINNIKIN, D. E., HUTCHINSON, I. G., CALDICOTT, A. B. & GOODFELLOW, M. (1980). Thin-layer chromatography of methanolysates of mycolic acid-containing bacteria. *Journal of Chromatography* **188**, 221–233.
- MORDARSKA, H., MORDARSKI, M. & GOODFELLOW, M. (1972). Chemotaxonomic characters and classification of some nocardioform bacteria. *Journal of General Microbiology* **71**, 77–86.
- MORDARSKI, M., GOODFELLOW, M., WILLIAMS, S. T. & SNEATH, P. H. A. (1986). Evaluation of species groups in the genus *Streptomyces*. In *Biological, Biochemical and Biomedical Aspects of Actinomycetes*. Edited by G. Szabó, S. Biró & M. Goodfellow. Budapest: Kiado Press.
- O'DONNELL, A. G. (1985). Numerical analysis of chemotaxonomic data. In *Computer-assisted Bacterial Systematics*, pp. 403–414. Edited by M. Goodfellow, D. Jones & F. G. Priest. London: Academic Press.
- O'DONNELL, A. G., GOODFELLOW, M. & MINNIKIN, D. E. (1982). Lipids in the classification of *Nocardioideae*: reclassification of *Arthrobacter simplex* (Jensen) Lochhead in the genus *Nocardioideae* (Prauser) emend. O'Donnell *et al.* as *Nocardioideae simplex comb. nov.* *Archives of Microbiology* **133**, 323–329.
- O'DONNELL, A. G., MINNIKIN, D. E., GOODFELLOW, M. & PIOT, P. (1984). Fatty acid, polar lipid and wall amino acid composition of *Gardnerella vaginalis*. *Archives of Microbiology* **138**, 68–71.
- PRIDHAM, T. G. & TRESNER, H. D. (1974). *Streptomyces* Waksman and Henrici 1943, 339. In *Bergey's Manual of Determinative Bacteriology*, 8th edn, pp. 748–829. Edited by R. E. Buchanan & N. E. Gibbons. Baltimore: Williams & Wilkins.
- RIDELL, M., WALLERSTRÖM, G. & WILLIAMS, S. T. (1986). Immunodiffusion analyses of phenetically defined strains of *Streptomyces*, *Streptovercillium* and *Nocardioopsis*. *Systematic and Applied Microbiology* **8**, 24–27.
- SADDLER, G. S., GOODFELLOW, M., MINNIKIN, D. E. & O'DONNELL, A. G. (1986). Influence of the growth cycle on the fatty acid and menaquinone composition of *Streptomyces cyaneus* NCIB 9616. *Journal of Applied Bacteriology* **60**, 51–56.
- 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. (1985). Future of numerical taxonomy. In *Computer-assisted Bacterial Systematics*, pp. 415–431. Edited by M. Goodfellow, D. Jones & F. G. Priest. London: Academic Press.
- SNEATH, P. H. A. & SOKAL, R. R. (1973). *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. San Francisco: W. H. Freeman.
- WELLINGTON, E. M. H. & WILLIAMS, S. T. (1978). Preservation of actinomycete inoculum in frozen glycerol. *Microbios Letters* **6**, 151–157.
- 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 General Microbiology* **129**, 1743–1813.
- WILLIAMS, S. T., GOODFELLOW, M. & VICKERS, J. C. (1984). New microbes from old habitats? In *The Microbe 1984*, II, *Prokaryotes and Eukaryotes*, pp. 219–256. Edited by D. P. Kelly and N. G. Carr. Cambridge: University Press.
- WOLD, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition* **8**, 127–139.
- WOLD, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* **20**, 397–405.
- WOLD, S. & SJÖSTROM, M. J. (1977). SIMCA: a method for analysing chemical data in terms of similarity and analogy. In *Chemometrics: Theory and Application*, pp. 243–282. Edited by B. Kowalski. ACS Symposium Series no. 52, Washington, DC: American Chemical Society.